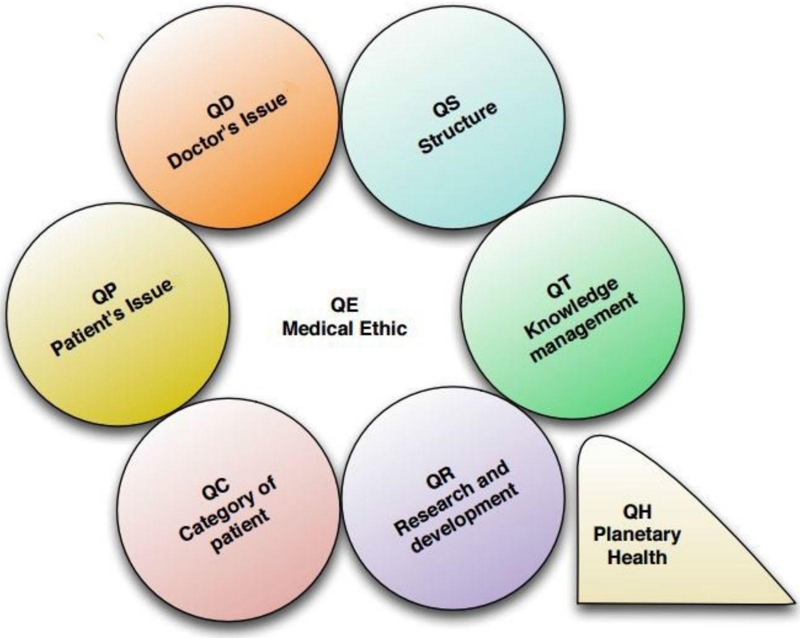
UNIVERSITY OF LIÈGE, BELGIUM

DOCTORAL THESIS

Towards a system of concepts for Family Medicine. Multilingual indexing in General Practice/ Family Medicine in the era of Semantic Web



Dr. Marc JAMOULLE

A thesis submitted in fulfillment of the requirements for the degree of Doctor in medical sciences

Département universitaire de Médecine générale

LIÈGE université

Unité de recherche Soins primaires et Santé

Academic year 2017- 2018

Towards a system of concepts for family medicine
Multilingual indexing in General Practice/ Family Medicine in the era of semantic web

Marc Jamoulle. MD.

Doctoral Thesis
University of Liège, Belgium
Département de médecine générale
Unité de recherche Soins Primaires et Santé
Academic year 2017-2018
December 8, 2017

- Chairman of the committee:
  - Professor Didier Giet, University of Liège, Belgium
- Supervisors:
  - Professor Marc Vanmeerbeek, University of Liège, Belgium
  - Associate Professor Ashwin Ittoo, University of Liège, Belgium
- Other members of the jury
  - Professor Robert Vander Stichele, University of Ghent, Belgum
  - Professor Anne-Françoise Donneau, University of Liège, Belgium
  - Professor Frank Buntinx, University of Leuven, Belgium & Maastricht, The Netherlands
  - Professor Kees van Boven, University of Nijmegen, The Netherlands
  - Professor Laurent Letrilliart, University of Lyon, France
  - Professor Michel Roland, University of Brussels, Belgium
  - Professor Stefan Darmoni, University of Rouen, France

Declaration of Authorship

I, Dr. Marc JAMOULLE, declare that this thesis titled, "Towards a system of concepts for Family Medicine.
Multilingual indexing in General Practice/ Family Medicine in the era of Semantic Web" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Marc Jamoulle MD, Family physician

Date: December 8, 2017

"L'information est à la médecine de famille ce que la technologie est à la médecine spécialisée"

"Information is to general practice what technology is to specialized medicine"

Adapted from Van Dormael (2001)

"The purpose of research is to discover and not to prove"

Hess (2004)

"Right term should be used for a clear communication "

Confucius (551 - 479 BC)

UNIVERSITY OF LIÈGE, BELGIUM

Executive Summary

Faculty of Medicine
Département Universitaire de Médecine Générale.
Unité de recherche Soins Primaires et Santé

Doctor in biomedical sciences

**Towards a system of concepts for Family Medicine.
Multilingual indexing in General Practice/ Family Medicine in the era of Semantic Web**

by Dr. Marc JAMOULLE

Introduction

This thesis is about giving visibility to the often overlooked work of family physicians and consequently, is about grey literature in General Practice and Family Medicine (GP/FM). It often seems that conference organizers do not think of GP/FM as a knowledge-producing discipline that deserves active dissemination. A conference is organized, but not much is done with the knowledge shared at these meetings. In turn, the knowledge cannot be reused or reapplied. This these is also about indexing. To find knowledge back, indexing is mandatory. We must prepare tools that will automatically index the thousands of abstracts that family doctors produce each year in various languages. And finally this work is about semantics[1]. It is an introduction to health terminologies, ontologies, semantic data, and linked open data. All are expressions of the next step: Semantic Web for health care data. Concepts, units of thought expressed by terms, will be our target and must have the ability to be expressed in multiple languages. In turn, three areas of knowledge are at stake in this study: (i) Family Medicine as a pillar of primary health care, (ii) computational linguistics, and (iii) health information systems.

Aim

- To identify knowledge produced by General practitioners (GPs) by improving annotation of grey literature in Primary Health Care
- To propose an experimental indexing system, acting as draft for a standardized table of content of GP/GM
- To improve the searchability of repositories for grey literature in GP/GM.

[1]For specific terms, see the Glossary page 257

Methods

The first step aimed to design the taxonomy by identifying relevant concepts in a compiled corpus of GP/FM texts. We have studied the concepts identified in nearly two thousand communications of GPs during conferences. The relevant concepts belong to the fields that are focusing on GP/FM activities (e.g. teaching, ethics, management or environmental hazard issues).

The second step was the development of an on-line, multilingual, terminological resource for each category of the resulting taxonomy, named Q-Codes. We have designed this terminology in the form of a lightweight ontology, accessible on-line for readers and ready for use by computers of the semantic web. It is also fit for the Linked Open Data universe.

Results

We propose 182 Q-Codes in an on-line multilingual database (10 languages) (www.hetop.eu/Q) acting each as a filter for Medline. Q-Codes are also available under the form of Unique Resource Identifiers (URIs) and are exportable in Web Ontology Language (OWL). The International Classification of Primary Care (ICPC) is linked to Q-Codes in order to form the Core Content Classification in General Practice/Family Medicine (3CGP). So far, 3CGP is in use by humans in pedagogy, in bibliographic studies, in indexing congresses, master theses and other forms of grey literature in GP/FM. Use by computers is experimented in automatic classifiers, annotators and natural language processing.

Discussion

To the best of our knowledge, this is the first attempt to expand the ICPC coding system with an extension for family physician contextual issues, thus covering non-clinical content of practice. It remains to be proven that our proposed terminology will help in dealing with more complex systems, such as MeSH, to support information storage and retrieval activities. However, this exercise is proposed as a first step in the creation of an ontology of GP/FM and as an opening to the complex world of Semantic Web technologies.

Conclusion

We expect that the creation of this terminological resource for indexing abstracts and for facilitating Medline searches for general practitioners, researchers and students in medicine will reduce loss of knowledge in the domain of GP/FM. In addition, through better indexing of the grey literature (congress abstracts, master's and doctoral theses), we hope to enhance the accessibility of research results and give visibility to the invisible work of family physicians.

Résumé exécutif

Université de Liège. Département de médecine générale.
Unité de recherche Soins Primaires et Sante.

Thése de doctorat en sciences médicales

Vers un système de concepts en médecine de famille. Indexation multilingue en Médecine Générale et de Famille à l'ère du Web sémantique.

Marc Jamoulle, médecin de famille 8 décembre 2017

Introduction

Rendre visible l'invisible contribution des médecins de famille à la connaissance est notre motto . Nous évoquons ici la littérature grise en médecine générale et de famille (MG/MF). La MG/MF ne semble pas considérée comme une discipline productrice de connaissances méritant une diffusion active. De nombreux congres sont organisés, mais la connaissance, assemblée un instant, disparait et n'est ni partagée ni réutilisée.

Pour conserver cette connaissance, l'indexation est obligatoire. Nous devons préparer des outils qui indexeront automatiquement les milliers de résumés que les médecins de famille produisent chaque année dans diverses langues. Le Web sémantique [2] semble offrir un horizon prometteur pour cette tâche qui semblait impossible auparavant.

Ce travail est une introduction aux terminologies de la santé, aux ontologies, aux données sémantiques qui préparent la prochaine étape: le Web sémantique pour les données de soins de santé. Notre point de départ sont les concepts et non plus les termes. Ces concepts doivent être exprimés dans les langues du quotidien de la relation médecin patient.

Objectifs

Considérant le domaine de la Médecine générale et de famille

- Identifier les connaissances produites par les généralistes en améliorant l'annotation de la littérature grise.
- Proposer un système d'indexation expérimental utilisable comme table des matières normalisée.
- Améliorer la capacité d'interrogation de la littérature grise.

[2] Un glossaire anglais/français des termes techniques est disponible à la page 257

Méthodes

La première étape a permis de concevoir une taxonomie en identifiant les concepts pertinents dans un corpus de textes en MF/MG. Nous avons étudié les concepts identifiés dans près de deux mille communications de médecins généralistes lors de congrès. Nous y avons exploré le contexte du métier, comme les questions d'enseignement, d'éthique, de gestion ou de structure. La deuxième étape a été le développement d'une ressource terminologique multilingue en ligne pour chaque catégorie de la taxonomie résultante, nommée Q-Code. Nous avons conçu cette terminologie sous la forme d'une ontologie légère, accessible en ligne pour les lecteurs et prête à être utilisée par les ordinateurs du web sémantique, adaptée à l'univers des Linked Open Data.

Résultats

Nous proposons 182 Q-Codes dans une base de données multilingue en ligne (10 langues) (url www.hetop.eu/Q) formant autant de filtres pour Medline. Les Q-Codes sont aussi disponibles sous la forme identificateurs de ressource uniques (URI) et sont exportables en langage ontologique Web (OWL).

La Classification internationale des soins primaires (CISP) est jointe aux Q-Codes pour former une classification du contenu essentiel de la MG/MF [Core Content Classification in General Practice] (3CGP). A ce jour, 3CGP est utilisée par des collègues en pédagogie, dans des études bibliographiques, l'indexation de congrès ou de thèses de master et d'autres formes de littérature grise en GP / FM. L'utilisation par ordinateur est expérimentée dans les classificateurs automatiques, les annotateurs et l'apprentissage en ligne.

Discussion

À notre connaissance, il s'agit de la première tentative d'élargir le système de codification de la CISP en y ajoutant une extension pour les questions contextuelles de la pratique de la médecine de famille. Il reste à prouver que notre terminologie aidera à gérer des systèmes plus complexes, tels que MeSH, pour soutenir l'activité de stockage et de récupération d'informations. Cet exercice est proposé comme une première étape dans la création d'une ontologie en MG/MF et une ouverture au monde complexe des technologies du web sémantique.

Conclusion

Nous avons créé une ressource terminologique pour l'indexation de la connaissance spécifique en MG/MF. Nous espérons qu'elle permettra de réduire les pertes de connaissances dans le domaine GP / FM. De plus, grâce à une meilleure indexation de la littérature grise, nous espérons améliorer l'accessibilité des trvaux de recherche et donner de la visibilité au travail invisible des médecins de famille.

Acknowledgements

I want to express my gratitude;
To my wife Nadia Hichy for her permanent support since 32 years, with the last 4 the most challenging.
To the president and members of my committee of theses who were also co-investigators. Prof. Didier Giet, Prof. Marc Vanmeerbeek, Prof. Ashwin Ittoo, Prof. Robert Vander Stichele, Prof Laurent letrilliart, Prof. Gustavo Gusso, Prof. Kees Van Boven, Prof. Julien Grosjean and Prof. Stefan Darmoni, Mrs Melissa Resnick, Mr Arthur Treuherz and Mss Elena Cardillo.
To Mss Meaghan Blanchard, my always listening copy-editor.
To Mss LianTze Lim, Latex expert at overleaf.com, always ready to help.
To all people listed below, colleagues, counselors, translators, testers, users.
Thank you all.

- Zekeriya Aktürk MD, PhD. Family physician, Professor of family medicine. Department of Family Medicine. Faculty of Medicine University of Sifa, Izmir, Turkey.
- Philippe Ameline; Ir. Computer scientist and terminologist. Paris, France
- Jong-Myon Bae MD, PhD. Professor of Preventive Medicine. National University of Jeju. College of Medicine and Graduate School of Medicine. Jeju, South Korea
- Mr Marie Dominique Beaulieu, MD, FCMF, MSc. Professeur titulaire, Université de Montréal, Canada.
- Heinz Bhend, MD. Family physician, Computer Science, Director of the Institute for Practice Informatics (IPI). Aarburg, Switzerland
- Mss Meaghan Blanchard, MS, Linguist and copy editor
- Johan Brouns MD. PhD. Family doctor. General practitioner and computer scientist, Ghent, Belgium.
- Olivier Bodenreider MD. PhD. Senior Scientist. Cognitive Science . NLM/LHNCBC, USA
- Frank Buntinx, MD, PhD. Professor emeritus of general practice, University of Maastricht, The Netherlands & University of Leuven (KUL), Belgium
- Elena Cardillo, PhD. senior researcher, health information sciences, Instituto di Informatica e Telematica (IIT) del Consiglio Nazionale delle Ricerche, Italy
- Ayça Çetinbaş MD. Family physician, assist. Professor of Family Medicine. University of Trakya Faculty of Health Sciences Department of Nutrition and Dietetics, Edirne, Turkey.

- Nguyễn Thùy Châu MD, MSc. Family Physician, Lecturer in Family Medicine. Phạm Ngọc Thạch Medical University, Ho Chi Minh City, Vietnam
- Mr Paul Chevalier, MA, historian and latinist, Charleroi, Belgium
- Prof Stefan Darmoni, MD. PhD. Professsor of medical computer sciences, Département d'Information et d'Informatique Médicale, Université de Rouen, France
- Dominque Dekeuster, Consultant, Computer sciences, Belgium
- Anne-Françoise Donneau, Biostatistics unit, Faculté de Médecine, University of Liege
- Lynn Dunikowski, Head, Canadian Library of Family Medicine; Betty Taylor Library. Western University Libraries. Canada.
- Nguyễn Thị Bích Duyên MD, MSc, Family Physician, Professor of Family Medicine. Phạm Ngọc Thạch Medical University, Ho Chi Minh City, Vietnam.
- Bernard Gay MD. Professor emeritus of general practice, University of Bordeaux, France
- Enrique Gavilán. MD. Family and community doctor. Centro de Salud Montehermoso. Cáceres, España.
- Prof. Didier Giet MD. Department of General Practice/Family Medicine, University of Liege, Liege, Belgium
- Luis Filipe Gomes MD, PhD. Family Physician. Professor of general practice at the University of Algarve, Portugal
- Julien Grosjean PhD. Professor of Computer Sciences, Département d'Information et d'Informatique Médicale, Université de Rouen, France
- Gustavo Gusso MD, PhD. Family physician. Professor of general practice at the University of São Paulo, São Paulo, Brazil
- Prof. Hector Falcoff MD, PhD. Family physician. Professor of general practice at the University Paris Descartes and Society of Training and Therapeutics of the Generalist (SFTG), Paris, France
- Bernard Henx, MD. Family physician. Luxembourg city, Luxembourg.
- Ashwin Ittoo, PhD. Professor of Information Systems, HEC, University of Liege
- Nora Jamoulle BA, MA. Linguist. Brussels. Translator
- Sophie Jamoulle BA, MS. Brussels Communication specialist. Web site editor

- Tarik Jamoulle BA, MS, PhD applicant in Neurosciences. Brussels. Translator
- Ana Kareli, MD. Family physician. specialist in Medical Quality Service. Tbilisi. Georgia.
- Gaetan Kerdelhue, MA. terminologist, Département d'Information et d'Informatique Médicale, Université de Rouen, France
- Frederic Ketterer PhD, sociologist. Department of Family Medicine. University of Liege, Belgium
- Daniel Knupp, MD. Family physician. Professor of family practice, Belo Horizonte, Brazil
- Ricardo La Valle MD. PhD. Family physician. Unit of Family Medicine, Hospital Italiano, Buenos Ayres, Argentina
- Mr Olivier Latignies, Computer scientist, Ontologist, Belgium
- Prof Laurent Letrilliart, MD, PhD. Family physician. Professor of General Practice, University of Lyon Est, France
- Mrs LianTze Lim, PhD. Overleaf LaTeX Community TeXpert at Overleaf.com
- Prof. Anne Marie Magnier, MD. PhD. Family physician. Professor of general practice, Faculté Pierre et Marie Curie, Paris, France.
- María Ana Mariño MD. Family physician. Sociedad Argentina de Medicina Interna General (SAMIG). Buenos Aires, Argentina.
- Susana Medeiros.MD. Family doctor. Head of general practice clinic at the University of Algarve, Portugal.
- Natia Natroshvili MD. pediatrician, Tbilisi Referral Hospital. Georgia.
- Armando Henrique Norman MD. PhD. Family physician and anthropologist. Brazil
- Patrick Ouvrard MD. Family physician and anthropologist. Society of Training and Therapeutics of the Generalist (SFTG), Paris, France.
- Miguel Pizzanelli MD. MS. Family physician, Associate Professor, Department of Family and Community Medicine. Faculty of Medicine UDELAR. Montevideo, Uruguay.
- Dr Ricardo Rebolho, MD, GP. Florianopolis, Brazil
- Mrs Melissa Resnick, MLS. Medical librarian. Ph.D. Candidate, University of Texas, Health Science Center at Houston, USA.

- Sebastien Rigaux, Master in Computer Sciences, Health Information Sciences, HEC, University of Liege .
- Michel Roland, MD, PhD. Family physician. Professor emeritus of general practice. Department of General Practice, University of Brussels, Belgium.
- Cyril Romain, Computer scientist, Ontologist, Belgium
- Joseph Roumier, Igr, Computer scientist, Ontologist, France
- Serdal Sanak. MS. Student in medicine. University of Brussels (ULB), Brussels, Belgium.
- Luis Miguel Santiago MD, PhD. Family physician, Professor of family medicine. General Practice Clinic, Faculty of Medicine, University of Coimbra, Portugal
- Matthieu Schuers, MD. Family physician. Département d'Information et d'Informatique Médicale, Université de Rouen, France.
- Christoffe Sepulchre, MD. Family physician. Professor of general practice, Lille University, France.
- Frank S. Shamenek, M.L.S., J.D. Consultant, New York, NY, USA.
- Christian Simon. MS. Computer Scientist. Angers, france
- Carl Steylaerts MD. Family physician. Past honorary treasurer, WONCA Europe, Maldegem, Belgium
- Ariana Tavares, student in medicine, General Practice Clinic, Faculty of Medicine, University of Coimbra, Portugal
- Benoit Thirion, senior terminologist, Département d'Information et d'Informatique Médicale, Université de Rouen, France.
- Arthur Treuherz. BA, MS. Supervisor of terminology and health classifications. Latin American and Caribbean Center for Health Sciences (BIREME / OPAS / WHO), São Paulo, Brasil.
- Kees van Boven, MD, PhD. Family physician. Professor of General Practice, Department of general practice, University of Nijmegen, The Netherlands.
- Robert Vander Stichele. MD, PhD. Family physician. Professor of Pharmacology, Heymans institute, University of Ghent, Belgium.
- Marc Vanmeerbeek. MD, PhD. Family physician. Professor of general practice, Department of General Practice, University of Liege, Liege, Belgium

- Raquel Vaz Cardoso MD, MPH. Family physician. Secretariat of the State of Health of the Federal District, Brasília, Brazil
- Trần Thị Hoa Vi MD, MSc. Family physician, Lecturer in Family Medicine. Phạm Ngọc Thạch Medical University, Ho Chi Minh City, Vietnam.
- Võ Thành Liêm MD, PhD. Family physician. Lecturer in Family Medicine. Phạm Ngọc Thạch Medical University, Ho Chi Minh City, Vietnam.
- Hamilton Wagner, MD. Family physician. Professor of family practice, Curutiba, Brazil
- Maxime Warnier, MS. Linguist, Computer scientist, Toulouse University, France
- Johan Wens MD, PhD. Family physician. Associate Professor, University of Antwerp, Antwerp, Belgium
- Ayoub Zayane, MD. Family physician in vocational training. Brussels, Belgium

Contents

List of Figures

List of Abbreviations

3CGP	Core Content Classification of General Practice / Family Medicine
ATC	Anatomical Therapeutic Classification
BIREME	Biblioteca Regional de Medicina (Sao Paulo)
BTNT	Broader than Narrower Term
CAQDAS	Computer-Assisted Qualitative Data Analysis Software
CISP	Classification Internationale des Soins Primaires [International Classification of Primary Care]
CISP-Club	French speaking association of ICPC users
CISMeF	Catalog and Index of French Language Health Resources on the Internet
CNGE	College National des Généralistes Enseignants [French association of teachers in family medicine]
CSV	Comma-Separated Values
DBGUI	Database graphic user interface
DeCS	Health Sciences Descriptors
DRC	Dictionnaire des Résultats de Consultation [Dictionary of encounters output]
ECMT v3	Extracteur de Concepts Multi-Terminologique version 3 [Multiterminologic annotator version 3]
EGPRN	European General Practice Research Network
EMR	Electronic medical record
ESAIP	Ecole d'ingénieur [Ingeneering school] , Anger, France
FAMLI	Family Medicine Literature Index
FIPA	Foundation for Intelligent Physical Agents
GP/FM	General Practice / Family medicine
GP	General practitioner
GP-LIT	Thesaurus of the Royal College of General Practitioners
HEC-Ulg	École de gestion de l'université de Liège [Liege university management school]
HeTOP	Health Terminology/Ontology Portal
HIS	Health information System
ICD-10	International Classification of Disease, tenth revision
ICHPPC	International Classification of Health Problems in Primary Care
ICNP	International Classification for Nursing Practice

ICPC-2	International Classification of Primary Care, second version
ICPC-Process-PC	International classification of procedures in primary care
ICPC-2 e ver. 6	International Classification of Primary Care second edition, electronic, version 6
ID	Identifier
IFP/GP SIG	International Family Practice-General Practice SNOMED Special Interest Group
ISO 639	Languages standards of the International Organization for Standardization
KOS	Knowledge Organization Systems
LILACS	Latin American and Caribbean Health Sciences Literature
LOV	Linked Open Vocabularies
MEDLINE	Medical Literature Analysis and Retrieval System Online.
MedDRA	Medical Dictionary for Regulatory Activities
MeSH	Medical Subject Headings of the National Library of Medicine.
NERC	Named Entity Recognition and Classification
NCCMERP	National Coordinating Council for Medication Error Reporting and Prevention
NCIt	National Cancer Institute terminology
NLM	National Library of Medicine (US)
NLP	Natural language processing
NTBT	Narrower Than Broader Term
OWL	Web Ontology Language
P4-SIG & O	Quaternary Prevention & Overmedicalisation WONCA Special Interest Group
PAHO	Pan-American Health Organization
PF	Preferred term
PHC	Primary Health Care
PubMed	Search interface of Medline
Q-Codes	Contextual classification in GP/FM, included in 3CGP
RCC	Read Clinical Code
RDF	Resource Description Framework
SBMFC	Brazilian Society of Family and Community Medicine
SciELO	Scientific Electronic Library Online
SDM	Shared Decision Making
SFTG	Société de Formation et de Thérapeutique Généraliste. [French society of teaching and therapy in family medicine]
SGD	Stochastic Gradient Descent

TF-IDF	Term Frequency – Inverse Document Frequency
SKOS	Simple Knowledge Organization System
SML	Supervised Machine Learning
SNOMED-CT	Systematized Nomenclature of Medicine – Clinical Term
SNOP	Systematized Nomenclature of Pathology
SVM	Support Vector Machine
TF-IDF	Term Frequency – Inverse Document Frequency
TFE	Travail de Fin d'Etude [final work]
TW	Text word
UMLS	Unified Medical Language System
URI	Uniform Resource Identifier
URL	Uniform Resource locator
VHL	Virtual Health Library
WHO	World Health Organization
WHO-ART	World Health Organization Adverse Drug Reaction Terminology
WICC	Wonca International Classification Committee
WONCA	World association of family doctors

Published papers & books

In the following list, personal papers published in international journal, books, posters, conference proceedings and databases edited for this thesis are preceded by an asterisk between brackets (*). They form the base of this work. Those bearing two asterisks (* *) are reprinted after the general introduction. On-line edited texts and material are used for other chapters and Appendix. Some publications and books are anterior to the beginning of the thesis but form its background and are quoted also in this list as my personal contribution to the science of taxonomy in Family Medicine.

Papers

(* *) Jamoulle, M., Resnick, M., Ittoo, A., Vander Stichele, R., Cardillo, E., Grosjean, J., Darmoni, S., & Vanmeerbeek, M. (2017). Indexing grey multilingual literature in General Practice in the era of Semantic Web. The Grey Journal.(accepted)
http://hdl.handle.net/2268/210490

(* *) Jamoulle, M., Resnick, M., Grosjean, J., Ittoo, A., Cardillo, E., Vander Stichele, R., Darmoni, S., Vanmeerbeek, M. (2017). Development , dissemination , and applications of a new terminological resource the Q-Codes taxonomy for professional aspects of General Practice / Family Medicine. EJGP, (accepted), 1–13.

(* *) Jamoulle, M., Resnick, M., Ittoo, A., Cardillo, E., Vander Stichele, R., & Vanmeerbeek, M. (2017, September). Analysis of definitions of General Practice/Family Medicine and Primary Health Care. British Journal of General Practice Open, 050.
http://hdl.handle.net/2268/210049

(* *) Jamoulle, M., Grosjean, J., Ittoo, A., Resnick, M., Treuherz, A., Vander Stichele, R. H., Cardillo, E., Shamenek, F., & Vanmeerbeek, M. (2017). A terminology in General Practice / Family Medicine to represent non-clinical aspects for various usages. In R. Randell et al. (Ed.), Informatics for Health: Connected Citizen-Led Wellness and Population Health (pp. 235:471-475). IOS Press.
http://hdl.handle.net/2268/206527

(*) Jamoulle, M. (2016, April 14). New 2016 MeSH addressing information gap, poverty, violence and danger of medicine set the tone for policy makers in patient care. International Journal of Health Policy and Management.
http://hdl.handle.net/2268/195481

(* *) Jamoulle, M., Vander Stichele, R. H., Cardillo, E., Roumier, J., Grosjean, J., & Darmoni, S. (2015). Semantic Web and the Future of Health Care Data in Family Practice. Merit Research Journal of Medicine and Medical Sciences, 3(12), 586-594.
http://hdl.handle.net/2268/189292

(*) Jamoulle, M., Gavilan, E., Cardoso, R. V., Mariño, M. A., & Pizzanelli, M. (2015). The words of prevention, part I: changing the model. Revista Brasileira de Medicina de Família e Comunidade, 10(35), 1-9.
http://hdl.handle.net/2268/176753

(*) Jamoulle M, Gavilán E, Cardoso RV, María Ana Mariño MA, Pizzanelli M, Grosjean J, & Darmoni SJ.(2015) The words of prevention, part II: Ten terms in the realm of quaternary prevention. Revista Brasileira de Medicina de Família e Comunidade. 10(35):1-1.
doi:10.5712/rbmfc10(35)1063.

(* *) Jamoulle, M., Cardillo, E., Roumier, J., Warnier, M., & Vander Stichele, R. (2014). Mapping French terms in a Belgian guideline on heart failure to international classifications and nomenclatures: the devil is in the detail. Informatics in Primary Care, 4, 189-198.
http://hdl.handle.net/2268/171599

(*) Cardillo, E., Warnier, M., Roumier, J., Jamoulle, M., & Vander Stichele, R. (2013). Using ISO and Semantic Web standards for creating a Multilingual Medical Interface Terminology : A use case for Hearth Failure. Terminology and Artificial Intelligence TIA 2013.
http://hdl.handle.net/2268/171534

(* *) Jamoulle, M. (2013). Using the International Classification for Primary Care (ICPC) and the Core Content Classification for General Practice (3CGP) to classify conference abstracts. Letter. The Portuguese Journal of General Practice (RPCG), 29(5),66-67.
http://hdl.handle.net/2268/171601

(* *) Warnier, M., Roumier, J., Jamoulle, M., Cardillo, E., Vander Stichele, R., & Romary, L. (2012). Publishing a Multilingual Medical Terminology According to Terminology Standards and Linked Data Principles.

SABRE Conference 2012.
http://hdl.handle.net/2268/171545

Roumier, J., Jamoulle, M., Vander Stichele, R., Romary, L., & Cardillo, E. (2011). Towards a terminologies support system in Primary Care (Letter to the editor). Informatics in Primary Care, 19, 257-258.
http://hdl.handle.net/2268/171544

Rosendal, M, L Letrilliart, K van Boven, A Grimsmo, H Britt, J K Soler, M Verbeke, M Jamoulle, R Simkus, and M Klinkman. (2010). "Comment on the article : Sixteen Years of ICPC Use in Noewegiam Primary Care by Botsis et Al." BMC Medical Informatics and Decision Making, no. 10(June):11.
https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/1472-6947-10-11/comments .

De Jonghe, M., Jamoulle, M., & Porignon, D. (2006). Utilisation et évaluation du thesaurus belge adapté au Rwanda. Revue Médicale de Bruxelles, 27(4), 274-8.
http://hdl.handle.net/2268/203450

Okkes, I., Jamoulle, M., Lamberts, H., & Bentzen, N. (2000). ICPC-2-E: the electronic version of ICPC-2. Differences from the printed version and the consequences. Family practice, 17(2), 101-7.
http://hdl.handle.net/2268/171600

Jamoulle M, editor La CISP, une classification en soins de première ligne et une méthode de traitement de l'information en médecine ambulatoire. Classifications des patients des secteurs stationnaire et ambulatoire', PCS Suisse, Société Suisse pour la Politique la Santé,; 2000/11; Berne.
http://www.ulb.ac.be/esp/cisp/berne2000.htm

Books

(*) Jamoulle, M., Brouns, J., Steylaerts, C., & Wens, J. (2016). Een meertalige terminologie in de huisartsgeneeskunde - Nederlandse versie. Jumet, Belgique: Care Editions.
http://hdl.handle.net/2268/202212

(*) Jamoulle, M., Marino, M. A., La Valle, R., & Pizzanelli, M. (2016). Terminologia multilingue en medicina general y de familia - Version en Español. Jumet, Belgique: Care Editions.
http://hdl.handle.net/2268/202211

(*) Jamoulle, M., Ouvrard, P., Schuers, M. (2016). Terminologie multilingue de médecine générale et de famille. Version française. Jumet, Belgique, Care Editions.
http://hdl.handle.net/2268/202210

(*) Jamoulle, M., & Resnick, M. (2016). General Practice / Family Medicine Multilingual Terminology. English version. Charleroi, Belgique: Care Editions.
http://hdl.handle.net/2268/202209

(*) Jamoulle, M., Treuherz, A., Gomes, L. F., Medeiros, S., & Gusso, G. (2016). Terminologia Multilingue de Clinica Geral / Medicina de Familia e Comunidade - Versao em Portugues. Jumet, Belgique: Care Editions.
http://hdl.handle.net/2268/202213

(*) Jamoulle, Marc, Nguyễn Thùy Châu, Trần Thị Hoa Vi, Nguyễn Thị Bích Duyên, và Võ Thành Liêm. (2016). Từ Điển Thuật Ngữ Đa Ngôn Ngữ về Y Học Tổng Quát và Y Học Gia Đình. Rio de Janeiro. Nov 2, 2016.
http://orbi.ulg.ac.be/handle/2268/202214.

Jamoulle, M. (2005). Contributor to, the International Classification of Family Medicine, Revised 2nd revision (ICPC-2-R) edited by the Wonca International Classification Committee. Oxford Medical Publications.

Jamoulle, M. (2003). Contributor to the Wonca Dictionary of General / Family Practice. In N., Bentzen (Ed.). Copenhagen, Dk: Maanedsskr.
http://hdl.handle.net/2268/171567

Jamoulle M, Roland M, Humbert J, et Brûlet, JF.(2000) Traitement de l'infor- -mation médicale par la Classification Internationale des Soins Primaires 2ème version (CISP-2), assorti d'un glossaire de médecine générale, préparé par le Comité International de Classification de la WONCA. Bruxelles Care Editions asbl, Bruxelles.

Roland, M., Jamoulle, M., & Dendeau, B. (1996). Approches taxinomiques en médecine de famille. Bruxelles, Belgique: Care edition asbl. 2 vol.
http://hdl.handle.net/2268/195844

Jamoulle, Marc, Michel Roland, François Mennerat, and Jacques Humbert. (1999). "Origine de La CISP et Mise En Application Actuelle Dans Les Pays Francophones." In L'informatisation Du Cabinet Médical Du Futur, edited by A Venot and Hector Falcoff, 201–12. Paris: Informatique et Santé. Springer-Verlag.
http://hdl.handle.net/2268/175668.

Jamoulle, M. (1998). ICPC-2. The International Classification of Primary Care, an Introduction.
http://orbi.ulg.ac.be/handle/2268/195813.

Jamoulle M, Roland M.(1992) Classification Internationale des Soins Primaires (CISP). Lacassagne.

Part of books

(*) Gusso, G., & Jamoulle, M. (2016). From Dictionary to Terminologies / Ontologies in General Practice/Family Medicine. The World Book of Family Medicine – Iberoamericana Edition 2016 (pp. 60). Iberoamericana Edition.
http://hdl.handle.net/2268/202788

(*) Soler, J. K., Jamoulle, M., & Schattner, P. (2015). The International Classification of Primary Care. In Stichting WONCA Europe, The WORLD book of family medicine (pp. 127-8). Ljubljana: European edition.
http://hdl.handle.net/2268/187050

Jamoulle, Marc. (1993) "Practical Experience with ICPC in Belgium." In The International Classification of Primary Care in the European Community., edited by Henk Lamberts, Maurice Wood, and Inge Hofmans-Okkes, Oxford University press, 101–5.

Posters & Conference proceedings

(*) Knupp, D. A., Rebolho, R., Jamoulle, M., Poli Neto, P., & Gusso, G. (2017). Coding of papers submitted to the 14th Congresso Brasileiro de Medicina de Família e Comunidade (abstract of oral communication).
http://orbi.ulg.ac.be/handle/2268/214643

(*) Jamoulle, M. (2017). Online ICPC-2 & Q-Codes: General Practice / Family Medicine On-line Multilingual Terminology & Knowledge Base. In WONCA International Classification Committee Open day (p. 18 slides). Lyon, France.
http://orbi.ulg.ac.be/handle/2268/214115

(*) Jamoulle, M., Treuherz, A., Gomes, L. F., Medeiros, S., & Gusso, G. (2016, November 02). Terminologia Multilingue de Clinica General / Medicina de Familia e Comunidade. Poster session presented at Wonca world Congress, Rio de Janeiro, Brazil.
http://hdl.handle.net/2268/203016

(*) Jamoulle, M., Kuehlein, T., Pizzanelli, M., Grosjean, J., Kerderhué, G., & Darmoni, S. (2015, October 23). International Classification of Primary Care in a cross-lingual terminology portal (Poster). Poster session presented at WHO- Family of International Classifications Network Annual Meeting 2015, Manchester, United Kingdom.
http://hdl.handle.net/2268/185131

(* *) Jamoulle, M., & Steylaert, C. (2015, October 22). Concept identification and context analysis of the Wonca Europe 2015 book by instance-based categorization of general practice activities. Poster session presented at Wonca Europe congress 2015, Istanbul, Turquie.
http://hdl.handle.net/2268/185776

(*) Jamoulle, M., Pizzanelli, M., Grosjean, J., Kerdelhue, G., & Darmoni, S. (2015, March 20). International Classification of Primary Care in a cross-lingual terminology portal. Poster session presented at 4° Congreso Iberoamericano de Medicina Familiar y Comunitaria Montevideo March 19-21 2015. Montevideo.
http://hdl.handle.net/2268/179499

(*) Schuers, M., Jamoulle, M., Kerdelhue, G., Grosjean, J., & Darmoni, S. (2015, March). Mise en ligne de la CISP en près de 20 langues au sein d'un portail terminologique de santé. Poster session presented at Congrès de la Médecine Générale, Paris, Paris, France.
http://hdl.handle.net/2268/207421

(*) Jamoulle, Marc. (2015 - by teleconference). Une étude bibliographique avec 3CGP [Bibliographic study with the Q-codes]. Cisp Club 2015, Rouen.
http://orbi.ulg.ac.be/handle/2268/187393

(*) Jamoulle, M., Grosjean, J., & Darmoni, S. (2015, November 13). Structure of the Q-Codes semantic knowledge base in General Practice/Family Medicine. Poster session presented at WICC annual meeting Hyderabad, India, Hyderabad, India.
http://hdl.handle.net/2268/187498

Jamoulle, M., & Dekeuster, D. (2007, October 19). Workshop : Towards an e-archive for Wonca documents. Paper presented at Wonca Europe congress 2007, Paris, France.
http://hdl.handle.net/2268/187135

Reports & working papers

Jamoulle, M. (2010). Terminologies, informatique et soins primaires. Belgique 1980-2010. Bruxelles, Belgique: Ministere belge de la Santé

Publique.
http://hdl.handle.net/2268/171570

Jamoulle, M., & Romain, C. (2010). Analyse des systèmes de gestion de données de médecine générale les plus usités en Wallonie et à Bruxelles, Belgique, 2009/2010. Bruxelles, Belgique: Ministere de la Santé Publique SPF-SP.
http://hdl.handle.net/2268/196340

Jamoulle, M. (2010). Some views about SNOMED-CT by a General Practitioner. Eprint/Working paper
http://docpatient.net/onto/doc/SNOMED_CT_study_MJ_2010.pdf
.http://hdl.handle.net/2268/173908

Databases & support

(*) Jamoulle, Marc, Ana Kareli, & Natia Natroshvili. (2017). Q-Codes, version 2.5., tabular list, Georgian version, copy desk. 2017.
http://hdl.handle.net/2268/215161 .

(* *) Jamoulle, M., Grosjean, J., & Darmoni, S. (2017). Access to multilingual individual rubrics in URI format for ICPC-2 and the Q-Codes.
http://hdl.handle.net/2268/211268

(*) Jamoulle, M. (2017). Guide d'indexation des travaux de fin d'étude en médecine générale.
http://hdl.handle.net/2268/208968

(*) Jamoulle M. (2017). 3CGP user guide. Towards Knowledge management in General Practice & Family Medicine; Guide to the indexing of master's thesis.Department of General Practice, University of Liège. 12p.
http://hdl.handle.net/2268/209379

(*) Jamoulle, M., Grosjean, J., & Darmoni, S. (2016). Online multilingual (en, fr, es, pt, nl, kr, vi) terminological database in General Practice/Family Medicine ; the Q-Codes .
http://hdl.handle.net/2268/202199

(* *) Jamoulle, M. (2016). Q-Code, version 2.5, multilingual tabular list (en, fr, es, pt, nl, ko, vi) copy desk.
http://hdl.handle.net/2268/202216

(*) Comité de Bibliografía en Prevención Cuaternaria, WONCA SIG P4 & O, Miguel Pizanelli, Ricardo La Valle, and Marc Jamoulle.(2017) Quaternary Prevention Library and Resources in Four Languages (Qp Library). https://tinyurl.com/P4-bib-resources .
http://hdl.handle.net/2268/212975

(*) Jamoulle, M., Gusso, G., Marino, M. A., Pizzanelli, M., Steylaert, C., Resnick, M., Bae, J.-M., Vo, T. L. (2016). A semantic web based resource for GP/FM in English, Portuguese, French, Spanish, Dutch, Vietnamese and Korean .
http://hdl.handle.net/2268/202439

Jamoulle, M., & Roland, M. (2000). CISP-2 Exemplaire de bureau. Wonca. http://hdl.handle.net/2268/196337

To my wife Nadia who is certainly co-authoring this work, at least in support, management, emotion, despair, hope, caring and love.

In Memoriam

- Bent Gunthorm Bentzen, Professor of general practice, Oddense, Norway.
- Henk Lamberts, Professor of general practice, Amsterdam, The Netherlands.
- Maurice Wood, Professor of family practice, Richmond, Virginia, USA
- Charles Bridges-Webb, Professor of general practice, Sydney, Australia.

Preamble

The approach of the thesis is threefold: expertise in a particular field of knowledge (family medicine), computer science (databases, semantic web), and management of language (terminology, linguistics, documentation science). A narrow collaboration between those three domains has been unavoidable in reaching my objectives.

My steps in learning the other two disciplines have been hesitant, but I have been supported by my fellow computer scientists, linguists, terminologists, and librarians in an extraordinary way.

I would especially like to thank Professors Marc Vanmeerbeek, Robert Vander Stichele, Ashwin Itoo, Elena Cardillo, Julien Grosjean and Stefan Darmoni as well as Mrs. Melissa Resnick and Mr. Arthur Treuherz, who have accompanied me step by step in this research.

In the field of family medicine, interest and support came from all over the world. When I started this work, it was unimaginable to think the results would one day be available in 10 languages, ready for the Semantic Web.

This gradual evolution is reflected in the progress of the research and the eventual writing of this document.

In the introductory chapter, I will attempt to describe the project as if it had been performed originally as a classic study, with a literature review, hypotheses, methods and results.

However, a different methodology was used in reality. The interactions have been strong between new knowledge, methods, and results. Gradually, I was able to see the three axes of this research: producing a specific indexing system, contributing to General Practice / Family Medicine (GP/FM) grey literature visibility, and contributing to the writing of the table of the subjects of family medicine.

The chapters following the introduction are a compilation of successive publications that illustrate the subject and show the extent of this research.

In the General Introduction, the reader will be invited to look deeper into the issues addressed, and then they will be asked to refer to chapters that describe the various issues that went into this research.

The profession of family doctor connects technological medicine to humans and protects them from the errors of science through expertise in the domain of information and communication.

By contributing to the management of information, which can be considered the heart of this activity, I hope to contribute to the development of my profession and the protection of my patients.

Préambule

Le présent travail est le fruit d'un long cheminement. La découverte de l'existence des nouvelles techniques de la gestion de connaissance a profondément influencé l'organisation de cette thèse.

La démarche est triple. L'expertise sur un champ de savoir particulier, la médecine de famille, confrontée aux techniques informatiques (bases de données, web sémantique) et à la gestion de la langue (terminologie, linguistique, documentologie) ont imposé une collaboration entre les trois disciplines.

Mes pas dans l'apprentissage des deux autres disciplines ont été hésitants mais j'ai été soutenu par mes collègues co-auteurs informaticiens, linguistes, terminologistes documentologistes.

Je veux ici remercier les Professeurs Marc Vanmeerbeek, Robert Vander Stichele, Ashwin Itoo, Stefan Darmoni et Julien Grosjean ainsi que Elena Cardillo, Melissa Resnick et Arthur Treuherz, qui m'ont accompagné pas à pas dans ma recherche.

Dans le domaine de la médecine de famille, c'est du monde entier que sont venus l'intérêt et le support. Il était inimaginable quand j'ai commencé ce travail que son résultat, les Q-Codes, soit un jour disponible en 10 langues et pret pour le web sémantique.

Cette évolution progressive se reflète dans le déroulé de la recherche et l'écriture du document. Dans le chapitre introductif je tenterai de refaire le projet comme s'il avait été pensé au départ de façon classique, avec une revue de la connaissance, des hypothèses, un méthode, des résultats.

La réalité n'a pas été celle-là pourtant. Entre nouvelles connaissance, méthodes et résultats les interactions ont été fortes et c'est petit à petit que ce sont profilés les trois axes de la recherche : produire un système d'indexation spécifique, contribuer à rendre visible la littérature grise de médecine de famille, contribuer à l'écriture de la table des matières de la médecine de famille.

Les chapitres qui suivent l'introduction sont une compilation des publications successives qui permettent d'illustrer le propos et de voir l'extension de la recherche. Dans l'introduction générale, le lecteur sera invité à approfondir les questions traitées en se référant aux chapitres indiqués dans le texte.

Le métier de médecin de famille relie la médecine technologique aux humains et les protège des errements de la science grâce à l'expertise du domaine, l'information et la communication. En contribuant à la gestion de l'information, cœur de l'activité, j'espère avoir contribué au développement de mon métier et à la protection de mes patients

Chapter 1

Indexing grey multilingual literature in General Practice in the era of Semantic Web

1.1 BACKGROUND

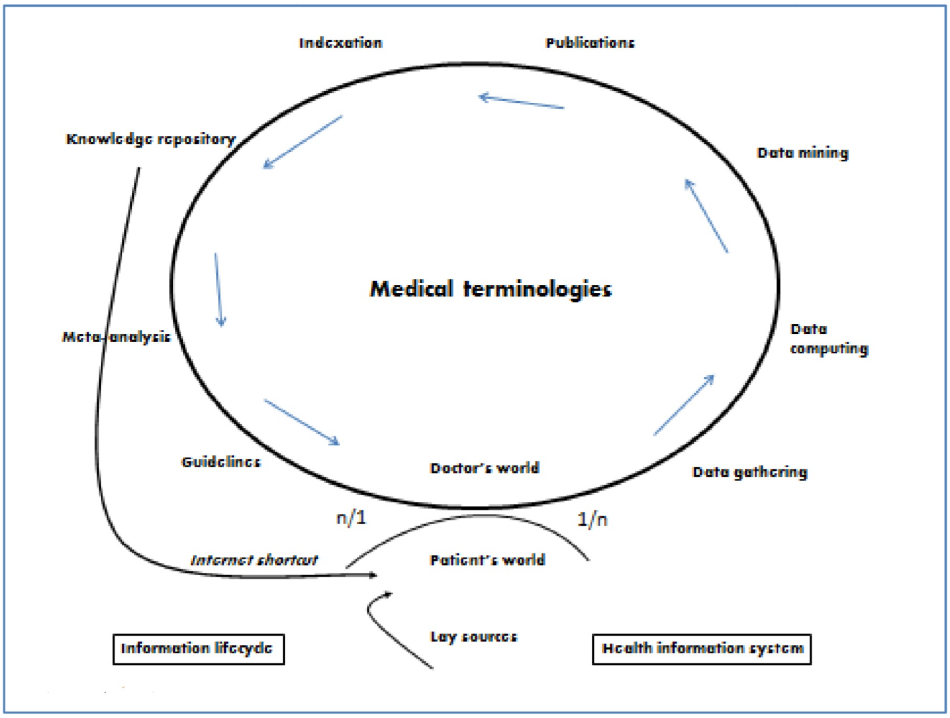
1.1.1 Need for information in family medicine

In the cycle of patient centered information (Jamoulle et al., 2015b), the General Practitioner (GP) is simultaneously a heavy user and producer of published/unpublished data (see Fig. 1.1). Data produced could be clinical, (i.e. dealing with symptoms, processes and diseases). Or data produced could be contextual, addressing some particular issue of the patient, which may influence the process of care (Schrans et al., 2016). However, contextual data can also deal with issues concerning the doctor, the managerial aspects of care. In particular, it could address the position of GPs within the health care system, the general concepts used in Primary Health Care (PHC), or the delivery services. In this work, focus will remain on these last contextual medical features of General Practice / Family Medicine (GP/FM), as its tools for training, research, ethics, inquiry, environmental issues, infrastructure and principle of care. These features are central to this field, and family doctors are used to exchange information over them when they meet in training sessions or during congresses.

The last two decades have seen the rise and maturation of Electronic Medical Records (EMR) and Health Information Systems (HIS) (Tharmalingam, Hagens, and Zelmer, 2016). Research in terminological fields is ongoing. Over time, locally produced vocabularies have been absorbed and replaced by huge terminologies. And now, data mining into gathered information is becoming more and more prevalent (Roumier et al., 2011; Roski, Bo-Linn, and Andrews, 2014; Troeung et al., 2016).

The realm of GP/FM differs from mainstream health care, as Family Physicians (FPs) address biological, technological, behavioral, sociological and anthropological domains. All of which have a deep impact on the terminologies needed (Helman, 2008; Thompson et al., 2014). As the creation of already available terminologies were focused on specialized domains,

FIGURE 1.1: Patient-Doctor relation is a consumer and producer of clinical and contextual information in GP/FM)(Jamoulle et al., 2015b)



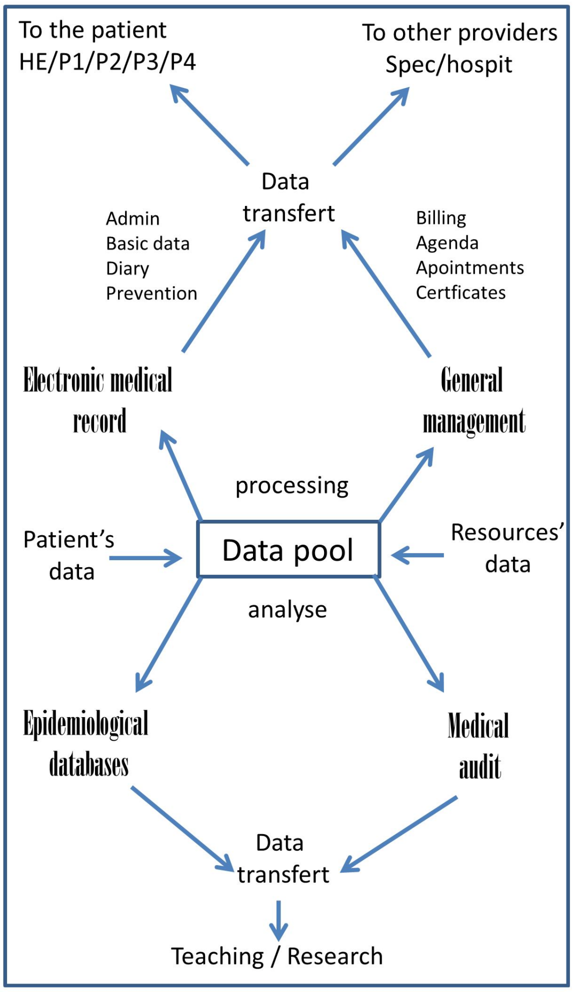
the biological and technological field of medical terminologies are now almost complete (Jonquet et al., 2016; Lelong et al., 2016). However, they sometimes fall short when applied to the field of GP/FM, which relies intensely on complexity and timeline issues (Liang et al., 2014; Madkour, Benhaddou, and Tao, 2016). Moreover, clinical issues are well documented. Professional contextual issues, like management, teaching, research, and ethics are very poorly documented for the first level of care (Jamoulle et al., 2017a).

Despite the speed of information flow and dramatic changes in resources since the Internet revolution (Sommerhalder et al., 2009), the general frame of information circulation in the core business of GPs - developed 30 years ago- is still useful (Jamoulle, 1986).

The Fig.1.2 shows patient and resources data alimenting a data pool. This pool of data could be processed [processing] or analyzed [analyse]. Processing data gives birth to the medical record, for now electronic, and to the general management of the activity of the primary care unit. That data is transferred to the patient (Jilka et al., 2015) and to the other providers through EMRs on Internet (Rosemann et al., 2006). Analysis of information produced in GP/FM field could potentially give birth to an epidemiological database, provided the information is structured and classified correctly

(Schmidt et al., 2015). Data could also be used for auditing activity and conducting research (Vellinga et al., 2016). Dissemination of information for teaching and research is the natural output of such an analytic process (vanBoven et al., 2011).

FIGURE 1.2: Data flow in GP/FM (Jamoulle, 1986)
HE: Health education. P1 to P4: primary, secondary, tertiary, quaternary prevention programs



1.1.1.1 GP/FM, a profession without clear limits

Despite elaborate definitions of GP/FM Allen et al. (2011) and Primary Care Physicians (PCPs) (AAFP, 2011), the manner in which the profession of GP/FM or PCP is defined and structured varies greatly across GP/FM textbooks (Casado Vicente, 2012; Gusso and Lopes, 2012; Kochen, 2012; Murtagh, 2011; Druais et al., 2009; David et al., 2013; Lakhani, 2003; McWhinney, 1997). This rings especially true in regard to managerial and contextual features. These textbooks have offered a top-down expert view of the profession as the authors themselves chose the subjects addressed. In this research, we rely on what doctor are interested in. In this sense, one can speak of a bottom-up approach.

If one examines the table of contents of these works, one sees an initial major division. The different body systems are reviewed, usually with a focus on the relevant specialties (i.e. rheumatology or endocrinology). Titles such as "Cardiology for General Practice" or "Dermatology for General Practice" are common. Remarkably, the book of Gusso and Lopes (2012) classifies the chapters according to the grid of ICPC. As far as general management and contextual background are concerned, the quoted books are absolutely different. One focuses on communication, the other on the systemic approach, and the third on the ethics of relationships. None gives a similar view of the scope and contextual scope of family medicine. Technology is often absent. Only one author or another approaches current technical processes in family medicine.

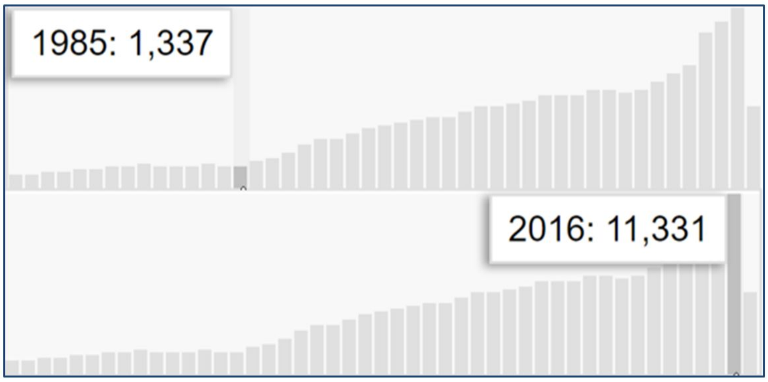
An extensive review of the vocational training programs in the specialty of General Practice was not done. The author, however, knows from experience and the many contacts he has in family medicine in Europe/the world that these programs have no homogeneity, despite the recommendations of EURACT, the WONCA Europe working group on education (Heyrman, 2005). On the other hand, when considering Continuous Medical Education, the drug industry's influence on the choice of subjects is decisive, which creates multiple conflicts of interests (Davis, 2004). Therefore, it seemed wise to develop an index of concepts dealing with family medicine by listening to practicing GPs, and to develop a bottom-up approach rid of conflicts of interest.

1.1.1.2 Published and unpublished in GP/FM, what's the meaning?

Medical Subject Headings (MeSH) is an ontology used for the indexing of PubMed and other databases. MeSH are normalized terms directed to enter queries in Medline, the bibliographic data base of the National Library of Medicine (NLM) through the interface PubMed (Lowe and Barnett, 1994). Currently, the PubMed interface gives access to 27.3 million citations. A search with 7 GP/FM relevant MeSH in June 2017 brings a little less than 200,000 citations. The same interrogation with 8 specific MeSH descriptors of Primary Health Care (PHC) gives 480,000 citations (PHC and GP/FM related MeSH are listed in Fig. 1.50, page 79). On August 17, 2017, the 8 PHC related MeSH and their entry terms gather 468,921 citations (1,17% of Medline content) while the 7 related GP/FM related MeSH and their entry terms gather 155,007 citations (0,56%). Together the 15 specific descriptors of the first line of care don't gather even 2% of Medline content.

However, the rise in publications has been inevitable since 1985 (see Fig. 1.3). The quoted publications considered relevant to GP/FM in this figure are obtained through a search strategy comprised of seven carefully chosen MeSH and their text words (see Appendix E, page 209 and Fig. 1.50, page 79). An argument against this is that this could be certainly appraised as *published* data and claimed white literature. This raises question about Pisa definition of grey literature, considered as *not controlled by commercial publishing* (GreyNet, 2014) as one can hardly pretend that papers published by the National Library of Medicine are always commercially controlled.

FIGURE 1.3: Evolution of publications in GP/FM (7 MeSH) from 1985 to 2016 on PubMed



Unpublished data in GP/FM are numerous. Yet, GP/FM organizations are heavy producers of continuous medical education (VanNieuwenborg et al., 2016), training sessions, multiple congresses at local, regional and national level in every country each year, research meetings (Buono et al., 2013), virtual conferences (Cavadas, Villanueva, and Gervas, 2010), websites, blogs. They are also active on social networking (Veuillotte et al., 2015). For local and national events, the local language is the rule.

It is not a secret that medical knowledge translation between doctor and patient is highly controlled by pharmaceutical companies (Moynihan, 2003) (Moynihan and Bero, 2017), with heavy impact on GPs behavior (Fickweiler, Fickweiler, and Urbach, 2017). Despite the activities of GP organizations, in some countries, domestic papers as medical newspapers remain the main source of information for practicing doctors (Tabatabaei-Malazy, Nedjat, and Majdzadeh, 2012). Usually, they are within an atmosphere of heavy, silent corruption (Angell, 2017). The translation of information by drug representatives is also a determining factor (Greenway and Ross, 2017) as well as predatory open access publications (Shen and Björk, 2015) or pay for publishing process (Quan, Chen, and Shu, 2017) which dismisses whole sectors of publication. This implies that knowledge management tools in GP/FM must be controlled by dedicated, unbiased GPs.

The movement no-gracias (Spain) (http://www.nogracias.eu/, the movement of free lunch doctors(USA) (http://www.nofreelunch.org/, the Medico Sin Marca (Chile and Colombia),
(http://www.medicossinmarca.cl/), the Therapeutic initiative (Canada) (http://www.ti.ubc.ca/) or more generally members of the International Society of Drug Bulletins (ISDD) (http://www.isdbweb.org) refuse to adhere the pharmaceutical industry influence. Their work merits distributing its grey literature in a professional multilingual indexing system that is accessible by all. Finding a way to publish and share information outside the default language of English is attractive to many GPs. However, sharing the results of research with General Practitioners is crucial for the survival of the discipline of GP/FM, which means a universal system must be created (McIntyre et al., 2016).

As an example of hidden grey literature, the website of the World family doctor association in Europe (www.woncaeurope.org) edits more than 30.000 non-indexed abstracts of European or world conferences in English. Each abstract often represents two years of work (Master's Theses are included in this), and if not published is lost work. This also represents missed opportunity to develop networking between authors sharing the same centers of interest. Only half of this production has the chance to be published in indexed medical journals (Van Royen et al., 2010; Hummers-pradier, 2007).

1.1.2 Producing Information at the Point of Care

It could be believed that a simple terminological subset may be sufficient to meet the needs of GP/FM computer systems. Moreover, special tools specifically dedicated to primary health care, such as the International Classification of Diseases, tenth revision, for primary care (ICD-10 PC) (Ustün et al., 1995) or the Statistical Manual of Mental Disorders, 4th ed., primary care version (DSM IV PC) (Pingitore and Sansone, 1998) have been developed and quickly nicknamed ICD-10 or DSM-IV for dummies.

Lack of visibility of the complexity of the work of family doctors have allowed such a partial, biased vision. Of course, the family physician often sees simple problems. But he is accompanying a set of patients throughout a lifetime. The family physician knows a patient more extensively than most specialists do. He also becomes a specialist in patients bearing rare diseases and of various cultural background. He will, therefore, have extensive terminological needs, even more extensive than many specialists.

An interpretation can be seen in the Pareto curve (Fig.1.5), showing the importance of the Long Tail that makes it possible to understand this phenomenon. GPs see many patients with common diseases (to the left of the figure). But throughout life, they see probably an equivalent amount of more rare situations or rare diseases (to the right of the figure). This implies the use of specialized interface terminologies when needed. It would be inappropriate to say to a patient; *your disease doesn't fit my computer knowledge*.

FIGURE 1.4: Your disease don't fit with my computer knowledge. Pen draw by Guy Barbier

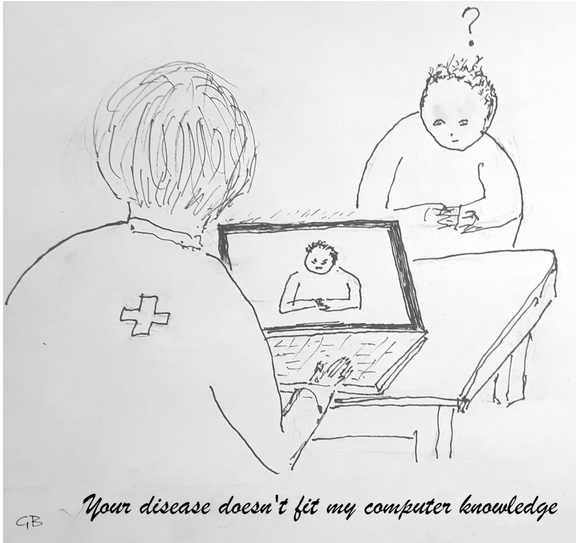
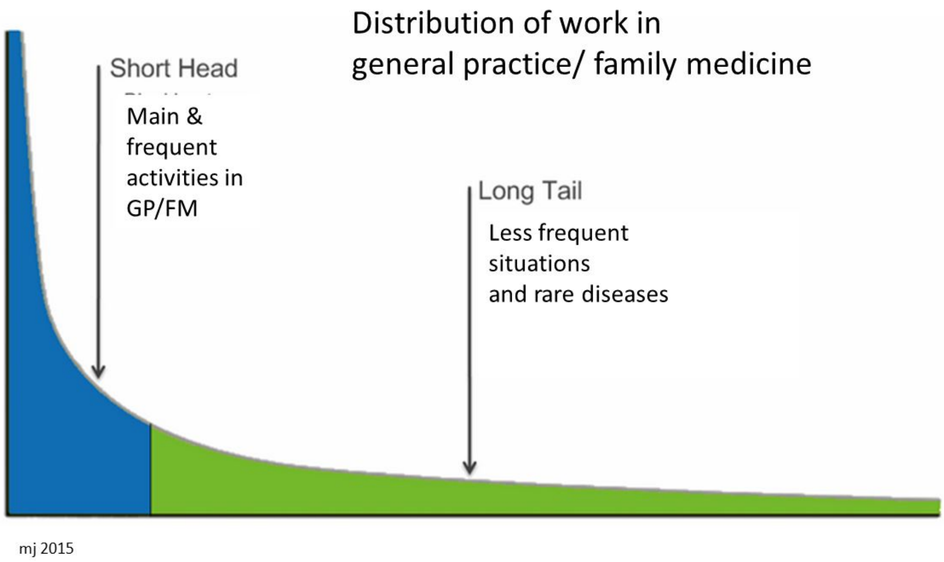


FIGURE 1.5: The Pareto curve shows the hypothetical distribution of patients seen along a physician career. At the left, frequent occurrences of problems, ate the right less frequent or rare occurrences. The surfaces of the two parts are equal.



1.1.2.1 Clinical Information

The adjective clinical deals here with patient related data, such as: reasons for encounter, symptoms, acts performed or requested and diagnosis. Terminology for clinical information is a highly specialized and difficult field of current medicinal research (Jamoulle et al., 2014). Clinical information is accumulated in EMRs and, if well organized, transferred to study centers where huge database (Charlton et al., 2010; Carey et al., 2004; Britt et al., 2003) may be used to teach medicine, analyze epidemiological data or be used for secondary searches (see Fig. 1.2).

Among others, studies worth citing are, mostly with good validity, produced by the Dutch Transition Project (Soler et al., 2012) (http://www.transhis.nl), data produced by the GP/FM led Belgian Intego project (Bartholomeeusen, Buntinx, and Heyrman, 2002) (https://intego.be/en) , the Beach project in Australia (Britt et al., 2016) or at a larger scale the UK General Practice Research Database (http://gprd.com) (Khan, Harrison, and Rose, 2010).

1.1.2.2 Professional Contextual Information

During the development of this work, the researchers used - successively - the terms *metaclinical* then *managerial* and finally *contextual* as a generic term for the name of the taxonomic product. A concept like *uncertainty*, the usual companion of the doctor or the concept of *quality assurance* or this one of *environmental health*, are all essential elements of professional practice. These are not clinical terms as they do not always deal with current patient problems. The term metaclinical, therefore, was deemed appropriate. As research progressed, general management items appeared to be prevalent and the term *managerial* was considered. However, the term *contextual* appeared the most relevant, as it was defined in the Meriam-Webster Dictionary as: *the interrelated conditions in which something exists or occurs*.

Answering the question: *What are they discussing?* in a meeting of two or twelve or several hundreds or thousands of GPs may give insight into the details of this well-defined (Jamoulle et al., 2017b) but not limited profession (see chapter 7, page 169). That has been the driving force behind this study. As stated by Cimino (1998) : *Part of the difficulty with using a standard controlled vocabulary is that the vocabulary was created independent of the specific contexts in which it is to be used*. By adding a contextual supplement to ICPC, the intention is to add precision to its domain of applicability. Controlled vocabulary *is a general term for a list of standardized terms used for indexing and information retrieval usually in a defined information domain* (Library and Archives Canada, 2017). In this case the Controlled vocabulary is also a *Vocabulary coding scheme* as defined in Dublin core (see further).

This approach to the doctor's professional context should not be confused with the contextual approach of the patient's universe, as developed by Schrans et al. (2016) who studied the elements of the patient's life context that influence his or her state of health and the health problems he shares with the doctor.

1.1.3 Consuming information at the point of care

As stated by James (2016) the Internet has triggered transformational change in the dissemination of science in the form of a global transition to open access (OA) publishing. GPs, the rank and file (Chinitz and Rodwin, 2014) workforce in medicine are using those resources extensively despite sometimes huge material difficulties to access the sources when working in rural or remote area (Salman Bin Naeem, Shamshad, and Amjid, 2013).

Finding information is sometimes difficult for researchers, even though they may not see patients and remain mostly in academic laboratories with access to expensive medical journals. So, what about *rank and file* physician on the spot who has a few seconds to check information and the relevance of his source (Hubbard, 2008)? Availability in real world, at the point of care access is clashing with the economic model (paid access) or with copyright issues (Myška and Šavelka, 2013).

Open access (OA) to researchable and usable information has several advantages (better metrics, end-user uptake, cross fertilization) (Eysenbach, 2006). Availability at the point of care is of utmost importance in GP/FM (Heilman, 2015) . It is necessary to maintain open access, point-of-care resources at the high level of quality that patient care demands (PLoS Medicine Editors, 2015). A potential issue with this is that GPs should consider that daily medical journals as well as major papers can also be manipulated by the pharmaceutical industry (Schwitzer, 2017; Dowden, 2015).

Another problem is the low quality of available usual medical information. Ioannidis et al. (2017) call health care professionals to *take up the responsibility to master skills and become teachers and trainers* to afford the *Medical Misinformation Mess to encompass the set of issues that relate to the low quality of medical information*.

1.1.3.1 Sources of information at the point of care in GP/FM

On-line, directly accessible major documentary databases in open access and local language are not numerous nor specific to GP/FM (Hubbard, 2008). This doubles the workload of a GP, while simultaneously making it more difficult to treat a patient accurately.

The US National Library of Medicine PubMed database (https://www.ncbi.nlm.nih.gov/pubmed/) gives access to millions of paid and open access publications. The Pan American Health Organization sponsors Scientific Electronic Library Online (http://www.scielo.org) SciELO is the South American champion of free access journals (PAHO Bireme Sao Paulo, 2016). The World Health Organization supports the Global Index Medicus (http://www.globalhealthlibrary.net). Language are also an issue. As an example, in France, the LiSSa database (http://www.lissa.fr) gives access more than 10^6 citations of French literature (Cabot et al., 2017a).

To say nothing of Google Scholar, the web is a considerable resource of information in medicine, especially in gray literature. Janamian et al. (2016) have identified and searched for 260 web sites as GP/FM sources. As stated by González-González et al. (2007) *"In primary care, each practitioner encounters more than 500 different clinical topics in any year"*. Ten years ago Internet base accounted for only 5% (ibidem) for search of information by Spanish GPs. This percentage has risen to 59% for German GPs in 2016 (Eberbach et al., 2016). Anyway the use of Internet is now so generalized that studies are performed on influence of Internet on the patient-physician relationship (Tan and Goonawardene, 2017).

1.1.3.2 Available indexing systems and descriptors in GP/FM

The 70s saw the birth of groups and organizations that would determine the future of general practice. The European General Practice Research Workgroup (EGPRW) which would later become the European General Practice Research Network (EGPRN) and the European Academy of Teachers in GP/FM (EURACT) took flight in this decade. WONCA, the world organization of family doctors, was founded in 1972. The Leeuwenhorst European Working Party (1980) would be decisive in this adventure. GP/FM would also claim its place as a separate discipline in this time. *General practice is a scientific discipline within medicine and has a specific place in a comprehensive health care system* (()European General Practice Research Workshop 1983).

The field of classifications is also well studied in this time (Bentsen, 1976; Bentsen, 1986; Braun, 1965; Braun, 1970). The year 1976 saw the first publication of the International Classification of Health Problems in Primary Care (ICHPPC) (Schneeweiss et al., 1977). In the meantime WONCA has launched a bibliographic committee which will be very active in publishing FAMLI during several years (see further). Nevertheless, the Medical subject Headings (MeSH) remain the most used indexing tool for GPs. In South-America, the Health Sciences Descriptors (DeCS) are the rule. This paragraph address those three tools. A deeper analysis of MeSH for GP/FM use is available at 1.3.7.3, page 77.

- The Family Medicine Literature Index (FAMLI) 1980 – 1992

  In 1980, the College of Family Physicians of Canada, the National Library of Medicine (U.S.) and the World Organization of family doctors (WONCA) provoke a significant movement in the treatment of the specific literature of GP/FM. They edit jointly, under the lead of Dr Jan McWinney, the Family Medicine Literature Index, an index for the literature specific to family medicine system (CFPC, 1980-1992). As pointed out by D. Fitzgerald: *Family physicians cannot always rely entirely on Index Medicus and Medline to provide the information they require. A number of family medicine journals, including Canadian Family*

Physician, are not indexed by these major sources. (Fitzgerald, 1980). This statement is emphasized by Lynn Dunikowski, Director of Library Services of the College of Family Physicians of Canada and second editor of FAMLI, who notes : *Because MeSH originally was intended for use with clinical documents, the problem has always been how to adapt it to cover the range of a family physician's non-clinical activities*. (personal communication Aug 2014).

E. Challis, Chairman of the Teachers of Family Medicine, College of Family Physicians of Canada underlines the role of WONCA ; *Dr. Ian R. McWhinney, the chairman of the WONCA Bibliography Committee, who was also instrumental in the development of this informative reference* (Challis, 1981). D. Fitzgerald gives more information about additional features of FAMLI, such as the five-year cumulative list of books written by or for family physicians. *Each annual cumulation of FAMLI includes a five-year cumulative list of books written by or for family physicians. Note that this was a comprehensive, rather than a recommended, listing. The main content of FAMLI was always the index of the journal literature; the list of books was secondary*. Moreover, the FAMLI index was edited under the name of WONCA as a quarterly publication, including articles from many journals not referenced in Index Medicus (Doherty, 1987), which shows the importance given to this area by the organization of family doctors at this moment.

In the 90s, the specific GP/FM literature began to boom (see Fig. 1.2) and lists of publications were regularly made available. (Dunikowski and Weston, 1986; Weston and Dunikowski, 1992; Verhoeven, Boerma, and Meyboom-de Jong, 1995). In September 1992, the volume 12, issue of the series FAMLI, comprising a set of keywords specific to GP/FM under the title *KeyWords in Family Medicine (A Thesaurus)*, was edited by Dunikowsky (1992). It was actually included in every volume of FAMLI, not just the 1992 volume. It was revised with every volume, taking into account new MeSH and new terms commonly used in the practice of family medicine.

Precisions Dunikowsky (1992) brings to the approach deserves mention, because she already fully notes some discrepancies between MeSH and the world of family medicine:

> *Some terms commonly used in the field of family medicine do not appear in Medical Subject Headings (MeSH). The thesaurus bas been designed to act as a bridge between these terms and the MeSH headings used in FAMLI. Family medicine terms appear in the thesaurus with the nearest equivalent MeSH term (or terms) used in FAMLI printed in bold-face type. The thesaurus also provides brief explanations of selected MeSH terms, to help in choosing the term that most closely corresponds to a selected*

> *topic. Some MeSH terms are listed together with closely related MeSH terms, to suggest ways of broadening a search.*

- Use of Medical Subject Headings (MeSH)

  For retrieval of scientific bibliographic information in GP/FM, Medical Subject Headings (MeSH) are used by all doctors and researchers. The MeSH is a huge terminology of 27.000 hierarchically managed descriptors- i.e., normalized terms, intended to index medical documents and growing yearly (Jamoulle 2016). While ICPC, Read Codes or SNOMED-CT or other medical terminologies are intended to retrieve data about people, indexing system like MeSH intends to retrieve documents or at least references to them (Vanopstal et al., 2011). It is a difficult exercise to identify in the MeSH descriptors appropriate to the GP/FM. The related MeSH concepts are vague and dispersed in many occurrences, compelling the construction of complex filters (Mendis and Solangaarachchi, 2005; Jelercic et al., 2010).

  The Index medicus, the forerunner of MEDLINE, was the bible of U.S. health care. (Greenberg and Gallagher, 2009). The MeSH were built on it, mainly in the second half of the 20th century and was influenced by the technological development of medicine. General medicine and primary care are the poor parents of this development. Mesh has been intended to index medical journals. Recently use for indexing civil data in health care domain is being studied; (Marc et al., 2015).

  General Practice is a profession generally regarded as part of the first line of care, Primary Health Care (PHC), a form of care organization. General practice and Primary Health Care concepts share the same extension but not the same intension. Intension designates, according to Sadegh-Zadeh (2015), *the informational content of an expression*. While it is commonly accepted that General Practice and Primary Health Care have as an extension; comprehensiveness, continuity and accessibility of personal care (Starfield, Shi, and Macinko, 2005), it is not possible to confuse a function with its organization of work. Yet, curiously enough, the confusion between the two is maintained in one of the most prominent sources of medical knowledge.

  This is largely due to the confusion found in the MeSH terminology base. Gill et al. (2014) states that: *Constructing a highly efficient search filter to identify primary care relevant articles is challenging, particularly due to the inadequate and ambiguous description of the clinical research setting in title, abstract and MeSH keywords*. Huang, Névéol, and Lu (2011) observes that: *manually assigning MeSH terms to biomedical articles is a complex, subjective, and time-consuming task*. Shultz (2007) argues that: *terminology was observed to be a major factor affecting retrieval and the ability of both systems to obtain unique items*. However, not all aspects of the broad field of GP/FM are covered in a specific area (Sladek et al., 2006). Despite this, interesting advance in MeSH indexing for GP/FM have been proposed (Mendis and Solangaarachchi, 2005) (Jelercic et al., 2010). Themes of interest searched for by GPs have also been studied. (Hong et al., 2016)

The future looks prepared as MeSH are now available in RDF, ready for Semantic web(Bushman, Anderson, and Fu, 2015). This subject is studied in depth at paragraph 1.3.7.3, page 77.

- Use of Health descriptors (DeCS)

  The controlled vocabulary of Health Sciences Descriptors (DeCS), initially a translation of MeSH into Spanish and Portuguese, has been translated also in French and expanded with new categories. It was also adopted into the indexing and multilingual search of the scientific and technical literature in South America. (`http://decs.bvs.br/I/homepagei.htm`). It's a by-product of the Latin American and Caribbean System on Health Sciences Information (BIREME), the Pan American Health organization (PAHO) network of libraries and documentation centers (Neghme, 1975).
  Doctors and student in South America are using DeCS as a standard controlled vocabulary for indexing scientific and technical health related documents in knowledge databases like the Virtual Health Library (`http://bvsalud.org/`) or SciELO (`http://www.scielo.org`). Both sources generally give open access to documents.
  Within the framework of this thesis, a formal cooperation was born between BIREME, the Pan American World Health Organization library and D2IM, which allowed: (i) the translation of the DECS into French and its integration in HeTOP, (ii) the use of DECS in the various tools of D2IM, in particular CISMeF and LiSSa, the French equivalent of PubMed for English, or LILACS for South and Central America (Spanish & Portuguese). Thanks to InfoRoute, any LiSSA query can be launched in LILACS, the most important and comprehensive index of scientific and technical literature of Latin America and the Caribbean (url http://lilacs.bvsalud.org/en/) This allows the Spanish and Portuguese speakers of France to have contextual access to LILACS.

1.1.3.3 ICPC for classifying clinical issues

The International Classification of Primary Care (ICPC) (Soler, Jamoulle, and Schattner, 2015) is routinely used by physicians around the world to categorize the problems encountered in their practice with patients- i.e., their clinical activity. We will see that ICPC has proved effective in many other uses and show that it is adapted for collecting clinical problems that doctors discuss at congresses. We are giving a short review of ICPC as well as its relation to the Systematized Nomenclature of Medicine - Clinical Term (SNOMED-CT). For further documentation on the issue of available terminology in health care see (Lagasse et al., 2001) and (Jamoulle, 2010). ICPC is presented in copy desk format in the appendix B, page 195.

- ICPC

  Primary care isn't specialized care. It encompasses specialized care and biotechnological vocabularies along with anthropological, family-based/personal information. A medical record in primary care must encompass all facets of health care, alongside personal and family environment knowledge. This is why, WONCA, the world organization of family doctors, through the ongoing work of its International Classification Committee (WICC), has developed the International Classification of Primary Care (ICPC) (Lamberts and Wood, 1987) (WONCA, 2005) (Lusignan, 2005).

  ICPC was not constructed to reflect the contextual aspects of the profession, such as: managerial, ethical, environmental, educational, or research aspects. The WICC was set up by the WONCA Council in 1972 in Melbourne for the Fifth World Conference (Jamoulle et al., 1999). WONCA recognized the need for classifying and analyzing data derived from clinical encounters with patients, unique to family medicine. This would become increasingly important with the advent of electronic health records. In 1987, the three existing classification systems, the International Classification of Health Problems in Primary Care (ICHPPC-2), the classification of procedure in primary care (ICPC-Process) and the classification of Reasons for Encounter (RFE) - were merged into a single one; the ICPC. ICPC-2 was published in 1998 by the WONCA International Classification Committee, after several years of revision, adding definition , exclusion and inclusion criteria and cross-mapping with the International Classification of Diseases, tenth revision (ICD–10) (Jamoulle, 1998; Okkes, Jamoulle, Lamberts, and Bentzen, 2000).

  ICPC is a coding and classification system that reflects the distribution and content of the domain of family medicine. It also acts as its ordering principle. It was designed as an epidemiological tool to classify and analyze data over three important elements of the health care encounter: the reasons for encounter (RFE), the diagnosis or problem, and the process of care in family medicine. ICPC-2 is available in more than 2O languages. It is compact, and the entire classification can fit onto two A4 pages. ICPC is compatible with problem-oriented clinical records (Shukor, 2017).

  WICC is developing the third version of the classification. ICPC-3 has been in development for many years, using data collected with ICPC and other coding systems to allow expansion and modernization of the classification without losing its core characteristics and values. Further information is available at www.ph3c.org, including the ICPC training module and an extended bibliography. (Soler, Jamoulle, and Schattner, 2015).

ICPC is not a terminology, but a classification system used - often in conjunction with - the International Classification of Diseases (ICD) in General Practice/Family Medicine (GP/FM) for usual clinical purposes - or with any terminological interface system.

ICPC-2-E, the electronic version of Chapter 10 of the book ICPC-2 (WONCA, 2005) , is to be used specifically in an electronic patient record and for research purposes. Several ICPC mapped terminologies like ENCODE-FM (Rayner, 2009), the Transhis Thesaurus (Becker et al., 2005) or ICPC Plus (Wang et al., 2008), have been developed. (See https://goo.gl/KjvY9C). ICPC-2 is to be used together with the first nine chapters of the book ICPC-2, indispensable to make a correct use of ICPC. ICPC-2-e is to be downloaded in its latest version, currently ICPC-2e-v.6.0 (6 April 2017) on the Norwegian e-health web site https://ehelse.no/icpc-2e-english-version.

The WICC is preparing ICPC-3 since more that 10 years.

In 2017, it was possible to establish a primary health care classification consortium which, to date, brought together WONCA, WONCA Europe, the Norwegian health authorities, the Brazilian Society of General Practice and the Dutch Society of General Practice. The objective is to publish ICPC-3 in the near future, under the guidance of the WONCA International Classification Committee.

- SNOMED-CT "SNOMED Clinical Term" (or "SNOMED-CT") is the name proposed for a new system of clinical coding, which fuses the Read Codes of the National Health Service (UK) and SNOMED-RT of the College of American Pathologists (U.S.). Thus, SNOMED-CT. It could potentially become the standard clinical terminology for Anglo-Saxon dependent countries, and it is currently marketed worldwide, as an expression of the U.S. soft power. First, it was marketed under the umbrella of a non-profit Danish association called IHTSDO. Since 2017, it is marketed by a non-profit society called SNOMED International (http://www.snomed.org/). SNOMED conveys the American medical subculture (Jamoulle, 2010). Although the product is not available in full, in French or Dutch, the Belgian e-health authorities have made the choice to impose SNOMED-CT as the de-facto standard. This will soon be compulsory in Belgian health informatics.

  SNOMED is a multi-axial systematic nomenclature. It allows to project medical concepts along several orthogonal axes. The variety of these axes distinguishes this nomenclature from a mono-axis classification such as the International Classification of Disease, which is essentially limited to one type of concept: diagnostics. Furthermore, each axis is itself hierarchically structured, with the various levels being linked by relations of specialization.

Several on-line browsers allow to consult SNOMED-CT. The latest in a series of browsers and terminology tools developed to demonstrate terminologies from the Read Codes through to today's clinical terminology SNOMED CT® (Hina, Atwell, and Owell, 2011) is CliniClue® Xplore (`http://www.cliniclue.com/cliniclue_xplore`).

WONCA and IHTSDO have developed a joint work-group for the creation of a SNOMED-CT / ICPC terminological subset.

- SNOMED-CT / ICPC subset

  The International Family Physician/General Practitioner Special Interest Group (IFP/GP SIG) was established (2010-2015) to suggest content for the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) related to general/family practice and to provide quality assurance for SNOMED CT content from the GP/FM practice perspective. The aims of this project were to create two interdependent products: an international general / family practice reference set (RefSet) of SNOMED CT (called the GP/FP RefSet), containing SNOMED CT concepts frequently used by GPs/FPs and a map from the GP/FP RefSet to ICPC-2. The use of the map from the SNOMED GP/FP RefSet to ICPC-2 is free within SNOMED member countries and available from the SNOMED CT local distribution center. Only those who have licenses for both SNOMED-CT and ICPC-2 can access the map from the GP/FP RefSet to ICPC-2. The results of the Field Test of the RefSet should be on the SNOMED International website (GP/FP SIG section). No publications on usability of this subset are available.

1.1.3.4 Paper & Pencil: Lambert's seminal idea of Q-Codes

In the Department of General Practice at The University of Amsterdam, in the 80s, the task of reading medical journals specific to General Practice was devoted to teachers and assistants. Articles relevant to the profession were indexed by ICPC and non-clinical subjects were indexed by Q-Codes. Using the letter Q, available in ICPC, the late Professor Henk Lamberts has opened this new category for use in bibliographic retrieval.

This, however, was before the Internet, even before Medline became available on CD-Rom. To the researcher's knowledge, this seminal list of Q-Codes has never been published outside of a copy from 1987, taken from the tables of the department library. This copy has been preserved. In Figure 1.21, page 48, one can see the list of Q-Codes proposed by Professor Lamberts, also a linchpin to the development of ICPC in the WONCA

International Classification Committee (WICC) and the WONCA working group in the classification field.

1.1.3.5 Grey Literature as a source of information

The price of access to international high-level journals, mostly exclusively in English, is prohibitive. The economic aspect is therefore a major obstacle to the spread of knowledge outside academic circles. However, the change of economic model is under way in the world of publishing, the cost of which is largely reflected on the author and not the reader. In the mean time, appeals for public access of research data continue to proliferate (Lin and Strasser, 2014). McKenzie (2017) considers that the explosion of the use of Sci-Hub facilities (`http://sci-hub.cc`) is the beginning of the end of the scholarly publishing. The sci-hub site, dedicated to the legal and illegal distribution of scientific literature, is sometime considered as the Robin hood of the literature and used by millions of researchers (Greshake, 2017).

Whatever the case, the open access grey literature in medicine is in full development (Swan, 2012). According to Schöpfel (2015), *The term gray literature remains ill-defined, imprecise, with fuzzy outlines. Its two handicaps are part of its definition: identification, access and acquisition are often difficult, and quality and reliability are not always assured.* Ferreras Fernández (2016, p.211-216) has done an exhaustive review of the definitions of grey literature. Those definitions share the negation of commercial involvement like Pisa's (GreyNet, 2014). Grey literature reviews are not always free of access like the Grey Journal itself (`www.greynet.org`).

Practically, when addressing the case of grey literature, authors exchange more pragmatic definitions than the Pisa's one as; *difficult to locate or retrieve* (Moher et al., 2000), or; *has not been formally published* (Hopewell et al., 2007) or; *there is no such peer review or passage through quality filters* (Silva, Garcia, and Cássia, 2009).

Interestingly, Hoffmann et al. (2011) points out that the *grey literature yields more substantial information* (than white literature) *on the content of interest*. This could be understandable, partly as white medical literature is not free from the influence of the industry (Gotzsche, 2013; Schwitzer, 2017).

We propose to consider grey literature in GP/FM publications that share the following characteristics:

- For the background:
  - Sharing knowledge specific to the field of GP/FM no matter the format (Paper, articles, memo, master thesis, PhD thesis, leaflet, abstracts of presentation, web pages, video, images, youtube, Facebook, Twitter, Google+, Linkedin, dataset.

- Being unreferenced in well known local or international medical databases (PubMed, LiSSa, Scielo, Lilacs, ORBI, etc.)
- Being submitted to a scientific quality assurance process (in anthropology or bio-sciences)

- For the format:
  - Being freely accessible in an Open access model
  - Using a systematic multilingual vocabulary encoding scheme (indexing system)
  - Relying to Dublin Core Metadata Initiative or equivalent standardization process
  - Being ready for machine use in the semantic web

1.1.4 Metadata and Vocabulary Coding Scheme

Metadata consists of statements we make about resources to help us find, identify, use, manage, evaluate, and preserve them (Sutton, 2007). Metadata may be interpreted by machines and people. Dublin Core Metadata Initiative (DCMI) (http://dublincore.org/) provides simple standards to facilitate the finding, sharing and management of information. Metadata are basic description mechanism for digital information that, can be used in all domains, for any type of resource, simple, yet powerful, can be extended and can work with specific solutions, making it easier to find information on the Web as it develops. DCMI participates in the development of the "new Web", the Semantic Web and Linked Data (Dekkers, 2009). Allen (2016) states that *The emergence of machine intelligence and machine reading in the second machine age will make it even easier to automate the production of metadata to help people find, filter and organize information.* . Catalogues must use standard tools, especially metadata, to describe and index resources (Darmoni et al., 2001). Quality of search results is dependent on the quality of the metadata in the original repositories of which high quality structured metadata are more accessible (Newbold and Grimshaw, 2010).

We are thus dealing with knowledge identification process by humans and by machine through well formalized denominations. So we are addressing here the concept of *Vocabulary Coding Scheme*, in other words a *Controlled Vocabulary*. The reader has be conscient that the same concept could bear different names. For instance the Australian Metadata Online Registry (MeteOR) uses the term *Classification scheme* for pointing the same issue. *A classification scheme is an official terminological system, recognised and endorsed by a national or international body, that is used to classify data.* (http://meteor.aihw.gov.au). Another standard mapped to dublin Core is MARC acronym for MAchine Readable Cataloging. This is a computerized method of recording the information needed in a cataloging record: the descriptive cataloging, subject headings and other access points, and classification numbers and other call number information. (congress, 2017).

1.1.5 Grey literature and semantic web opportunities

Metadata allow the retrieval from data from dedicated repositories. Nevertheless, as stated by Goggi et al. (2016), *documents may contain important information that has not been encoded in the metadata.* Extracting key concepts from unstructured texts is the following step, done by semantic annotators, by-product of research in Natural Language Processing (Cabot et al., 2017b). Key concepts could be added to indexing facilities or tagged as identifiable information for use in Linked Open Data (LOD). This open the *possibility of enhancing the visibility and accessibility of grey literature via its connection to the data it describes and to an advanced full text indexing* (Goggi et al., 2016).

1.1.6 The semantic world

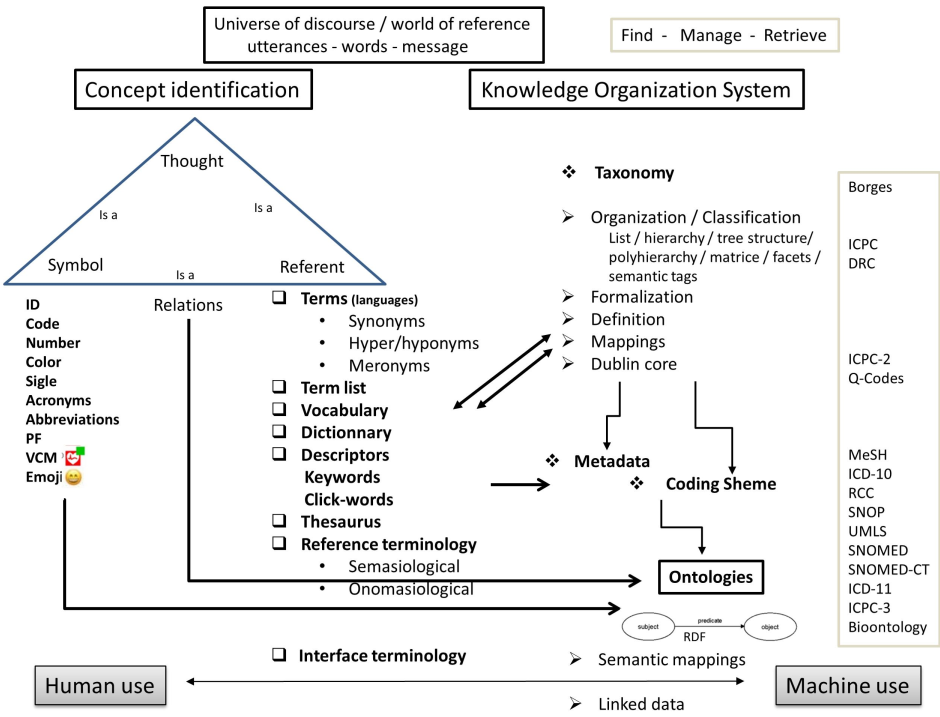
Figure1.6 attempts to synthesize the semantic universe and the birth of ontology. Managing knowledge (finding, managing, and saving it) of a domain of human activity implies the identification of its universe of discourse, which is also called its world of reference. This discourse is expressed by utterances, terms, and messages. The concept and its symbolic identification are decisive in initiating this process. Languages are only human referent. Terminology science is the standardization of language. The semasiological organization of reference terminologies allows for the use of interface terminology and for an exchange between human and machine.

The other pole of knowledge management is organization. Taxonomic science allows for the structuring, formalization and standardization needed for computer use. By following the arrows in Fig. 1.6, we see, that in a given domain, the alignment of the identifying symbols, relations, semiological terminology, metadata and classified nomenclatures induce the genesis of ontologies. This is represented by the graph of a predicate and expressed in a specific computer language for the management of humans / machine relations and distributed data in the Linked data universe. The lateral strip of the figure shows the evolution of the complexity of some classifications, in particular the field of health information systems. Let's go deeper in those concepts.

1.1.6.1 Universe of discourse, World of reference

Health-care has become a world-wide information-intensive industry. The growing dominance of the computer industry in health-care requires a standardization of the interface between man and machine; therefore, terminologies appear to be unavoidable (Chute, Cohn, and Campbell, 1998). Nevertheless, the vocabulary in medical terminologies can have very different meanings. A consultation between a patient and a doctor is the meeting of two *Universe of discourse* or two *worlds of reference*. Those terms are used in both computer sciences and computational linguistics. Boole (1854) stated,..*whatever may be the extent of the field within which all the objects of our*

FIGURE 1.6: Out of complexity in Knowledge management; the path to ontologies



discourse are found, that field may properly be termed the universe of discourse (see also Moffett, 1967)

The semioticians are using the term *world of reference*, a concept addressed by Umberto Eco in his book *The role of the reader* (Eco, 1979; Smith, 1980). We are addressing here the *universe of discourse* or *world of reference* of the profession of General Practitioners, also called Family Physicians. This profession is generally presented by its acronym GP/FM for General Practice/Family medicine. The duality of this profession is already present in this acronym. GP, meaning General Practice, is about mastering all processes of care. Generally, it is a term used more in The United Kingdom. FM meaning Family Medicine, more used in the USA, specifies that the doctor is caring for families, i.e. people. Thus, the acronym itself address the balance between process and people, reflecting the balance between technology and anthropology and highlighting the somewhat hazy character of this field.

Semantic conflict between science and culture are at the heart of terminological sciences (Meneghini and Packer, 2007), but variation of meaning in health-care terms - along with their worlds of reference and conditioned by historical perspectives -, have not been analyzed properly. The Unified Medical Language System of the US National Library of Medicine

(UMLS) followed by SNOMED-CT (Jamoulle, 2010) of the former International Health Terminology Standard Development Organization (IHTSDO), now SNOMED INTERNATIONAL, are the leading terminologies in health care. These reservoirs of knowledge are not yet adequate, concerning the world of reference of GP/FM. Each has its own history, which explains some of their discrepancies (Jamoulle et al., 2014).

Additionally, the historical approach of the Medical Subject Heading, a huge terminology constructed over time (Lipscomb, 2000), explains striking particularities in information retrieval in the Medline online library when dealing with GP/FM issues.

The key factor in defining which universe of discourse one is dealing with is the entrance into the world of semantics. We know that the family physician is the connecting point between science and patients. This position can often be antithetic, as medical discourse is often built purely on the biosciences approach, while patient discourse is built on his or her phenomenological experience.

Gunnarsson, Linell, and Nordberg (2014, p.99) hypothesize that the professional universe of discourse is created by three layers : (i) the specific knowledge based or cognitive layer; (ii) the social representation layer establishing their political and economic patterns and supported by a specific language; (ii) the internal cohesion layer establishing the group identity and the distance to other groups.

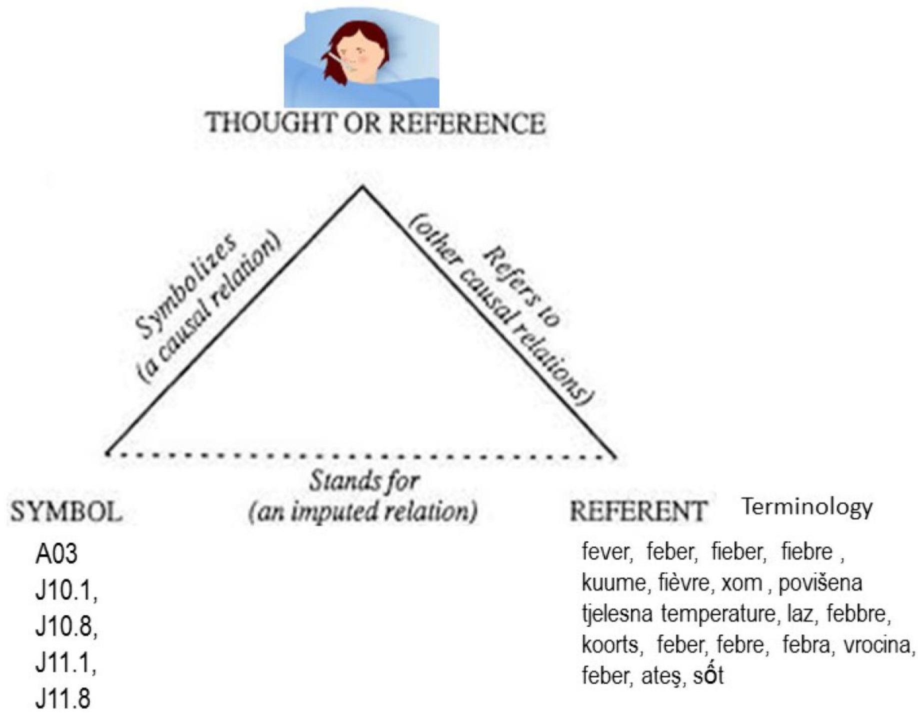
A pre-existing set of properties are referred to by the reader (Eco, 1979) and defined as his/her world of reference, which is subjective to a person's own. This will be transformed if the couple (author/reader - doctor/patient) do not have the same world of reference. This may also manifest as an approximation of meaning, since the author and reader may not refer to the exact same concept, though they may be similar. This is the case in the doctor/patient relationship, and this will pave the way for interfaces between lay mans terms and professional terminologies in health care. (Cardillo, 2011)(Cardillo, 2015b)(Vander Stichele et al., 1991)(Smith and Fellbaum, 2004).

1.1.6.2 Concepts

Organizing a world of reference implies dealing with concept identification (SYMBOL), concepts (THOUGHT) and content - called definitions- scope notes or criteria (REFERENT). Identifiers, concepts and content are connected in a triangle, as proposed by Richards and Ogden (1930) well discussed by (Smith et al., 2006) (see Fig.1.7 and Fig1.6, upper left).

Cimino (1998) states that *most systems that report using controlled vocabulary are actually dealing with the notion of concepts* and that *The unit of symbolic processing is the concept*. This implies that a taxonomy must be concept-based and, consequently, independent of language.

FIGURE 1.7: The concept of fever along the Ogden & Richards triangle (Richards and Ogden, 1930) Fever is visually represented by a kids with a thermometer (THOUGHT OF REFERENCE), symbolized by an Identifier (SYMBOL) in the classification ICPC-2 & ICD-10 and referred to Entry terms of ICPC in 20 languages (REFERENT).



1.1.6.3 Symbol & codes

Relationships between symbol and meaning could be regarded as arbitrary. However, symbols could be integral to organizational life and communication (Rafaeli and Worline, 1999). Symbols, for example, could solely be numbers of the alphanumeric alphabet, but members of organizations will recognize it, just like acronyms or abbreviations.
As recognized by Cimino (1998) *the temptation is to chose hierarchically distributed code as they are several problems with using the concept identifier to convey hierarchical information. Reclassification is difficult and missing place is usual.* Codes could have structure with levels of specification like in UMLS (ex; Concept Uniform Identifiers (CUI) contains the letter C followed by seven numbers) or a determined meaning like in ICPC (a letter for each chapters, two digits for each rubrics).

1.1.6.4 Coding and redundancy

The use of the concept of redundancy in Information sciences comes from the Information Theory. As expressed by Shannon (1951); *redundancy is* *related to the extent to which it is possible to compress the language*. Loss of redundancy in signifier/signified (i.e. code/term) relationships will make the message incomprehensible to humans. The reader will find an example of loss of redundancy in this series; redundancy / rdundncy / rdndnc / rddc. The information in the meaning is lost to humans, but could still be significant to the machines. Organizing the loss of redundancy is the base for the normalization process, aimed at diminishing the loose of information when transmitting the message by machines. In a sense, one can say that normalization is the inverse of redundancy.

Redundancy (see Fig. 1.8) could be obtained by codes conveying visual meaning. More abstraction implies the loss of visual redundancy. Fever in the International Classification of Primary Care (ICPC) is [A03] which conveys (for insiders) the information A, general and 03 a rubric under 30 and thus a symptom; R50.9 could be recognized as pertaining to the International Classification of Disease (ICD-10) by sub-specialists in coding; The code C0015967 is recognizable as pertaining to Unified Medical Language System (UMLS) as a concept (letter C) but only machine are able to recognize 386661006 as the code of Fever in SNOMED-CT. We can also consider the language as a code conveying visual information. Note that, for an non Vietnamese locutor, the Vietnamese term *Sốt* [Fever] has no redundancy at all.

FIGURE 1.8: Coding conveys redundancy. The arrow show the gain in redundancy in languages, emoticons and coding systems in health care.

Redundancy	CATEGORY of CODE	SYMBOL	REFERENT	TYPE
	English language		Fever	language (en)
	Vietnamese language		Sốt	language (vi)
	Emoji		Fever	draw
	VCM		Fever	pictogram
	ICPC-2	[A03]	Fever	symptoms &complaints
	ICD-10	[R50.9]	Fever, unsp.	symptoms and signs
	UMLS Concept:	[C0015967]	Fever	sign or symptom
	SNOMED-CT Concept:	[386661006]	Fever	findings

Recently, Visual Concept in Medicine (VCM), an iconic language that may ease information retrieval tasks (Griffon et al., 2014) and emoticons (Emoji) have enriched the range of coding possibilities. Emoticons, small digital images or icons, are mostly used for the expression of emotion, for strengthening the verbal part of a message, and for expressing humor (Derks, Bos, and Grumbkow, 2008). Fundamentally, they are a kind of visual code, a shared meaning under a symbolic format that does not require a similar medical vocabulary between patient and doctor. It is worth noting that Emojis and pictograms could have a higher redundancy than terms, as they are language independent. Also, Cimino (1998) states that using a *name as a unique identifier for a concept limits our ability to alter the preferred name when necessary*. The unique identifier name is called Preferred Term (PF). It is often expressed in English, neutral or masculine, in lowercase, acting as a code. It is defined in the Unified Medical Language System (UMLS) as *The string preferred in a source or in the Metathesaurus as the name of a concept, lexical variant, or string.* (NLM, 2017). The National Information Standard Organization (NISO, 2005) defines PF as *One of two or more synonyms or lexical variants selected as a term for inclusion in a controlled vocabulary. (formerly known as descriptor).*

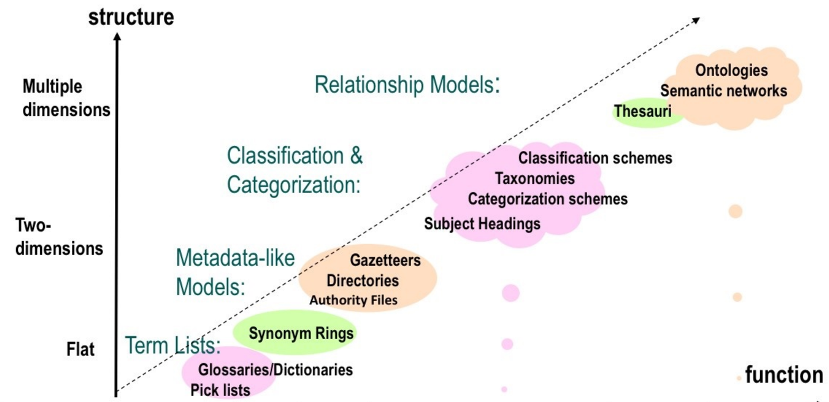
1.1.6.5 Classification, Taxonomy & Nomenclature

Knowledge Organization System (KOS) is a generic name which encompasses all subdivisions of knowledge, from the simplest to most complex. Lei Zeng (2008) proposes *four (although not mutually exclusive) groups, identified and ordered from simpler to more complex structures and also by functionality, following the figure of the semantic staircase: (i) Term Lists, (ii) Metadata-like Models, (iii) Classification and Categorization, and (iv) Relationship Models.* We will attempt to review the main terms used in the KOS world which, as shown in the figure 1.9, are growing in complexity.

Classification & Taxonomy are types of Knowledge Organization Systems (KOS). KOS, also known as *information retrieval language* are defined as vocabularies. They are more or less structured tools of digital network communication mainly used in Information Systems (Bratková and Kučerová, 2014). These KOS are used for researching, document indexing, electronic health records coding, and teaching.

Vanopstal, Vander Stichele, Laureys, and Buysschaert (2011) argues that *the terms taxonomy, classification, thesaurus, ontology and controlled vocabulary are used in many different contexts, including linguistics, bibliographic information retrieval (IR) and knowledge management, including medical coding.* One sees in the Fig 1.6 that terms, terminologies and thesauruses are used to determine the referent in the triangle of Ogden and Richard, and that they differ by their complexity and usage-rules. Used to identify information, they are modulated by rules of organization to become classification tools, or ontologies, as shown in the right part of the fig. 1.6.

FIGURE 1.9: Four groups of Knowledge Organization System (KOS). (Lei Zeng, 2008) (http://www.isko.org/cyclo/kos)



Major functions

eliminating ambiguity	xxx		xxx	xx	xxxx	xx
controlling synonyms		xxxx	xxx	xx	xxxx	xx
establishing relationships: hierarchical			x	xxxx	xxx	xxx
establishing relationships: associative					xxxx	xxxxx
presenting properties						xxxxx

A taxonomy is a controlled vocabulary managed as a subject-based classification, arranged in an hierarchical way (Garshol, 2004). It is a hierarchical classification of words, labels, and terms, organized into groups based on similarities. A taxonomy may be defined and centrally managed by one or more individuals. Taxonomies are useful as they provide a logical, hierarchical structure of meta-data that can be used to classify information consistently (Microsoft, 2017).

Hierarchical classifications contain a readily available wealth of category definitions, plus a hierarchy. They also reflect some degree of community consensus (Hepp and Bruijn, 2007). A classification, like the International Classification of Primary Care (ICPC) (Lamberts and Wood, 1987) limits a field and must follow rules as to hold mutually exclusive categories. Criteria and inclusion and exclusion criteria help to maintain the concepts very clear for those who are developing and for those who are using a classification. As pointed out by Bonnet (1999), *classification is a mental process which aims at the ordering of the universe and experience and which is based on a reference system*.

A nomenclature, like ICD, is a set of vocabularies that supports a singular classification system. The definitions, like those stated in the Dictionary for General/ Family practice (GP/FM) (Bentzen N.(ed), 2003) are delimiting the world of reference of the family physician. They are, in turn, delimiting the context of the activities.

A classification system must follow three basic rules (RCGP, 1973):

- *It has consistent, unique classificatory principles in operation*
- *Its categories are mutually exclusive*
- *Its system is complete*

The view developed by Lambe (2007) is more exhaustive. He describes nine key criteria for taxonomy validation. He cites that a classification system must be:

- *Intuitive ; reflecting natural working or usage habits*
- *Predictable ; demonstrating internal consistency for easy category retrieval*
- *Relevant ; reflecting user perspectives*
- *Unambiguous ; exhibiting a minimum of difficult choices*
- *Hospitable ; having the ability to accommodate potential new content*
- *Parsimonious ; exhibiting no redundancy*
- *Meaningful ; having terms that reflect common usage*
- *Durable ; not exhibiting frequent change*
- *Balanced ; being harmoniously populated*

On the other hand, a nomenclature doesn't need to follow those principles and *an agreed-upon naming scheme* (RCGP, 1973). As an example, ICD that one considers as a classified nomenclature, can't avoid the problem of redundancy. In turn, ICD uses the dagger and asterisk system to relate overlapping multiple entries.

Bentzen N.(ed) (2003) wrote in the preface of the dictionary of GP/FM that *during the development of ICPC-2 it became apparent that there is a great need to ensure that the concepts and terms we use should be defined in the same precise way*. His base was other initiatives that started in the seventies, when general practice (which is synonymous with family medicine in the dictionary) was growing in importance, but with a different scope in different countries. This implied a need for *standardization*.

Such tools as ICPC and the Dictionary have been developed in English, as it is the working language most used in scientific world. The availability of ICPC-2 in more than 20 languages shows the worldwide distribution of primary care concepts and allows doctors to receive and interpret information in the language of their patients.

But ICPC and the Dictionary are out of the Gutenberg age. The global knowledge network implies *a classified, mapped and interrelated dictionary-named terminologies or thesaurus for human reading and ontologies for machine use*. (Tao, Pathak, Solbrig, Wei, and Chute, 2013).

1.1.6.6 Terminology & thesauri

Once a concept has received a number and is translated into many languages, one tries to connect medical terminology to other languages. Classifications like ICPC or nomenclature (i.e. classified terminology) like ICD-10 are addressing mostly clinical concepts. Clinical means: in relation to specific care issues like symptoms, findings, processes done or asked, or diagnosis.

Clinical classifications and terminologies cover a large array of activities in health care. From procedural ones, like drugs (Anatomical Therapeutic Classification) to paramedicals, like International Classification for Nursing Practice (ICNP), all are in use in electronic Health Information Systems (HIS). The market for this kind of product is important, and the interface terminologies, those which make the link between computer and physicians, are in full development. As an example, Intelligent Medical Object (IMO) (`https://www.e-imo.com/`) is a Clinical Interface Terminology which bridges the gap between clinical language and complex coding systems.

Non-clinical classifications are not frequent in HIS and are sometimes confined to dictionaries. A non-exhaustive list of terminologies and dictionaries for non-clinical use is listed in the book *Terminology in Family Practice* (Jamoulle and Resnick, 2016) and reproduced in the Appendix G, page 223.

Medical Subject Headings (MeSH) thesaurus of the National Library of Medicine is particular, as it contains clinical and contextual entries. The particularity of the WONCA dictionary (Bentzen N.(ed), 2003) is its way of addressing non-clinical issues, such as the definitions of common organizational concepts like: quality, accessibility/continuity, medical ethics, content of teaching or environmental health in primary care. These represent contextual tasks and knowledge that General Practitioners are using daily.

Considering he ISO 16175-2 definition of a thesaurus in the ISO Online Browsing Platform (OBP) (`https://www.iso.org/obp/ui/`) : *Thesaurus ; controlled vocabulary and structured vocabulary in which concepts are represented by terms , organized so that relationships between concepts are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms*, we can argue that a such a thesaurus prepared in an ontology or web ontology language (OWL) for use by machine could be named an ontology.

1.1.6.7 Ontology

Cimino (1998) was incredibly perceptive, declaring that it *is likely that vocabularies will become concept-oriented, using non-semantic identifiers and containing semantic information in the form of a semantic network, including multiple hierarchies*. The Semantic Web aims to connect all concepts in order to provide decision support, to map different vocabularies/resources, and to allow machines to manage information (Jamoulle et al., 2015b) (see chapter 6, page 159). With an exponentially growing amount of information in all fields, it is no longer possible for humans to retrieve these concepts.

Ontologies are the machine based semasiological expression of all possible occurrences of meanings and links of a set of words in a defined domain (i.e. all terms of sailing ships or all terms related to electrocardiography) within the scope of human language. This possible multilingualism is due to the fact that the terms are related not to their lexical representation, but to the concept they represents. This may be expressed in various linguistic formats.

The notion of *concept* varies, and there is no consensus. For Nelson (2009) a Concept is a unit of thought, a specific meaning which refers to an object. Concept in MeSH is defined as the bearer of linguistic meaning (`https://www.nlm.nih.gov/mesh/xml_data_elements.html`). However, as suggested by Clark et al. (2012), we use it here in its abstract sense. Usually, the ontologies are in English, as a working language, but specifications sometimes allow for the interoperability of languages. For further multilingual insight, see `www.babelnet.org`

When complex terminology is developed to allow machines to speak to other machines, one creates an ontology. A good identification of the concepts at stake and their definitions will be the ground of a specific epistemology of general practice and will prepare the transition form terminology to ontology. (Gusso and Jamoulle, 2016).

There are almost as many definitions of ontology there are authors over said concept. A confusion could exist between the term "Ontology" used in philosophy, written in upper case, and "ontology" of computer science, written in lower case. Ontology is sometimes written *ontologie°*. The second, ontology° is a derivative of the aforementioned, restricted to use in philosophy (Sadegh-Zadeh, 2015, p.799).

- Ontology deals with what really exists.
- ontology° deals with representation of interconceptual relations.

1.1.6.8 Defining Ontology

Gruber (1993)'s definition is quoted here in full : *An ontology is a (formal),explicit specification of a (shared) conceptualization [. . . .] When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge.*

The term conceptualization is well explained by Uschold and Gruninger (1996); *Ontology is the term used to refer to shared understanding of some domain of interest. An ontology necessarily entails or embodies some sort of world view with respect to a given domain. The world view is often conceived as a set of concepts, their definitions and their interrelationships; this is referred to as a conceptualization [. . . .] It will include a vocabulary of terms and some specification of their meaning.*

This definition of Ontology by the W3C group is more understandable: *Ontologies are formalized vocabularies of terms, often covering a specific* *domain and shared by a community of users* (W3C OWL Working Group, 2012). The OWL Web Ontology Language developed by the W3C Web Ontology Working Group specifies the definitions of terms by describing their relationships with other terms in the ontology. Any OWL 2 ontology can also be viewed as an RDF graph. The primary exchange syntax for OWL 2 is RDF/XML.

FIGURE 1.10: The difference between taxonomy and ontology (Uschold, 2006) (Abreviations: Ctld , controlled; atts; attributes, Gnl; General, B/N; broader/narrower, Nrl lang def's; natural language definitions, w/ fml; without formal)

	Ctld. Vocab	Taxonomy	Thesaurus	Ontology
Definition	Defined terms, controlled	Controlled vocab. in a hierarchy.	Controlled vocab. in a network.	*specification of a conceptualization*
Notation	Free text, Definition structure varies.	Strict: tree Or: multi-parent	Broader/narrower (maybe taxonomy) Gnl. association;	Logics, Taxonomy as backbone + atts. & relations.
Meaning	Nrl lang def's Dictionary; common usage	Nrl lang def's + meaning of link Strictness & Precision varies. Isa, partOf, similarTo …	Nrl lang def's + meaning of links. B/N: various mng's Gnl Assoc'n: no specific meaning	Logics w/ fml. semantics. Isa hierarchy; Dom/Range constraints; cardinality. Nrl. language comments in the ontology.
Purpose	Human communi-cation (HC)	HC + Structure info. base; browsing	HC + Structure digital libraries; indexing, browsing & search	Union of all the others & more.

1.1.6.9 Difference between Taxonomy and Ontology

Following Garshol (2004), taxonomy means *a subject-based classification that arranges the terms of a controlled vocabulary into a hierarchy, based on essentially one relationship: the broader/narrower relationship used to build the hierarchy.* Ontologies, in computer sciences are a model for describing the world that consists of a set of types, properties, and relation-ship types. Gruber (1993) argues that Ontologies are often equated with taxonomic hierarchies of classes, class definitions, and the subsumption relation, but ontologies need not be limited to these forms.

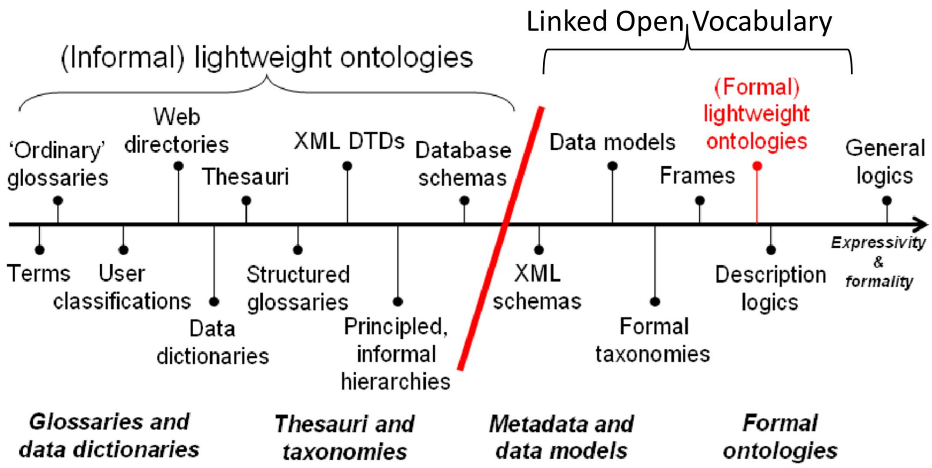
1.1.6.10 Lightweight ontologies

Lightweight ontologies are *typically defined as more hierarchical or classificatory in nature. Like their better-known cousins of taxonomies, but with greater connectedness, lightweight ontologies are often designed to represent subsumption or other relationships between concepts. They do not have many/ overly-complicated predicates (relationships)* (OSF Wiki 2014).

Lightweight ontologies could be considered as informal, basic ontologies consisting of backbone taxonomies only. (Giunchiglia and Zaihrayeu, 2009)
However, taxonomy, with multiple link types that each have a precise meaning, is also usually called an *ontology* (Uschold, 2006). The Fig.1.10 show an interesting comparison proposed by Uschold, in which the complexity increases from left to right and from top to bottom.

FIGURE 1.11: Spectrum of informal lightweight ontologies and Linked Open Vocabulary, adapted from Giunchiglia and Zaihrayeu (2009) and Uschold (2004)

.



1.1.6.11 Heterogeneity of ontologies

In the present text, ontology is used to refer to that of computer science. Ontologies can be viewed as metadata schemes that provide controlled vocabulary terms, each defined explicitly by a machine computerizable semantics (Chiaro and Damonte, 2005).

As stated by (Cardillo, 2015a) : *During the last ten years ontologies and the use of Semantic Web technologies has been seen as a better solution to semantic interoperability because this allows describing the semantics of information sources and makes its contents explicit by providing a shared comprehension of a given domain of knowledge [. . . . . . .] Unfortunately, ontologies and their structure are not really familiar and natural to most health care providers and their use raises* *heterogeneity problems to a higher level.*

The same author proposes to classify *Heterogeneity* into different categories:

- Syntactic Heterogeneity, which occurs when two ontologies are not expressed in the same ontology language;
- Terminological Heterogeneity, which occurs when different medical terms represent the same concept in different ontologies (e.g. the use of heart and cardiac to represent the same concept);
- Semantic Heterogeneity, which occurs whenever two contexts do not share the same interpretation of information (e.g. homonyms and synonyms) for different problems such as the difference in coverage, in granularity and perspective; and, finally,
- Semiotic Heterogeneity, which is caused by the subjective interpretation by humans of the used terms.

1.1.6.12 Linked Open Vocabularies (LOV)

Linked Open Vocabularies (LOV) could be considered as a kind of informal Light Weigh Ontology (see Fig. 1.11). The definition of the term Vocabulary in LOV is not linguistic but related to its use in the Semantic Web. Vocabulary is basically synonymous of ontology. The term of informal and formal refers to the degree of precision of the terms. Formal language is a language with mathematically precise construction rules applicable by computers. However, vocabulary differentiates from ontology by characteristics enabling reuse and integration by other vocabularies; small size, low formal constraints, few instances except for examples, rich user documentation (Labels, comments, definition, description, etc.) (Vandenbussche and Vatant, 2011). Linked Open Vocabularies (LOV) gathers definitions of a set of classes and properties (together simply called terms of the vocabulary), useful to describe specific types of things, or things in a given domain or industry, or things at large but for a specific usage. The definitions of terms provided by the vocabularies bring clear semantics to descriptions and links, thanks to the formal language they use (some dialect of RDF such as RDFS or OWL). In short, vocabularies provide the semantic glue enabling Data to become meaningful Data (LOV, 2017).

The Fig.1.6, page 20 proposes an attempt to understand the complexity of the constituent of an ontology. A semasiological set of entries (i.e. entries organized along meaning) are gathered in a on-line repository. Each entry has a Uniform Resource Identifiers, called URI, and is linked by an extended markup language (XML) called Resources Description Frameworks (RDF).

The building blocks of an ontology are triplets. A triplet is a set comprising a subject, a predicate and an object. (For example: Charles is_a man) as

FIGURE 1.12: The Triplet (design; R.Vander Stichele)

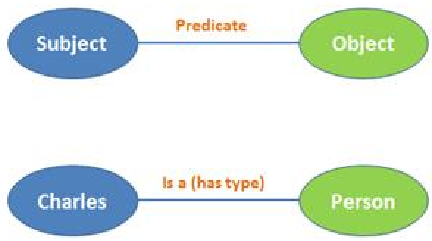
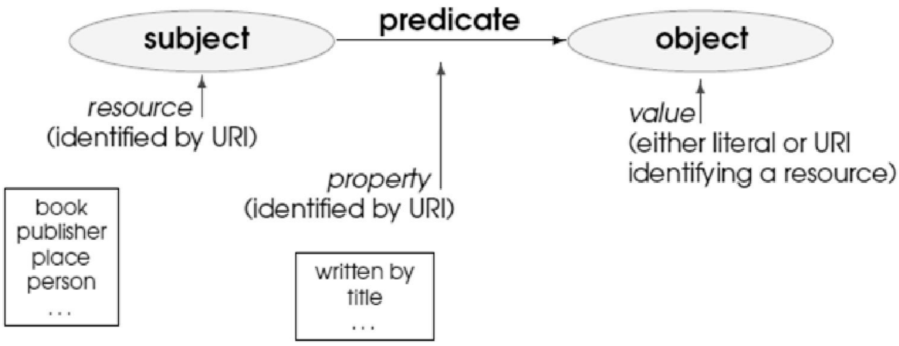


FIGURE 1.13: Example of a RDF Triple; the subject is a resource, the relation a property and the subject a value (Courtesy E.Cardillo) (Cardillo, 2011)



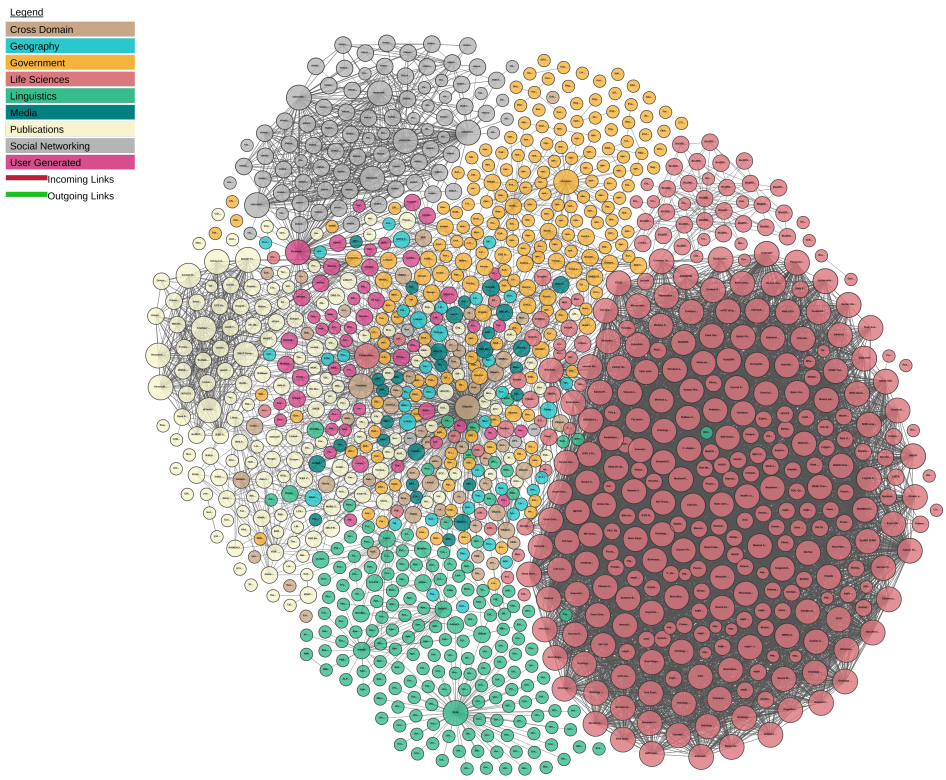
shown in Fig. 1.12. Possible links are infinite, and normalized terms could be linked semantically (literally "by their meaning") to codes or ID, or to other terminologies or definitions as shown in Fig. 1.13. The fig. 1.14 shows the RDF expression of the concept *Feeling tired* in the Ontology of Emotion. (Hastings et al., 2017). These points are highlighted in Chapter 6, page 36.

FIGURE 1.14: The concept *feeling tired* (normalized text underlined) in the Ontology of emotion in RDF in Obofundry http://www.obofoundry.org/

```
<!-- http://purl.obolibrary.org/obo/MFOEM_000080 -->

    <owl:Class rdf:about="http://purl.obolibrary.org/obo/MFOEM_000080">
        <rdfs:subClassOf
rdf:resource="http://purl.obolibrary.org/obo/MFOEM_000006"/>
        <obo:IAO_0000115>The subjective emotional feeling of tiredness,
needing sleep.</obo:IAO_0000115>
        <obo:MFOEM_000165>tired</obo:MFOEM_000165>
        <rdfs:label>feeling tired</rdfs:label>
    </owl:Class>
```

FIGURE 1.15: The Open Linked data network (Abele et al., 2015). Circles are web resources in specific fields (colored). (http://lod-cloud.net/)



1.1.7 A new format for knowledge management; The Semantic Web

Since the seminal proposal of Berners-Lee, Hendler, and Lassila (2001) to turn the Internet of documents into an Internet of data, giant steps have been made by numerous researchers. Simultaneously, the power of the laptops/computers has grown exponentially. Gradually, the health knowledge management field has become a multidisciplinary case, involving domain specialists, terminologists, taxonomists computer scientists and computational linguists. The technology of the Semantic Web and Linked Data have emerged as a future solution to exchange data distributed between multiple providers and family physicians around the globe.

The Semantic Web, sometimes referred to as Web 3.0 (Giustini, 2007), demonstrates the multiple possibilities of distributed data and how it has exploded in various field, such as: clinical guidelines (Kumar et al., 2004), information resources for consumer health (Smith and Fellbaum, 2004), mappings between medical classification systems (Cardillo et al., 2008), Linked data (Bizer, Heath, and Berners-Lee, 2009), ICD-11 development

(Tudorache et al., 2010), medical education (Blaum et al., 2013), terminological resources and Natural Language Processing (Neveol, Zeng, and Bodenreider, 2006), NLP and ontology (Liu, Hogan, and Crowley, 2011) (Liu et al. 2011), biomedical data integration (Smith et al., 2007a), interoperability (Qamar, 2008).

The Semantic Web relies on domain modeling. Numerous terms have emerged in reference to the concept of a domain model, such as: Semantic Network, Ontology, Concept Map, Conceptual Graph, Taxonomy, etc. (Clark et al., 2012). The medical domain expert does not need to have mastered all of these technologies, but he or she must, at least, understand the underlying knowledge that would facilitate a discussion with computer scientists and computational linguists.

This list of acronyms, taken from Liyanage et al. (2013) is key to the understanding of the domain:

- RDF (Resource Description Framework): A family of World Wide Web Consortium specifications originally designed as a metadata data model
- OWL (Web Ontology Language): A family of knowledge representation languages for authoring ontologies
- SPARQL (SPARQL Protocol): An RDF query language that is a semantic query language for databases, able to retrieve and manipulate data stored in RDF format

This massive amount of distributed data written in RDF, as shown in the Fig.1.14, sometimes referred too as Linked data (see Fig.1.15), can be questioned by machines, trained to *reason* through the scope of various ontological information and answer complex questions. Such ontologies are interlinked and form a gigantic new web, relying on relationships between data. (Linked Open data) (Abele et al., 2015).

The considerable development of medical ontologies demonstrates the vitality of this field of discovery. It has become increasingly clear that advances in information technology -or ontologies- and new languages such as RDF (Allemang and Hendler, 2008) or SPARQL (Salvadores and Horridge, 2012) , are transforming the Internet into a huge distribution data base. It is associated with Natural Language Processing techniques (Ittoo and Bouma, 2013a), which will allow for strides in information management. It is hoped that this will become integral in information management systems of GP/FM. The understanding that that information is to general practice what technology is to specialized medicine (Van Dormael, 2001) may finally be within reach.

Data encoded in OWL or RDF ares hard to interpret for those who are not experts in these formalisms. There have been some research of building question-answering systems over linked data. Such a system would typically take as input a natural language question, e.g. what is the cause of AIDS, convert it into an appropriate query in a specific language and then query an appropriate ontology in the linked-data cloud. The query language is called Protocol and RDF Query Language (SPARQL) (Unger, Freitas, and Cimiano, 2014; Lopez et al., 2013; Walter, 2017).

1.1.7.1 The Uniform Resource Identifier (URI)

Another contribution of Sir Tim Berners Lee to humanity was the Uniform Resource Identifier (URI), *a compact sequence of characters that identifies an abstract or physical resource* (Berners-Lee, 1998; Berners-Lee, 2002). It is a string of characters used to identify a resource (Miller, 1998). A URI identifies a resource by either location, name, or both. In addition to identifying a web resource, a URI specifies the means of acting upon or obtaining the representation of it. The Uniform Resource Identifier (URI) has been the building block of ontologies. The URL is the Internet address of a specific document. The URI is the address of a specified data. For example, a person (first data) and their weight "80kg" (second data linked to the first). The second most important proposal is RDF (Resource Description Framework) and OWL (McGuinness and Harmelen, 2004). Hyper Text Mark-up Language (HTML), the Internet language, is familiar and is used for linking documents. Resource Description Framework (RDF) is the name of the language which allows the linking of data with other data. The Web Ontology Language (OWL) is based on RDF.(see FIG. 1.3.8.2 87

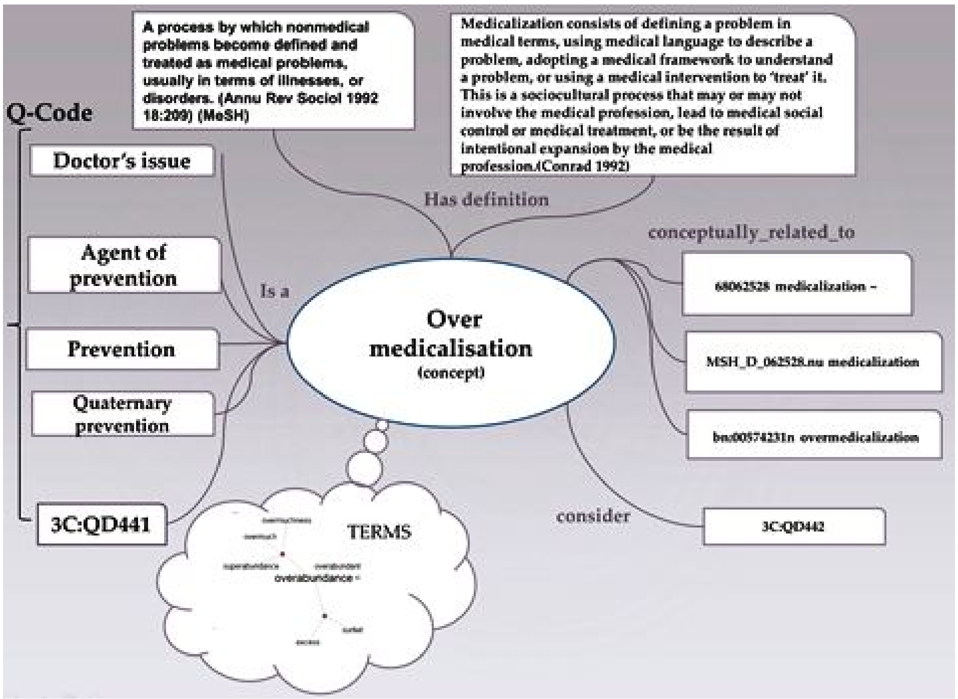
1.1.7.2 Data Structure Diagram

The Data Structure Diagram is a graphic technique, based on a type of notation dealing with classes of entities and the classes of their relationships (Bachman, 1969). In the Fig.1.16, the concept is central. Here *Overmedicalisation*, is linked by its relations (is a - consider - has a definition, conceptually_related_to) to other formally defined fields of knowledge. This kind of structure is machine readable and forms the basic structure of our taxonomy. It is presented in Excel format in the Fig.1.40, page 68.

1.1.7.3 Linked Open Data (LOD)

Examining the cloud of linked data on `http://linkeddata.org/`, one can see that billions of data-sets are now interlinked by RDF and then submitted to queries by dedicated robots (see the Fig. 1.15, page 33). Whatsoever the activity- business, cars, sales, book, civil affairs, management or health care - there are many sites already managing linked data. If one queries the linked data website of the BBC, perhaps asking for information about

FIGURE 1.16: Data structure diagram (DSD) of a Q-Code, showing the map of concepts and their relationships (conceptual data model)



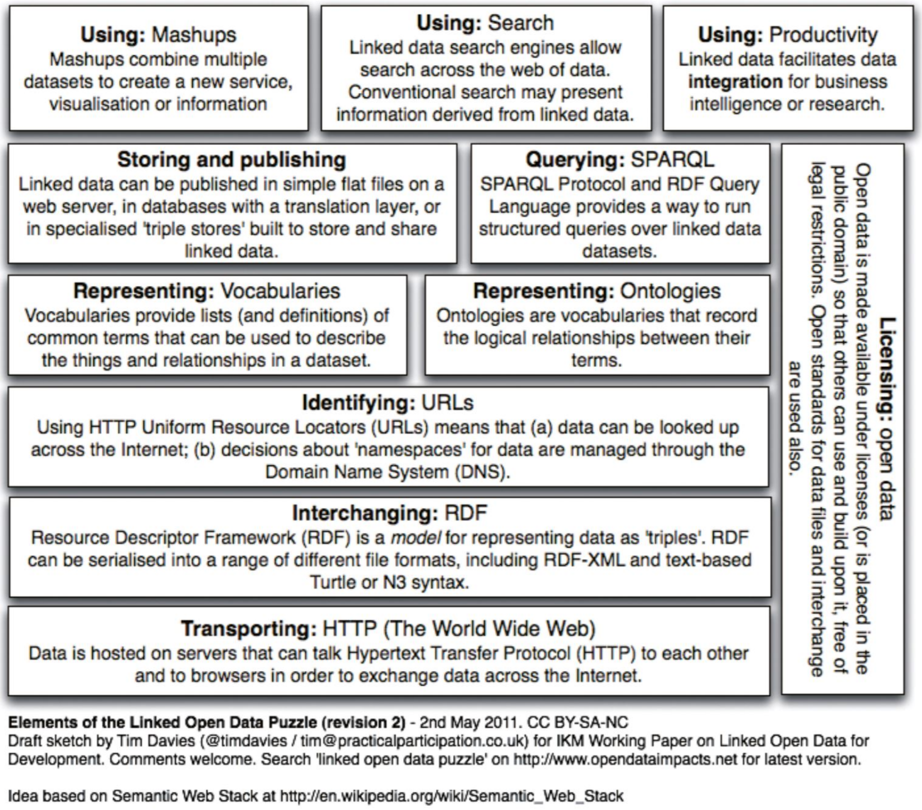
an artist, he or she will not receive a link to a prepared page. Rather, the page will be created instantly to match the query, using prepared semantic relationships. The screen is prepared instantaneously. This is achieved by copying data from the relevant cloud.

The computer scientist, however, doesn't say 'copied'. It uses the strange term *dereferencing data*. This allows for the retrieval of all knowledge needed on a subject, while the data is not gathered in a database, but dereferenced at a glance. The data is reachable through the Internet. Internet is now the database, and this is the explanation of the word "distributed data". The basic idea of a semantic web is to provide cost-efficient ways to publish information in distributed environments. For Bauer and Kaltenbock (2012), the term *data* refers to electronically-stored information or recordings, including but not limited to documents, databases, transcripts, and audio/visual recordings. Public data is data that is not subject to valid privacy, security or privilege limitations, as governed by other statutes.

The Fig. 1.17, shows the various component of the Linked Data universe. Following Bauer and Kaltenbock, 2012, they have to be; 1/ complete, 2/ primary, 3/ timely, 4/ accessible, 5/machine-processable, 6/ non-discriminatory access, 7/ non-proprietary format, 8/ license-free, 9/ permanent, 10/ at minimal cost.

As stated by Pierce (2014) Pierce et. all: *Semantic Web technologies offer*

FIGURE 1.17: Elements of the Open Data Puzzle (Bauer and Kaltenbock, 2012)



the potential to revolutionize management of health care data by increasing interoperability and reusability while reducing the need for redundant data collection and storage . This is the future of health care information systems and Electronic Health Records (EHR) and it is close at hand (Fernández-Breis et al., 2013). That's why scientists are working so hard to develop medical ontologies like Open Biomedical Ontologies consortium (OBO)(Smith et al., 2007b), the National Center for Biomedical Ontology (Musen et al., 2012), the Bioportal, which provides access to a library of biomedical ontologies and terminologies developed in Web Ontology Language (OWL) (Whetzel et al., 2011) and Linking Open Drug Data (LODD) for pharmaceutical research (Samwald et al., 2011).

1.1.7.4 Interlinked Publications in GP/FM world

In the same way, the abstracts presented by doctors in conferences, grey literature, could be tagged by semantic web specifications. Once these abstracts are uploaded into the local database of conference organizers or local organizations, a semantic web robot could traverse these sites, dereferencing the asked information through a common indexing system. This kind of system is already in use in clinical settings (Colliers et al., 2016) and one hopes to apply such techniques to an indexing system for the communication of family doctors in congresses and related grey literature. This could potentially interlink them through the use of a dedicated ontology in a semantic web, GP/FM oriented universe.

1.2 GENERAL AIM

1.2.1 A new taxonomy of contextual aspects of GM/FM fit for the semantic web

In 2008, Starfield (2008) argued that Health Services Research, by now a mature field, has *never developed precise terminology for most of the characteristics of health systems and services. Like Alice [in Wonderland], its practitioners use terms to mean what they want them to mean, no more, no less*. This is partly true, as shown in our study on terms used in guidelines (Jamoulle et al., 2014) (see chapter 5, page 147). However, the WONCA classification committee has already produced the Glossary of GP/FM (Bentzen, 1995) and its successor, the Dictionary of GP/FM (Bentzen N.(ed), 2003). The World Health Organization has authored its glossary of terms for community health care (Singh et al., 2007) and the Royal College glossary has been available since 1973 (RCGP, 1973). But, these tools were also expert-based views, not conceived as terminology, and only available in print.

The researchers in this field were confronted with the fact that indexing systems such as MeSH, are not multilingual, nor fully adapted to the particular field of GP/FM. Considering the absence of appropriate domain-specific terminologies or classification systems, this dissertation proposes a new multilingual indexing system. The existing International Classification of Primary Care (ICPC) is currently used for clinical purposes. In this thesis an attempt is made to create an extension of ICPC- named Q-Codes, a taxonomy related to professional contextual aspects such as education, research, practice organization, ethics or policy in GP/FM, which are currently not captured.

The set ICPC-2 + Q-Codes is proposed under the name Core Content Classification in General Practice (3CGP). ICPC-2 has been already adapted as a light weight ontology (Cardillo, 2011). The Q-Codes will also be developed as a lightweight ontology under the name of 3CGP, with more complex relations and properties.

As seen above, there are numerous classifications and terminologies for patient data retrieval. However, one must know whether they could be used to index documents and retrieve documents in a consistent way. Indeed, units of knowledge managed in historically different terminologies and classifications are interlinked and address the same reality seen by various eyes and interests (Bowker and Star, 1999).

The absence of adapted concepts and descriptors for contextual aspects of GP/FM is one of the reasons why the scientific work of family physicians is hard to retrieve from mainstream bibliographic systems. In addition, more than 50% of the scientific output of general practitioners at conferences is never published (Van Royen et al., 2010). There are no dedicated indexes of grey literature (Mahood, Eerd, and Irvin, 2014), and congress abstracts or collections of dissertation are often not properly indexed in this field (Lawrence et al., 2014).

The general aim of this project is to gain more insight into the complexity of the family doctor profession and to create a taxonomy in order to retrieve, show, and teach this complexity, using current computer based knowledge management techniques.

The practical relevance of this thesis in General Practice /Family Medicine and Primary Health Care is :

- To improve annotation of grey literature.
- To facilitate indexing of this reservoir of knowledge
- To improve the searchability of repositories for these information artifacts.

1.2.2 Specific Aims

The specific aims of the thesis are :

- To identify knowledge produced by GPs
- To provide a draft for a standardized table of content of GP/GM.
- To propose an experimental indexing system fit for GP/FM

1.2.2.1 Research questions pertaining to the identification of Knowledge gathered by GPs

- How identify the knowledge gathered by GPs?
- Can an ontology for GP/FM, integrated with MeSH, be devised based on concepts identified in communications of GPs ?
- Can the detection of these clinical and contextual terms from GP/FM be automated ?

1.2.2.2 Research question related to the creation of a standardized table of contents for GP/FM

Indexing the field of knowledge in GP/FM contribute to identifying the core conceptual and operational content and areas of expertise of GP/FM. Knowledge of GP/FM is hard to delineate, and so far, no attempt has been made to formalize or ontologize this domain. The aim is to efficiently formalize GP/FM knowledge, especially the contextual professional aspects, and this in extension of the clinical aspects, addressed by the International Classification of Primary Care (ICPC).

- What knowledge representation formalism must one use (a simple lexicon, a database, an ontology: lightweight or full-fledged) and how will one construct the chosen knowledge representation approach?
- Could this exercise contribute to define the table of subject of GP/FM?

1.2.2.3 Research question pertaining to the production of an experimental indexing system fit for GP/FM

- Is it possible to produce a new indexing system, based on the structure of the International Classification of Primary Care- second version (ICPC-2)? Could this be complemented by a new contextual classification called Q-Codes version 2.5, which is intended to facilitate the manual or automatic indexing of grey literature in GP/FM?
- Does the facilitation of online publication with a multilingual indexing system allow for visibility of GP/FM grey literature?
- Could this, if other universities and associations of GPs use the same system of indexing, produce a link between knowledge and networks of national and international doctors/researchers?

1.3 METHODS

1.3.1 Qualitative methods and METHONTOLOGY

To identify key concepts in a domain-specific taxonomy, data is analyzed in a grounded theory approach (Glaser and Strauss, 1999). This approach is often used in disciplines such as: economics, law, medicine and social sciences (Wells, 1995; Denzin and Lincoln, 2000). It involves the construction of a hypothesis or discovery of concepts through data analysis (Faggiolani, 2011; Martin and Turner, 2016) .

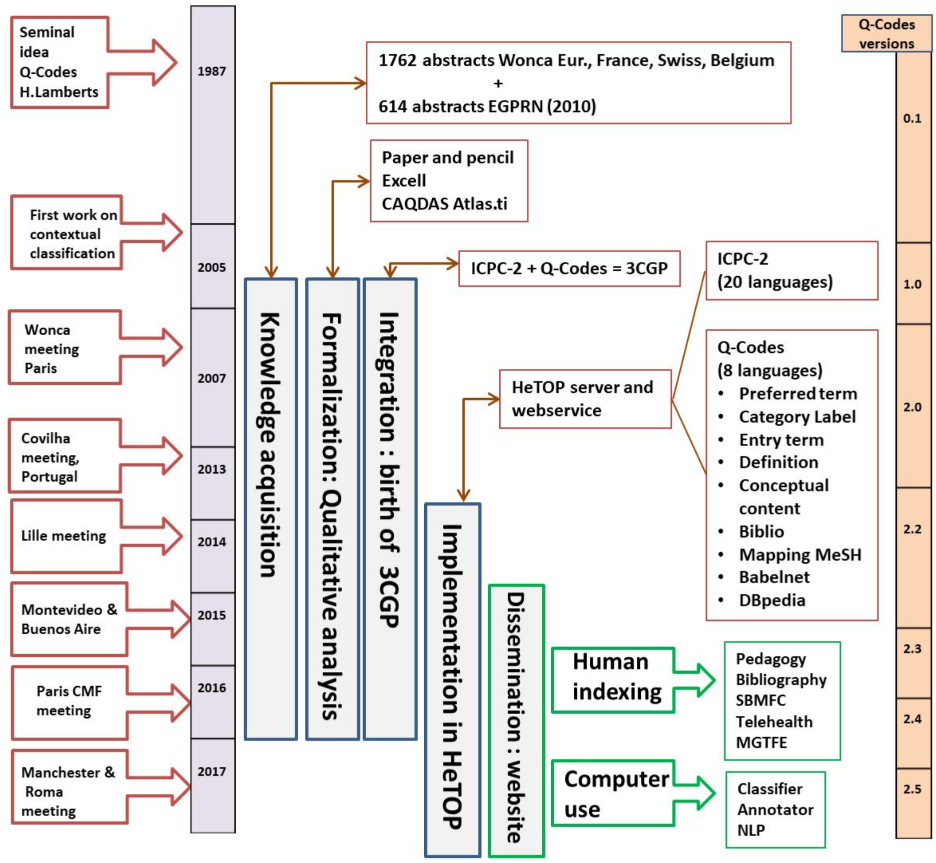
In addition, since the aim was to develop a domain-oriented taxonomy (the simplest form of an ontology, i.e. a lightweight ontology), methodology for ontology construction was included. We explored the methods proposed by Uschold and King (Uschold and King, 1995) and Gruninger and Fox (Grüninger and Fox, 1995) also KACTUS (Bernaras, Laresgoiti, and Corera, 1996), METHONTOLOGY (Fernandez, Gomez-Perez, and Juristo, 1997; Gómez-Pérez, Fernández-López, and Corcho, 2003), and SENSUS (Swartout et al., 1997) methodologies.

Each methodology was evaluated for its suitability to our task. Among the various potential methodologies, we chose METHONTOLOGY (Fernandez, Gomez-Perez, and Juristo, 1997; Fernández-López and Gómez-Pérez, 2002). This was chosen due to its: (i) a mature methodology; (ii) recommendation by the Foundation for Intelligent Physical Agents (FIPA) (http://fipa.org) for ontology/taxonomy construction tasks (Gómez-Pérez, Fernández-López, and Corcho, 2003) ; (iii) widely employment for constructing domain-specific ontologies in fields such as law (Corcho et al., 2005) (Corcho, Fernández-López et al. 2005), and chemicals (Fernandez-Lopez et al., 1999)

Furthermore, METHONTOLOGY is considered an effective and generally applicable method for ontology construction and validation (Fernandez-Lopez et al., 1999). It provides a user-friendly approach to knowledge acquisition, conceptualization, and formalization by non-knowledge engineers (López, 1999).

In addition, there is a clear, almost 1:1 correspondence between the steps of METHONTOLOGY and the steps of a grounded-theory analysis. In the grounded theory approach, the researcher works through a set of overlapping steps (Dick, 2014). Specifically, the note-taking step in grounded analysis corresponds to the knowledge acquisition phase of METHONTOLOGY. While coding, memoing, and sorting in grounded analysis corresponds to METHONTOLOGY's conceptualization (formalization) phase.

FIGURE 1.18: The phases of development of the project on a timeline. The four main phases of the METHONTOLOGY process and dissemination



The phases of development of the project are shown on Fig. 1.18 along the time line. Qualitative analysis of communications of GPs during congresses has induced the creation of a controlled vocabulary organized in a taxonomy. To develop a domain-oriented taxonomy (the simplest form of an ontology - i.e., a lightweight ontology), methodology for ontology construction was included (Gómez-Pérez, Fernández-López, and Corcho, 2003). The four main phases of the METHONTOLOGY process are shown Vertically : knowledge acquisition, formalization and integration were added in 2005. The implementation phase in the on-line HeTOP server began in 2014. We have added a dissemination phase through Internet and publications.The right column shows the versioning of Q-Codes.

1.3.2 Knowledge Acquisition Phase

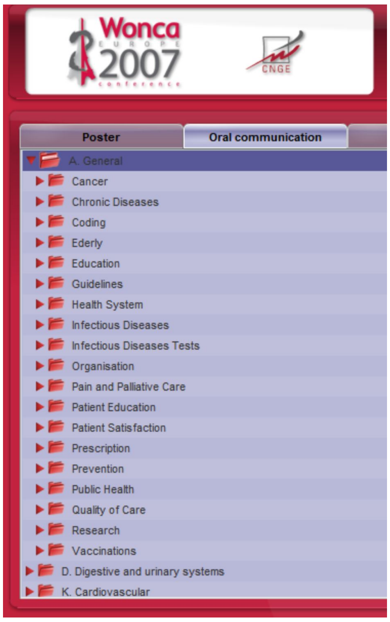
In 2005, after 25 years of work in the GP/FM classification field, and with some interest in Quality assurance (Roland, Prevot, and Jamoulle, 2002), the idea to analyze doctor's activities surged and not only patient problems. With so many abstracts in GP/FM congresses, the material reflecting GPs' activity was abundant. The choice of conferences analyzed (see Fig. 1.19) was guided by the availability of abstracts prior to the conference. This was to assure that the author was able to present and discuss the results at said conference with the audience. This kind of access is not easily granted, for which all involved in this project are incredibly thankful.

FIGURE 1.19: Sources of the abstracts analyzed

Year	Data Origin	Language(s)	Source	#
2007	Online. Access to reviewer pages before the congress	English	WONCA Europe conference	998
2013	Print, received during the congress	Portuguese	Portuguese 18th national conference of family medicine	128
2013	Excel file, abstracts received after the congress	French	Congres de la Confédération des Généralistes Enseignants (CNGE)	205
2014	Online. Access to reviewer pages before the congress	French	Congres de la Confédération des Généralistes Enseignants (CNGE)	289
2014	Print, published in Primary care, 2014	English, French, German	SwissFamilyDocs 2014	45
2014	Excel file	French	Belgian Congress 2014	37
2015	Print, published Family Practice 27(4) 459 :67	English	EGPRN Study	614

Examining the organization of medical conferences, one can see that there is no universally organized table of contents for GP/FM. From 2007 to 2010, French medical conferences in GP/FM have followed the ICPC distribution, using the letter A as a rag-bag for non clinical issues. At the WONCA Europe conference in 2007, Professor Bernard Gay, chairman of the scientific committee, had proposed to classify the abstracts with ICPC-2, which was achieved. The poster session was organized along ICPC-2 chapters. ICPC was well-suited for indexing clinical issues. But, the coding itself was, for lack of a better word, a mess. People were erroneously using the A chapter (General and unspecified) of ICPC-2 for all non-clinical issues. To be more precise, they were using their own subjective key-words. As a consequence, the list of A items was full of issues, and the results were absolutely meaningless, see fig 1.20). The 17 chapters of ICPC are also used as headings for the table of contents in the 2010 conference at http://videos.overcome.fr/mg2010/CDROM/index.html. (click "communications orales"). One can click on chapter A and see there the great diversity of the subjects which are incorrectly classified.This experience was not fruitful and was not repeated in the following conference.

FIGURE 1.20: WONCA 2007, CD-Rom of abstracts. Screen copy showing the use of ICPC chapters as ordering principle. The Chapter A is open and is used as a rag-bag. The list of non-clinical themes classed under the A chapter is impressive.



In the meantime, without forward-looking of the academic work this would become, the author began to read the abstracts presented by GPs from all over world who were present at the same conference. With the help of Anne Marie Magnier, Professor of General Practice and other colleagues, the author gained the chance to access the abstracts of the 2007 conference before its opening. This also ensured enough time to analyze the content of these abstracts as well. The careful analysis of 998 abstracts was presented at the same conference (Jamoulle and Dekeuster, 2007). The version of Q-Codes, proposed by H. Lamberts (ver.0.1) (see Fig. 1.21), evolved from 2005 and took shape at the 2007 Wonca congress (ver 1.0) (Reproduced in the Appendix A, page 191). The very unfavorable reception of the presentation led to the temporary burying of project. The consensus expressed by some attendants was that the MeSH was adequate and that a new classification, specific to GP/FM, was not necessary.

Five years later, invited by colleagues to the 2013 annual conference of Portuguese GPs in Covilha, Portugal, the author was struck by the richness of communication between these young GPs. There was a need to continue researching and analyzing the abstracts of this meeting. The analysis, performed during the night, was presented at the end of the same meeting, and the enthusiasm of the audience determined the fruition of this project. Conclusions have been published in the Journal of the Portuguese Association (Jamoulle, 2013) (see chapter 3, page 137). It was then that this project took shape and turned into more formal research. This was the starting point and catalyst for the realization of this thesis.

The French association of teachers in GP/FM (CNGE College National des Généralistes Enseignant - www.cnge.fr) annual conference was due to be held in Lille in 2014. This conference brings together French speaking teachers and researchers in the GP/FM field. With the help of colleagues from the French Société de Formation et de Thérapeutique Généraliste (SFTG - http://sftg.net/) and of Christophe Berckhout, professor of General Practice in Lille, the author was, once more, given the chance to gain access the to-be-presented abstracts beforehand. The analysis of accepted and refused abstracts has helped to develop the domain: Training and Teaching (QT).

Professor Laurent Letrilliart, head of General Practice at Lyon University, gave the idea to also analyze the abstracts of the 2013 CNGE conference. The purpose was twofold. First, show the focus of the meeting. Second, by indexing before the review process, make it possible to examine the contents of both rejected and accepted abstracts. This comparison would inform the conference organizers and participants themselves. The analysis of about 500 communications and their evolution between 2013 and 2014 have been presented at the same conference, which the audience regarded favorably (Jamoulle and Berkhout, 2014).

Soon after, the work was followed by the analysis of 45 abstracts of Swiss Family Docs 2014 and 37 abstracts of the Belgian research congress in general practice in 2014. Finally the analysis performed by EGPRN on 614 abstracts in 2010 has been used to control this work in the QR (Research) domain. (See Appendix D, page 207). Therefore, one can consider the current list of concepts found in GPs work as coming from 2316 different abstracts. It was not humanly possible to do more by hand.

The reader will, thus, understand why this research took 10 years and is still evolving. This research has transitioned from paper and pencil, to spreadsheet, and then to qualitative analysis software, while preparing the track for automated textual analysis and the semantic web.

1.3.3 Formalization Phase: Qualitative Analysis of Abstracts.

1.3.3.1 Reprocessing Lambert's Seminal Idea of Q-Codes

From 2005, initial Lambert's approach, (see1.1.3.4, page 16), has been filled with personal experience, along with 30 years of practice, research and teaching. With Lambert's first grid (see Fig.1.21) as an empirical document, we have attempted to fill in the gaps and modify the content of classification, using publications on GP/FM, pair experience, critics, and application to real practice. indexing exercises on published documents have also proven to be a sufficient way to verify the applicability of these classifications. An initial research proposal was presented to the WONCA International Classification Committee annual meeting in Dunedin, New Zealand (Jamoulle, 2005). First results of the preparatory work appeared, in which 8 domains had been designed. Three were originally in Lamberts' design. QO & Q1 of the Lamberts proposal are procedures that could be classified within ICPC processes. Q4 Personal functioning has been included by the researcher in QD8 work-life balance (see further). In 2007, the 8 domains remained the same, with slight changes in denomination and the addition of a rag-bag QO, "Other" (see further 1.3.5.1, page 63). For mnemonic reasons, the researchers have chosen to conserve the H of Hazard for the H of planetary health and to change the M of management in the S of structure in QS. The letter T for training and teaching remains to design the more wide Knowledge management for acoustic reasons.

FIGURE 1.21: Q-Codes; University of Amsterdam Dep. of gen practice. Prof.H. Lamberts. (circa 1987)

QO Care process (patient linked)
Q1 Care process (not patient linked)
Q2 Support task
Q3 Personal functioning
Q4 Patient's category
Q5 Research
Q6 Teaching

FIGURE 1.22: Evolution of the domains of Q-Codes from 2005 to 2007

	Domains of Q-codes	2005	2007
1.	Patient issues	QP patient	QP patient
2.	Provider's issues	QD doctor	QD doctor
3.	Structure of practice	QM management	QS structure
4.	Patient's categories	QC categories	QC categories
5.	Hazards	QH hazards	QH planetary health
6.	Ethics	QE ethics	QE ethics
7.	Training, teaching, editing	QT training	QT knowledge management
8.	R & D tools	QR research	QR research
			Q0 other

1.3.3.2 Using A Spreadsheet to Analyze WONCA Abstracts

In 2007, the author had the opportunity to access to the website of abstracts accepted by the reviewers of the scientific committee to the WONCA Europe Conference(http://www.woncaeurope.org/). As the abstracts were deposited, a digital copy has been taken and pasted it into an Excel database, with a unique ID. When the submission period was closed, a capture of 998 abstracts was available for coding purposes.

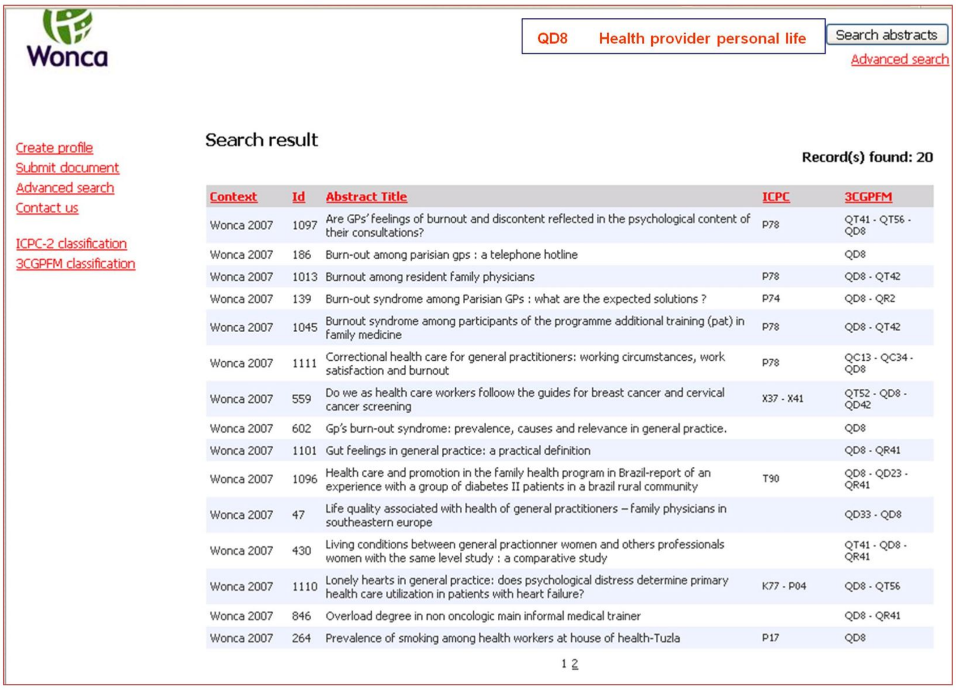
FIGURE 1.23: Extract of the Excel database of the Wonca 2007 conference, showing the titles of the abstracts and the coding with Q-Codes ver.1.0 and ICPC-2

ID	Title of the abstract	Q-Codes 1	Q-Codes 2	Q-Codes 3	ICPC-2 1	ICPC-2 2	ICPC-2 3
282	Evaluation of dispeptic and genitourinary symptoms during pregnancy and their relation with laboratory findings	QD33			W78	D87	U71
283	Depression among COPD patients in primary care	QD33			R95	P76	
284	Chronic back pain and depression, joint prevalence in primary care	QD33			L03	P76	
285	Why worry about a sleeping pill? First time benzodiazepine prescription in family practice: a qualitative study on patients' perspective.	QD31	QR41		P02	P18	
286	The lesser evil? Initiating a benzodiazepine prescription in general practice	QD31	QR41		P02	P18	
287	Use of proton pump inhibitors in nursing-home patients	QC14	QD32	QR21	D50		
288	A collaborative, multi-dimensional educational model:	QC22	QT11	QS21			
291	Euroaction: the achievement of blood pressure, lipid and glucose targets in a cardiovascular prevention programme in general practice – one year results	QD32	QD35	QR21	K50		
292	Suicidal ideation in depression: the role of comorbid generalized anxiety disorder	QD33			P76	P77	

Using ICPC-2 and Q-Codes version 1.0, the analysis went slowly. Several months later, a database, ready to be edited, had been created. With the help of Dominique Deceuster, a Belgian computer scientist, and financial support from the WONCA International Classification Committee, the designing and editing of an on-line query database for presentation during the WONCA world conference was possible. Unfortunately, after the presentation, the web domain and all the on-line material were lost due to lack of resources. Nevertheless, some screen copies remain to show how useful the tool could potentially be. Reproduced on fig. 1.24 is one slide of the PowerPoint presented at the WONCA Europe 2007 conference, during a workshop titled *Towards an e-archive for WONCA* (Jamoulle and Dekeuster, 2007).

Paper and pencil technology was also used five years later with the same tool of analysis for the abstracts of the Portuguese 18th National Conference of Family Medicine on September 28, 2013 (Jamoulle, 2013). The 128 abstracts, full of interesting work done by many young and enthusiastic GPs, have been indexed with Q-Codes version 1.0 and ICPC-2.

FIGURE 1.24: Screen copy of the WONCA Europe conference on-line, in which 998 abstracts were presented. The database shows 15 of the 20 results obtained with the query QD8 Health provider life. The right column shows the ICPC-2 and Q-Codes associated to QD8 in the abstracts. The acronym 3CGPFM was used erroneously for Q-Codes version 1.0



Context	Id	Abstract Title	ICPC	3CGPFM
Wonca 2007	1097	Are GPs' feelings of burnout and discontent reflected in the psychological content of their consultations?	P78	QT41 - QT56 - QD8
Wonca 2007	186	Burn-out among parisian gps : a telephone hotline		QD8
Wonca 2007	1013	Burnout among resident family physicians	P78	QD8 - QT42
Wonca 2007	139	Burn-out syndrome among Parisian GPs : what are the expected solutions ?	P74	QD8 - QR2
Wonca 2007	1045	Burnout syndrome among participants of the programme additional training (pat) in family medicine	P78	QD8 - QT42
Wonca 2007	1111	Correctional health care for general practitioners: working circumstances, work satisfaction and burnout	P78	QC13 - QC34 - QD8
Wonca 2007	559	Do we as health care workers followw the guides for breast cancer and cervical cancer screening	X37 - X41	QT52 - QD8 - QD42
Wonca 2007	602	Gp's burn-out syndrome: prevalence, causes and relevance in general practice.		QD8
Wonca 2007	1101	Gut feelings in general practice: a practical definition		QD8 - QR41
Wonca 2007	1096	Health care and promotion in the family health program in Brazil-report of an experience with a group of diabetes II patients in a brazil rural community	T90	QD8 - QD23 - QR41
Wonca 2007	47	Life quality associated with health of general practitioners – family physicians in southeastern europe		QD33 - QD8
Wonca 2007	430	Living conditions between general practionner women and others professionals women with the same level study : a comparative study		QT41 - QD8 - QR41
Wonca 2007	1110	Lonely hearts in general practice: does psychological distress determine primary health care utilization in patients with heart failure?	K77 - P04	QD8 - QT56
Wonca 2007	846	Overload degree in non oncologic main informal medical trainer		QD8 - QR41
Wonca 2007	264	Prevalence of smoking among health workers at house of health-Tuzla	P17	QD8

1.3.3.3 Qualitative methods used for developing the taxonomy

With grounded theory, sources of data are varied, but can include interviews, lectures, seminars, and expert group meetings (Ralph, Birks, and Chapman, 2014). The initial step of grounded theory is that of note-taking. It involves examining a chunk of data, such as a sentence, transcripts of focus groups, in depth interviews, or, in this case, abstracts of GP/FM conferences. The next step of grounded theory is that of coding. Coding begins with the constant comparison of data to data and attribution of a code to a sentence of the abstract. The information from this comparison is written, in the margins of the note-taking, as identified themes or variables (domain concepts). (See Figure 1.25 and 1.26).

In this research project, a 4 level framework with constraints at each level (max. 9) was used, effectively yielding a concept classification tree or a taxonomy. "Keeping it simple" has been the leitmotif. More levels would have provoked intractable situations only relevant to a heavy-weight ontology. On the other hand, it is necessary to obtain a certain level of granularity, but one must also know how to stop at the essential concepts; As

Bowker and Star (1999) says: *The decision not to collect is the most difficult to take for people maintaining any sort of collection based on a classification system.*

As proposed by Bradley, Curry, and Devers (2007), focus remains on strategies for analysis of qualitative data that are especially applicable in the generation of taxonomy, themes, and theory. Citing those authors, our experiment intends to:

- identify themes: *recurrent unifying concepts or statements about the subject of inquiry.*
- develop a taxonomy : *a formal system for classifying multifaceted, complex phenomena, to a set of common conceptual domains and dimensions.*
- build a theory : *a set of general propositions that help explain , predict, and interpret events or phenomena observed.*

1.3.3.4 Inductive or Deductive Content Analysis?

Within grounded theory, one can work with inductive or deductive content analysis or both simultaneoulsy (Elo and Kyngäs, 2008). The inductive process works from the specific to general, and it identifies categories using the bottom-up approach (Pope, Ziebland, and Mays, 2000; Bradley, Curry, and Devers, 2007; Soiferman, 2010). Although Bradley, Curry, and Devers (2007) states that it is recommended for grounded theorists to use only inductive analysis, others believe that both inductive and deductive methods can be used, proceeding from data to category and from category to data (McMillan, 2009).
In deductive content analysis, the researcher begins with preset aims (Pope, Ziebland, and Mays, 2000; Elo and Kyngäs, 2008) or themes (Bradley, Curry, and Devers, 2007). According to Bradley, Curry, and Devers (2007), the initial step is to define these themes/aims. Then, the data is reviewed, looking for patterns that fit into the pre-determined themes. In other words, moving from general to specific, using a top-down approach (Soiferman, 2010). Coding means *to bring similar data according to themes, concepts, etc. Generate code from the data level (inductively) or according to existing ideas (deductively) as necessary* (Silver and Lewins, 2014). This implies inductive capabilities (linking strings of text to existing codes) and deductive ones (looking for missing concepts in texts and attributing new code) are used.

In this research project, the deductive method was used in the beginning, as the project started from a historical predecessor. Then, the inductive method was used during the analysis of the abstracts. As soon as the restrained framework of the taxonomy was increasingly populated with concrete Q-Codes, the deductive method was used again. One can argue that the qualitative approach to the coding process is both inductive and deductive, an approach sometimes called abductive (Silver and Lewins, 2014).

1.3.3.5 Choice of the Computer-Assisted Qualitative Data Analysis Software (CAQDAS)

Following a suggestion from Dr. Frederic Ketterer, sociologist at the Liège University Department of General Practice, and after a careful study of existing products, the qualitative analysis software (ATLAS.ti® http://atlasti.com/) was used, as it enabled the required analyses to be executed at a relatively low cost.

FIGURE 1.25: Congress CNGE 2013: After being scanned, coded themes appear on the right column: Q-Codes (QS41, QC3, QP31) and one ICPC code (Z01). Here, the theme identified is the offer of family medicinal services in vulnerable populations. At the bottom left, the software proposes the list of pre-registered Q-Codes. (ATLAS.ti® Software)

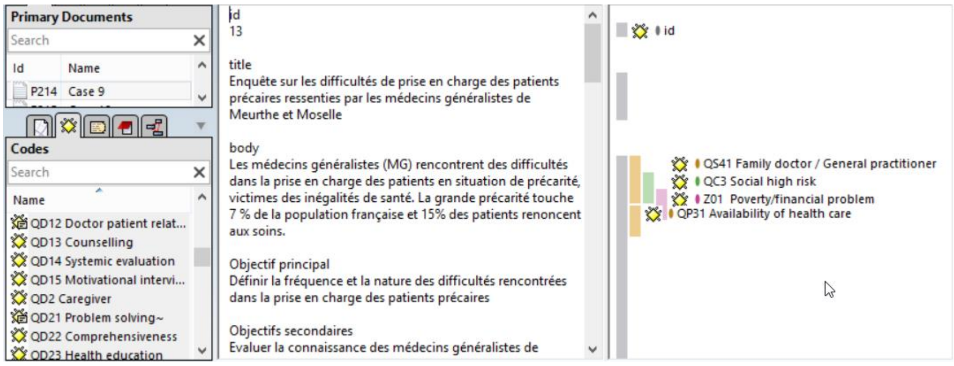
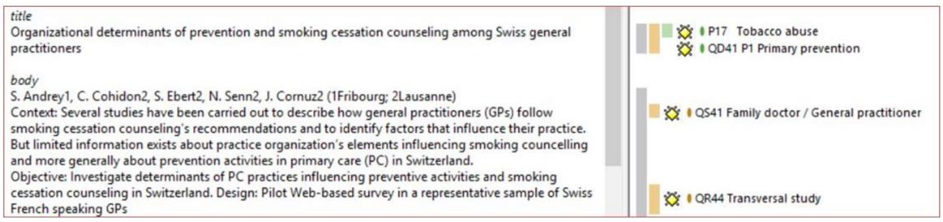


FIGURE 1.26: Example of the coding process of an abstract (Andrey, S. et al., 2014), using the software ATLAS-ti (coding by ICPC-2 and Q-Codes, version 2.3).



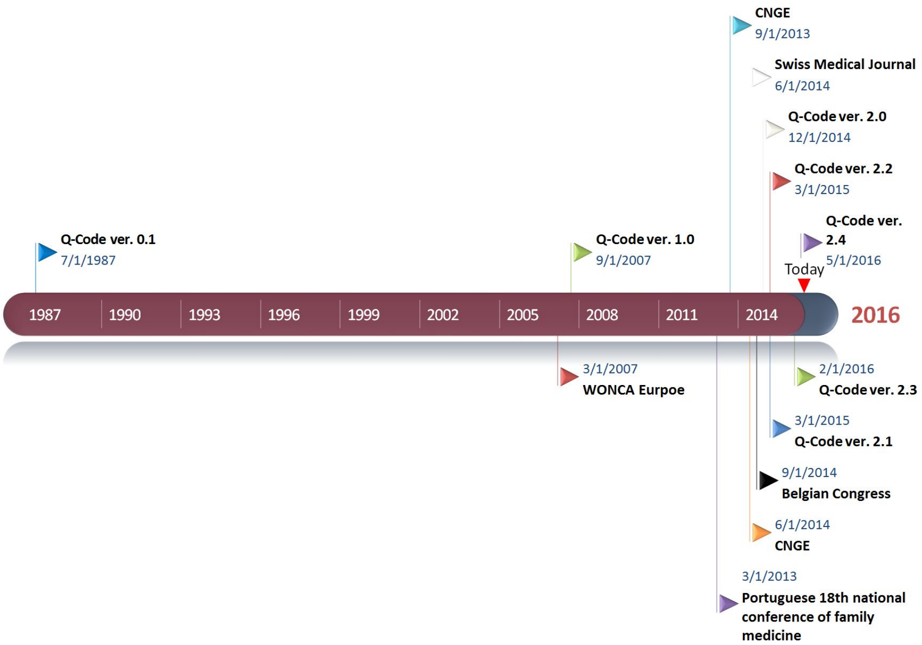
ATLAS.ti enabled the ability to map specific words to already-defined ICPC-2 and Q-Codes terms. Furthermore, the classification process enabled the discovery of new themes. (see also Fig. 1.31, page 56). The same theme could not reappear in the same abstract more than once, and (generally) no more than six themes were identified in each abstract.

1.3.3.6 Versioning

Hereunder, the successive versions of the works between 2005 and 2016 is presented.

- Version 0.1 (1987)
  Version 0.1 is comprised of Domains only. In transitioning from version 0.1 to version 1.0, all 7 Q-Codes (the letter Q followed by a number) from version 0.1 were deleted, and 153 new Q-Codes (the letter Q followed by a second letter) were added in order to create version 1.0.

FIGURE 1.27: Evolution of the Q-Codes between 1987 and 2016 along the time line (Design Melissa P.Resnick)



- Version 1.0 (2007)
  Version 1.0 contains 9 domains. Domains QD and QR each contain 8 Categories. Category QT5 contains 8 Sub-Categories, which allows for the addition of only one more additional Sub-Category before the maximum number of unique identifiers is reached.

- Transitioning from Version 1.0 to Version 2.0
  In transitioning from Version 1.0 to Version 2.0, 10 Q-Codes were deleted, and 53 Q-Codes were added, yielding a net increase of 43 Q-Codes.

- Version 2.0
  Version 2.0 contains 9 domains. Domains QD and QR each contain 8 Categories. Category QR4 contains 9 Sub-Categories, which is the maximum number of unique identifiers. There is one error in the assignment Q-Code Sub-Sub-Categories, where the numeration begins with "0" instead of "1": QD440.

FIGURE 1.28: Analysis of the abstracts along the time line from 2007 to 2016) (Design Melissa P.Resnick)

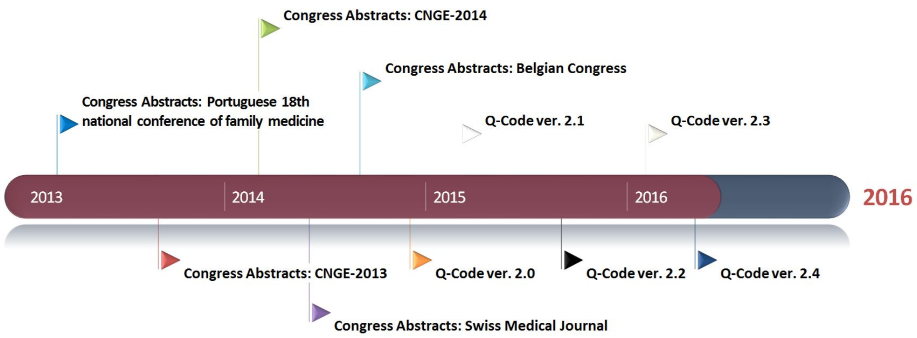


FIGURE 1.29: Evolution of the number of Domain, categories and subcategories of the Q-Codes from version 0.0 to version 2.4 (Design Melissa P.Resnick)

Q-Code Ver.	Circa	# Ragbag	# Domains	# Categories	# Sub-Categories	# Sub-Sub-Categories	Total Q-Codes
0.1	~1987	0	7	0	0	0	7
1.0	~2005	1	8	41	103	0	153
2.0	~2014	1	8	48	128	11	196
2.1	May 2015	1	8	48	118	12	187
2.2	Nov 2015	1	8	48	113	13	183
2.3	Jan 2016	1	8	48	112	15	184
2.4	Apr 2016	1	8	48	109	21	187

- Transitioning from Version 2.0 to Version 2.1
  In transitioning from Version 2.0 to Version 2.1, 27 Q-Codes were deleted, 12 Q-Codes were reassigned, and and 18 Q-Codes were added, yielding a net decrease of 9 Q-Codes.

- Version 2.1
  Version 2.1 contains 9 domains. Domain QD contains 8 Categories. Category QD2 contains 8 Sub-Categories. Sub-Category QD2 contains 9 Sub-Sub-Categories, which is the maximum number of unique identifiers.

- Transitioning from Version 2.1 to Version 2.2
  In transitioning from Version 2.1 to Version 2.2, 9 Q-Codes were deleted, 1 Q-Code was reassigned, and 5 Q-Codes were added, yielding a net decrease of 4 Q-Codes.

- Version 2.2
  Version 2.2 contains 9 domains. Domain QD contains 8 Categories. Categories QD2, QR4, and QT2 each contain 8 Sub-Categories. Sub-Category QD44 contains 10 Sub-Sub-Categories, which is more than the maximum number of unique identifiers.

FIGURE 1.30: Change in number of Q-Codes per Domain, over version number (version 03 was only numerical) (Design Melissa P.Resnick)

Ver	QC	QD	QE	QH	QP	QR	QS	QT	QO	Total
0.1	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	7
0.3*	27	23	5	6	27	7	12	9	0	116
1.0	23	29	8	6	28	13	15	30	1	153
2.0	26	42	8	6	28	24	24	29	9	196
2.1	26	41	8	6	22	23	23	29	9	187
2.2	27	41	8	6	23	23	21	29	5	183
2.3	27	42	8	6	23	23	21	29	5	184
2.4	28	42	8	6	23	25	21	29	5	187

- Transitioning from Version 2.2 to Version 2.3
  In transitioning from Version 2.2 to Version 2.3, 0 Q-Codes were deleted, 12 Q-Codes were reassigned, and 1 Q-Code was added, yielding a net increase of 1 Q-Code.

- Version 2.3
  Version 2.3 contains 9 domains. Domain QD contains 8 Categories. Categories QD2, QR4, and QT2 each contain 7 Sub-Categories. Sub-Categories QD32 and QD44 contains 6 Sub-Sub-Categories.

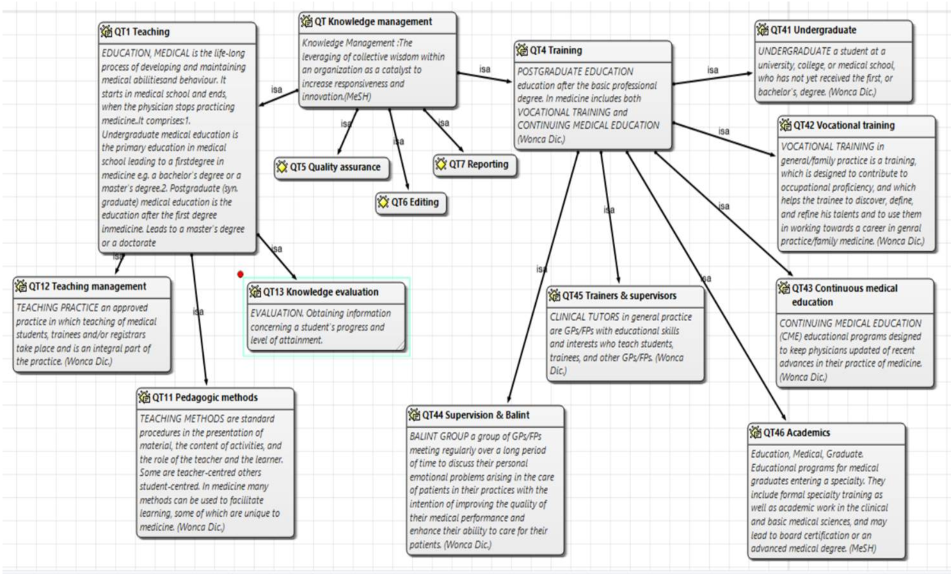
- Transitioning from Version 2.3 to Version 2.4
  In transitioning from Version 2.3 to Version 2.4, 8 Q-Codes were deleted, 7 Q-Codes were reassigned, and 11 Q-Code were added, yielding a net increase of 3 Q-Code.

- Version 2.4
  Version 2.4 contains 9 domains. Domain QD contains 8 Categories, which allows for the addition of only one additional Category before the maximum number of unique identifiers is reached. Category QD2, contains 7 Sub-Categories, which allows for the addition of only one additional Sub-Category before the maximum number of unique identifiers is reached. Sub-Categories QD32 and QD44 contains 6 Sub-Sub-Categories, which is more than the maximum number of unique identifiers. There are no errors in the assignment of Q-Codes.

- Version 2.5
  Some changes in the definitions, suggested by Vietnamese and Belgian colleagues, not the numbering of categories, has led to the Version 2.5 which is ongoing.

Note that comparison between data of the various conferences are not always possible due to the evolution of the numbering. The 8 main domains only have had no change since 2007.

FIGURE 1.31: Distribution of some categories of the QT domain with their definitions. Q-Codes ver. 2.3. Software Atlas.ti



At this stage, the steps of grounded theory are complete, along with the specification and conceptualization phase of METHONTOLOGY (Fernandez, Gomez-Perez, and Juristo, 1997), with the outcome being a description of a taxonomy. Two additional steps are required to complete the lightweight ontology (taxonomy) construction process according to METHONTOLOGY, namely: (i) Integration; and (ii) Implementation.

1.3.4 Integration phase ; born of the Core Content Classification in General Practice (3CGP)

The main proposal of this dissertation is to deliver a lightweight ontology integrating three levels of knowledge in GP/FM:

(i) Patient and doctor components- i.e., components 1 and 7 of ICPC-2, representing symptoms and diagnosis

(ii) Process components- i.e., the processes done or asked to perform by GPs, representing components 2 to 6 of ICPC-2, known under the name ICPC-2 Process.

(iii) Contextual professional components- i.e., the interrelated conditions in which something exists or occurs, represented by the Q-Codes.

These three combined taxonomies form a whole. It is hoped that this set allows for the meeting of the most frequently encountered and discussed situations by family physicians. The set has been designated as the Core

Content Classification of GP / FM under the acronym 3CGP. We can therefore express this set in the form: 3CGP = ICPC-2 + ICPC-2 Process + Q-Codes

1.3.4.1 ICPC-2

The International Classification of Primary Care (ICPC) (Okkes et al., 2000) can be considered a viable alternative to support the indexing of GP/FM content. This classification has been used, among others, for structured documentation of episode-oriented care in primary care (Lamberts and Hofmans-Okkes, 1996) , and for epidemiological research in sentinel practices (Lamberts, 1991). Interface terminologies like ICPC-Plus (Britt, Scahill et al. 1997) or more recently SNOMED-GP/FP RefSet (Jamoulle et al., 2015b) have been created to facilitate coding within ICPC. Over the years, extensions for nutritional advice (Binsbergen and Drenthen, 1999), procedures (Parkerson et al., 1996), medication (Mil, Brenninkmeijer, and Tromp, 1998), and chiropractic medicine (Charity et al., 2013) (Testern, Hestbæk, and French, 2015) have been developed.

However, ICPC only offers a partial solution, as it covers only the clinical issues of GP/FM (Verbeke et al., 2006; Boot and Meijman, 2010). Thus, it fails to capture the non-clinical professional issues, which are predominantly concerned with the organization and managerial aspects of GP/FM. Note, this does not address the personal context of the patient. To capture the personal context of the patient, tools like PeRI codes for patient ideas, concerns and expectations, are under development (Schrans et al., 2016)

ICPC has been progressively developed (Jamoulle, 1998; Soler et al., 2008). In 1987, the three existing classification systems: (i) Reason for Encounter (RFE) classification, (ii) ICPC-Process-PC, (iii) International Classification of Health Problems in Primary Care, defined (ICHPPC-2-d) were merged into a single system: the ICPC. ICPC-2 was published in 1998 by WONCA and the electronic version in 2000 (Okkes, Jamoulle, Lamberts, and Bentzen, 2000) with the International Classification of Diseases (ICD–10). ICPC has 3 components: symptoms, process and diagnosis. ICPC-2 is also available in more than 25 languages. It is compact, and the entire classification can fit onto two A4 pages (see Appendix B, page 195). The underlying principle used to develop the criteria in ICPC-2 was focused on *providing the most concise inclusion criteria possible, which would minimize variability in coding* (Jamoulle, 1998). ICPC has been developed to index frequent or heavy burden situations at the first level of care. Less frequent and rare disease are classified into categories ending with the digit 9 and crossed to ICD-10. The use of ICPC-2 for indexing less frequent and rare situations implies a link to ICD-10, which is not always available, see chapter 5, page 147.

FIGURE 1.32: Extract of ICPC-2e-v.6.0 (6 April 2017). Example of opening of 3 ICPC-2 process codes.

Code	Title	inclusion	exclusion	criteria	consider	note
-42	Electrical tracing	electrocardiogram; exercise electrocardiogram; Holter monitoring; electromyogram (EMG); electroencephalogram (EEG); electronystagmography (ENG)		a test used to measure the electrical activity of an organ (e.g. heart, nerve, brain, muscle)		
-44	Preventive immunization /medication	active/passive immunization; prophylactic treatment with drugs e.g. malaria prophylaxis; desensitisation		immunization/ desensitisation/ preventive medication used to avoid conditions/reactions, not present at the time of prescribing	-50 medication; contraceptive medication W11. If contraceptive medication is prescribed for medical reasons, use the code for medication -50.	If the patient has a condition (e.g. hypertension), use the code for medication -50.
-67	Referral to physician/ specialist/ clinic/hospital	Referral to another GP/FP/specialist; referral to disease-specific out-/in-patient clinics	referral to service for rehabilitation -68		-32 to -43 if ordering a specific procedure, which will be performed by another physician	

1.3.4.2 ICPC-2 Process, version 2016

In 2016 the WICC Process Group produced an extension of the ICPC-2 Process, which was not realized when publishing ICPC-2 in 1998. The Process Group has produced criteria, including and excluding criteria when appropriate for the 40 process classes pertaining to the component 2 to 6 of ICPC-Process. This work has been included in the latest release of the master file ICPC-2 version: ICPC-2e-v.6.0 (6 April 2017) on the web site of the Norwegian Directorate of e-Health , Department of Healthcare Classifications (https://ehelse.no/icpc-2e-english-version. The title of the rubrics of the ICPC-2 Process codes encompasses large domains, such as: imaging (code ICPC-2 41), electrical tracing (code ICPC-2 42) or referring to a hospital (code ICPC-2 67). These large domains are not very useful for indexing.

To allow the indexing of processes performed or requested by the provider, the items contained in the criteria, both inclusions and exclusions of texts elaborated by WICC, have been carefully analyzed. The procedures cited were identified and classified into a terminology one by one. They are offered in the form of a numbered list. In a first step, a sequential code opening in the format nn.00x is chosen, that is two digits for the procedure code followed by a period followed by a three-character sequential number (see Fig. 1.32). The development of this list has revealed some inconsistencies in the proposals of the Procedure group. These observations were transmitted and have been discussed at WICC meeting in August 2017. Meanwhile the list has been edited in a beta version (see Appendix C, page 203. The author is aware that this way of proceeding makes us cross a precipice - one that separates classification and terminology (Vanopstal, Vander Stichele et al. 2011). We go from a description of categories of concept (Fig. 1.32) to a list of terms organized sequentially (Fig. 1.33). This is done for two reasons. First, the aim is didactic, so that users of the coding system understand its content. Second, it provides a crucial step to the creation of a real ontology.

FIGURE 1.33: Same Process codes as in previous figure. The terms have been identified and code in a pick list format. Available in 4 languages (en-fr-es-pt) (See http://3cgp.docpatient.net/communications-publications/)

-42 Electrical tracing	42.000	electrical tracing
	42.001	electrocardiogram (ECG)
	42.002	exercise electrocardiogram
	42.003	electroencephalogram (EEG)
	42.004	electromyogram (EMG)
	42.005	electronystagmogram (ENG)
	42.006	Holter monitoring

-44 Preventive Immunization /Medication	44.000	preventive immunization
	44.001	preventive medication
	44.002	desensitization
	44.003	active or passive immunization
	44.004	malaria prophylaxis
	44.005	prophylactic treatment with drugs

-67 Referral to physician / specialist / clinic/ hospital	67.000	referral to physician
	67.001	referral to a specialist
	67.003	referral to a clinic
	67.004	referral to hospital

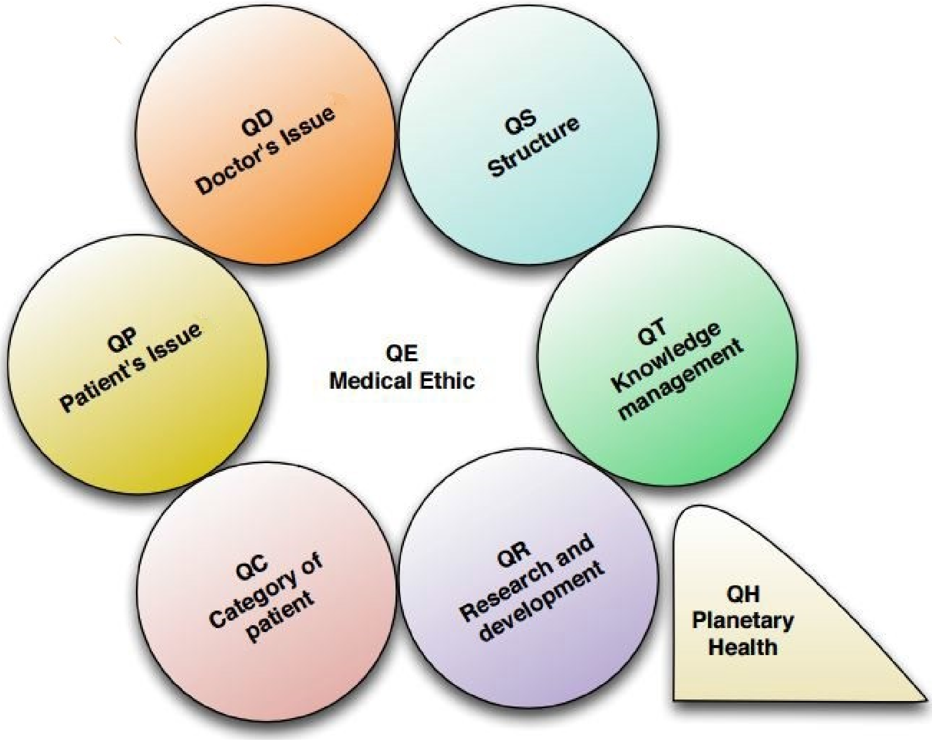
1.3.4.3 Q-Codes

As we have seen earlier, the Q-Codes taxonomy was elaborated on the model of ICPC, using the letter Q to categorize the contextual elements, for the letter Q was unemployed in ICPC-2. The study of the abstracts of French, Belgian and Swiss conferences have allowed augmenting the 2007 Q-Codes version 2 into a more furbished tool of 156 entries (version 2.2, 2015). Adding the domain QO, then comparing with the EGPRN 2010 study (see AppendixD, page 207) gave the version 2.3 which was edited on the HeTOP website and translated. The additions and corrections of the translators provoked a slight evolution of the tool, which has been numbered 2.5 in late 2016 and printed with this version number. But versioning applies formally only for printed versions, as the on-line database is continuously evolving. Corrections and additions, for instance of new MeSH 2016 (Jamoulle, 2016) and 2017, make this a living tool. Naturally, we hope that future conferences will allow for a surge in new concepts and new entries in the Q-Codes classification.

The Q-Code domains

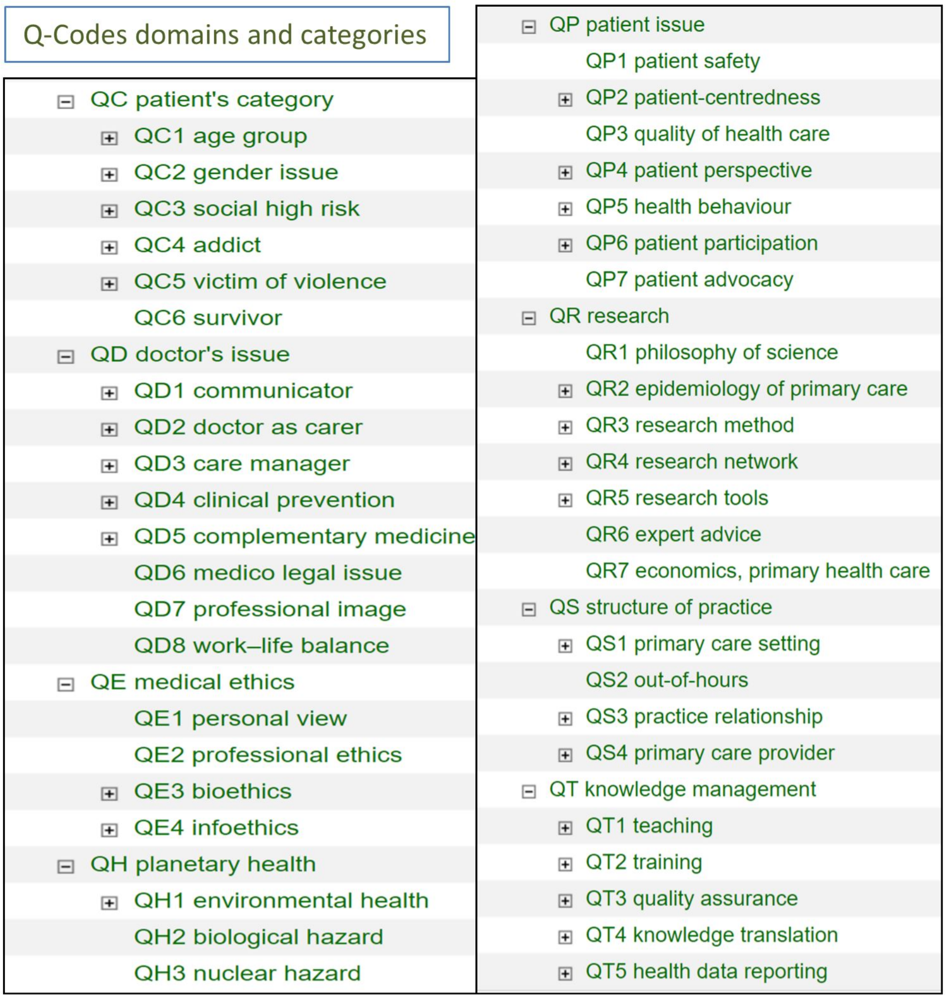
Concerning the Q-Codes' domains, the taxonomy starts with the QC domain, which represents "Patient's category", and covers topics such as age,

FIGURE 1.34: The Q-Codes matrix in the shape of a Q-Letter. Note that the Q's tail, which is the Planetary Health, prevents the wheel from turning endlessly. A very nice demonstration of the importance of the environment. (Graphic design Patrick Ouvrard)



gender issues, and victimhood. The second is the QD domain, representing *Family doctor's issue*, which covers issues such as disease management, communication, clinical prevention, and medico legal issues. QE represents *Medical Ethics*. This domain covers bioethics, professional ethics, and infoethics. The fourth domain is QH, representing *Planetary Health*, which deals with such areas as environmental health, biological hazards, and nuclear hazards. The next domain is QP, *Patient Issue*, which includes patient safety, patient centeredness, and quality of care. The QR domain is *Research*, covering research methods, research tools, and epidemiology of primary care. QS is the *Structure of Practice* Domain. This domain covers topics such as primary care settings, primary care providers, and practice relationships. Finally, the QT domain is *Knowledge management*. This domain deals with teaching, training, and knowledge dissemination. There are three clusters for the total number of Q-Codes (Categories, Sub-Categories, and Sub-Sub-Categories) within a given Domain. The low-outlier cluster consists of Domains QH and QE, which average a total of 7 Q-Codes within each of these Domains. The average cluster consists of Domains QS, QP, QR, QC, and QT, which averages a total of approximately 25 Q-Codes within each of these Domains. The high-outlier cluster consists of Domain QD, which contains

FIGURE 1.35: Q-Codes Eight domains, version 2.5, opening in categories. Copy screen of the hierarchy on the HeTOP server www.hetop.eu/Q



:;

42 Q-Codes within its Domain. The eight domains and the categories are edited at the Fig. 1.35.

The question of safety

For safety, we've chosen to distribute the concept over all the classifications when adequate, rather than creating a special category. Each conference was affirmed this thought and allowed for the addition of new elements. The fact that new concepts have emerged within Q-Codes has two reasons. First of all, the issue was addressed several times in conference abstracts. Secondly, the expertise in the field confirmed that the discussed issue was important. The general knowledge of the field helps to define the 8 broad domains of the coding scheme, which take an inductive approach with the data- i.e. the content of the abstract, guide him to identify, deductively, the categories and subcategories completing the coding scheme.

Intension and extension of a concept.
This approach is combined with the use of the intensional and extensional content of the categories. Intension, *the attributes of the individual object it denotes* (Sadegh-Zadeh, 2015), helps to specify the content of a category. While extension, "*the set of all object it applies to* (ibidem) gives the distribution of the concept. As an example, we can state that if an abstract deals with the concept of quaternary prevention and its definition, we will classify it in the category QD44 with the same label. We see here the MeSH approach, considering the item *quaternary prevention* as a topic. Instead, if the abstract deals with overdiagnosis, overscreening or overtreatment, which contains the ideas of both children and quaternary prevention, we will consider this as an extension of the concept and propose related subcategories in the coding scheme. As stated in the introduction, the *world of reference* (Eco, 1979) chosen has a deep influence over the hierarchy.

A matrix format
The presentation of the Q-codes under a matrix format is shown on Fig. 1.34. The matrix takes the shape of the letter Q, representing the 8 domains of the Q-Codes: On the left, the people related domains - Doctor's issue, Patient's issue and Category of patients; On the right, the managerial related domains - Structure, Knowledge management including Teaching and Training, and Research and development; Hazards are the underlying Planetary health conditions represented by the downward oblique tail stylized as a triangle but which are in reality the back-ground of the GPs work; in the center, joining all, Medical Ethics.

1.3.5 Overlap between ICPC-2 and Q-Codes

We see that ICPC and Q-Code could be in relation of synonymy (ex;-45 & QP51), meronymy (part of) (ex; -62 & QD6) or hyponymy (included in) (ex; P15 & QC4) By comparing them with Q-Codes, one can infer that ICPC-2 is a mixed classification, encompassing clinical patient related situations but also some professional management issues like A98 Health Maintenance and some processes like advice, referral, administrative and follow-up. In fact, those Q-Code contextual themes are statements from the doctor over activities related to the patient. Q-Codes are intended to identify thoughts of the doctor, what he has to manage, what he is discussing in texts, while ICPC-2 is intended to identify the problem experienced by the patient. On the other hand, ICPC-Process is related to the GP as manager. The overlap is unavoidable in certain cases. We exclude here the correspondences between diseases or symptoms semantically related to Q-Codes like P86 Anorexia & QP51 Patient alimentation.

In the coding process of abstract or texts, the kind of descriptor will depend upon the subject addressed. If the text is dealing with patient advice during a clinical contact, ICPC code -45 will be preferable. If the issue

FIGURE 1.36: Overlap of concepts between ICPC-2 and Q-Codes ver. 2.5.

ICPC-2	Q-Codes version 2.5
-45 Observe/Educate/Advice/Diet	QP51 Patient alimentation
-45 Observe/Educate/Advice/Diet	QD23 Health education
-46 Consultation with Primary Care Provider	QS33 coordination of care
-58 Therapeutic Counselling/Listening	QD13 therapeutic counselling
-62 Administrative Procedure	QD6 medico legal issue
-63 Follow-up Encounter Unspecified	QD25 continuity of care
-65 Encounter Initiated third person	QS31 practice collaboration
-66 Refer to Other Provider (EXCL. M.D.)	QS32 referral
-67 Referral to Physician/Specialist/ Clinic/Hospital	QS32 referral
-68 Other Referrals NEC	QS32 referral
A20 Euthanasia request/discussion	QD26 palliative care
A98 Health maintenance/prevention	QD41 primary prevention; QD42 secondary prevention; QD43 tertiary prevention; QD44 quaternary prevention
P09 Sexual preference concern	QC23 Sex difference; QC24 Transgender
P15 Chronic alcohol abuse	QC4 addict
P16 Acute alcohol abuse	QC4 addict
P17 Tobacco abuse	QC4 addict
P18 Medication abuse	QC41 prescribed drug addict; QC42 street drug addict
P19 Drug abuse	QC42 street drug addict
P29 (Burnout)	QD8 work–life balance
Z01 Poverty/financial problem	QP24 affordability
Z04 Social cultural problem	QP44 patient culture ; QP23 cultural competency
Z10 Health care system problem	QP21 accessibility
Z25 Assault/harmful event problem	QC54 victim of torture; QC51 gender-based violence

is managing how and when to advise, (alimentation for instance), patient Q-Code QP51 will be preferred. The context, clinical or managerial, will orient the choice despite that in some case both choices could be equivalent. In HeTOP, this kind of mapping could be indicated as *see also*.

1.3.5.1 Creation of a Rag-Bag; the Q-Code *Other*

The structure of what might be called an authoritative list of the major concepts encountered in the texts analyzed took shape slowly. Derived from Lamberts previous proposal, the categories have been divided into 8 major domains (see Fig. 1.34), opened in sub-domains to a maximum of four levels. It follows a parent-child organization. The ninth domain is called Other, and it aims to retrieve new potential concepts that can not otherwise be categorized.

To manage the triage of abstracts, a domain identified as *QO* was created as the rag-bag domain. When abstracts were identified with content that did not fit well into any existing domain, category or subcategory, they were placed provisionally into this rag-bag domain. Those QO codes are also available under URI format at page 89. When several abstracts, describing the same issue, appeared, they were assigned to the Q04 category, waiting for the creation of a new category presented also under a new URI format (see further in the paragraph 1.3.8.2). Four levels of QO were developed to assist in this process:

- QO1 was used for abstracts in which the themes and issues were unclear.

FIGURE 1.37: Acronyms, underlined in red in an abstract presented at the CNGE congress 2013. This abstract would be categorized into Q02 acronyms, not understandable.

SASPAS et **ETP**

Introduction : **L'ETP** fait partie intégrante de la loi **HPST**. Nous proposons déjà, aux **IMG**, des ateliers sur la posture éducative et les programmes **d'ETP**. Comment aller plus loin ?

Méthode : Une association médicale pluridisciplinaire **(AFRET)** s'est proposée pour être lieu de **SASPAS.** L'**AFRET** réalise des programmes **ETP** validés par **l'ARS** pour des patients atteints de : **DT2**, **HTA** et **BPCO**. **L'AFRET** est composée de **MSU**, maîtres de stages hospitaliers, **IDE** libérales et hospitalières, pharmaciens d'officine et hospitalier, kinésithérapeutes libéraux, diététiciennes libérales et hospitalières, psychologue, professeur en spot adapté. Tous ont reçu, ensemble, la formation initiale de 40h en **ETP**. Actuellement sont en cours des programmes **DT2** et dès septembre 2013 **BPCO**. Ce lieu de formation a été agréé par **l'ARS**. Les internes en **SASPAS** y effectueront un tiers temps et interviendront, du diagnostic éducatif aux ateliers **ETP**, jusqu'à la synthèse finale. Ils participeront à l'élaboration, la mise en place et la finalisation des programmes. Tous les 2 mois un étudiant en pharmacie sera présent au cours de ce stage, en accord avec la faculté de pharmacie.

Résultat escompté : découverte et création de programmes **ETP,** participation active aux ateliers, acquisition des compétences avec rétroaction par les **MSU.** Stimulation pour la formation initiale en **ETP** (40h).

Discussion : la formation en **ETP** au cours d'un **SASPAS** entre-t' elle dans les acquisitions d'un **IMG** ? Oui car l'approche centrée sur le patient, la relation et communication, l'éducation en santé sont 3 des compétences à acquérir. De plus la situation autour de patient chronique et la polymorbidité à forte prévalence (**DT2**, **HTA**, **BPCO**) font partie des grandes familles de situations cliniques.

Conclusion : Un stage **d'ETP** est essentiel pour acquérir cette compétence professionnelle. Cela se réalisera en **SASPAS** en pluridisciplinarité avec respect de l'autonomie de **l'IMG**.

- QO2 was used for abstracts that contained an acronym which was not understandable. Acronyms are used as signs of belonging. They give to their user the feeling of dealing with a particular tribe. Acronyms in a title of a communication are not acceptable as they presuppose the reader is of the same tribe, or understanding, as the author (see Fig. 1.37).
- QO3 was used for abstracts whose content was out of the scope of GP/FM. These include mainly hospital-based studies unrelated to GP-FM and incorrectly accepted at the GP/FM conferences.
- QO4 was used for abstracts until they were assigned a new domain, category, or subcategory

1.3.5.2 Influences of colleagues in the elaboration of the Q-Codes

- EURACT work
  The definitions published by EURACT Europe (http://www.euract.eu/), the WONCA group dedicated to teaching, have been carefully collected and reused in the QT domain.
- EGPRN work
  Kruschinski, Lange, Lionis, Weel, and Hummers-Pradier (2010) on behalf of the European General Practice Research Network (EGPRN) have studied the content of EGPRN conferences abstracts under the title *Themes and methods of research presented at European General Practice Research Network conferences*.

  The authors of the EGPRN 2010 study, stated that ; *The EGPRN study was performed as a part of the project on developing a WONCA/EGPRN* *research agenda. It analyzed the research themes covered within recent EGPRN conferences with a focus on study design and methods that were used. Abstract of past EGPRN conferences were classified on the basis of content and methodology by content analysis. This study revealed a broad range of research themes that had been addressed.* Quoting the WONCA 2007

FIGURE 1.38: Comparison and mapping between sixteen EGPRN 2010 research categories and Q-Codes version 2.2. Underlined in yellow the changes induced in Q-Codes

EGPRN Study. Research aspects	Q-Code	Remarks
6.4. Research methods	QR3	
6.4.1. Report for example the development of a guideline	QT32	Guideline is in the Q-Codes
6.4.2. Original study	_	this approach not considered in Q- Codes and could hardly be considered as it is a temporal judgment
6.4.2.1. Qualitative	QR31	
6.4.2.2. Quantitative	QR32	and subcategories of QR32
6.4.2.2.1. Intervention study	QR325	
6.4.2.2.1.1. with randomization	QR325	Included in QR325 this approach was not considered in Q-Codes ver2.2, included in entry term on HeTOP
6.4.2.2.1.2. without randomization	QR325	Included in QR325 this approach was not considered in Q-Codes ver2.2, included in entry term on HeTOP
6.4.2.2.2. Observational study.	QR321 to QR323	Observational is Implicit in the four following. This approach was not considered in Q-Codes ver2.2 included in entry term on HeTOP
6.4.2.2.2.1. cross-sectional surveys	QR323	
6.4.2.2.2.2. case-control studies	QR321	this approach was not considered in Q-Codes ver2.2
6.4.2.2.2.3. cohort studies (both prospective and retrospective)	QR322	
6.4.2.2.2.4. longitudinal (prospective data collection without control group)	QR324	
6.4.2.2.2.5. other, for example case reports	QR36	Assuming other means other quantitative. Only case report taken in account
6.4.2.3. Qualitative and quantitative	QR33	Qualitative and quantitative = Mixed study
6.4.2.4. Instrumental research. validation and reliability questionnaires	QR52	QR52 includes scale and questionnaire

e-archive proposal (Jamoulle and Dekeuster, 2007), they stated also that their results were similar to mine. Through the careful analysis of this work, the author has been able to more precisely develop the research domain (QR). The proposals of the EGPRN authors to classify the abstracts presented during EGPRN conference have been numbered (full study available in appendix D, page 207). In this publication 57 entries have been identified of which 21 are classed in categories and subcategories in the research domain QR (see fig 1.38)

This exercise has permitted to check the consistency of the Q-Codes proposal particularly in the QR domain as one has to be able to map mainly all EGPRN categories to Q-Codes as shown in the Fig. 1.38. Three modifications of the Q-R domain have been issued from this study and are presented in the Q-Code version 2.3. ; (i) Addition of randomization (with and without) as entry term in QR325 Intervention study; (ii) addition of QR321 Case control study; (iii) addition of the term observational in QR321 to QR323.

- Manifesto on Planetary Health
  The domain QH, initially titled Health Hazard, then Environmental Health and lastly Planetary Health was modified in 2016, after publication of the Lancet *Manifesto on Planetary Health* (Whitmee et al., 2015; Horton and Lo, 2015). Nevertheless, the abstracts that deal with environmental issues are rather infrequent in the congresses that we have analyzed.

- The Contribution of Translators
  The international network of family physicians and members of the World organization of family doctors (WONCA) are a fantastic group of dedicated volunteers. The head researcher worked both in French and English, with Mrs. Melissa Resnick, medical librarian and PhD candidate at University of Texas, Health Science Center at Houston, USA, checking for all language subtleties in English. Soon after, Dr Thành Liêm Võ from Phạm Ngọc Thạch University, Ho Chi Minh City, was keen to use Q-Codes in an e-learning program. In turn, he has begun to translate this into Vietnamese with a team of colleagues. But terms without context are difficult to translate. So, the Vietnamese colleagues have also translated all definitions attached to each rubric into Vietnamese, to ensure that contextual meaning is communicated (see further). This resulted in a healthy emulation between the volunteer translators of various backgrounds/nationalities. After a few months, the terms of the Q-Codes and their definitions were available in 10 languages: French, Dutch, English, Spanish, Portuguese, Turkish, Korean, Georgian, German & Vietnamese. The translation in Georgian is available since October 2017. German, Ukrainian, Greek and Italian translations are under construction. Two terminologists Melissa Ressnick (USA) and Arthur Treuherz, from PAHO / BIREME (Sao Paulo,Brazil) reviewed the translations for three of these languages. One terminologist evaluated the English translation, and the second one evaluated the Spanish and Portuguese translations. This process has led to minor modifications in term denomination and change in some definitions/ MeSH mappings (see further). The resulted version was named Q-Codes 2.5 when printed.

1.3.6 Implementation Phase; Creation of the terminology on the HeTOP server

The huge influence of semantic web technology has grown, and ontologies are taking precedence in terminologies for managing knowledge in the new era of distributed data (Cheung et al., 2009). Nevertheless, creating an ontology of GP/FM is not included within this scope of work. After gaining experience working in the field of comparison of terminologies (Jamoulle et al., 2014) and being aware of the future development of the semantic health data (Jamoulle et al., 2015b) (see chapter 5, page 147), it was concluded that the path of ontologies must be initiated. This can be done through the creation of a lightweight ontology.

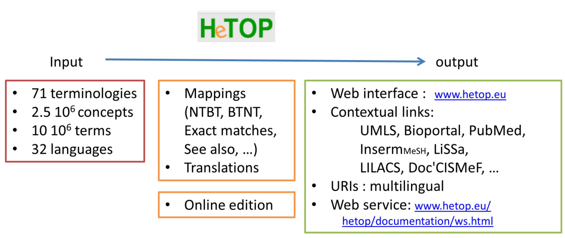
1.3.6.1 Seminal role of D2IM Rouen and the HeTOP interface

Contact with Professor Darmoni and his team have been facilitated by their membership to the CISP-Club (http://www.cispclub.org), the association of French users of ICPC . The welcome at the D2IM department was fantastic. The contribution of the Department of Medical Information and Informatics (D2IM) (http://www.chu-rouen.fr/cismef/d2im) of the University of Rouen, France, in the domain of ontologies editing is well known and substantial (Griffon et al., 2011; Grosjean et al., 2012; Merabti, Grosjean, and Darmoni, 2017). . It was possible to establish close collaboration with the members of this laboratory, which brings together experts from fields including family doctors, linguists, terminologists and computer scientists.

HeTOP is a multi-terminological portal that not only allows one to search for concepts in several terminologies (and several languages) at the same time, but represents their interoperability. (Friedman et al., 1999). Terminologies and ontologies are now widely used across the world for various applications. They are also generally specific to a domain (anatomy, rare diseases, medical devices, etc.). HeTOP is based on a multi-terminology meta-model that integrates all terminologies and ontologies into its data core. It is cross-lingual since terminologies and ontologies are often available in several languages.

The web site can be used by both humans and machines via a dedicated web service: www.hetop.eu/hetop/documentation/ws.html.

FIGURE 1.39: Terminologies are mapped in HeTOP and expressed in various formats



There was a need for a meta-model that is able to integrate any Knowledge Organizing System (KOS) and then create a corresponding server. The HeTOP server (for Health Terminology/Ontology Portal), in turn, was developed. The final goal of this server is to provide an accessible service on the Internet for both humans and machines. It now offers secure and personalized access to KOS in a suitable environment for intensive daily use by humans. In order to best meet the needs of the community, a Web Service provides the same functionalities as HeTOP for machines (Grosjean, 2014). The hetop.eu interactive web site offers cross lingual multi-terminological mappings on a semantic basis. Developed initially on MeSH and French mapping of MeSH for French speaking users, it has evolved towards a dynamic semantic interface, supporting a two-dimensional navigation across terminologies/ontologies and languages.

The HeTOP server currently contains 71 health terminologies and ontologies (only 17 are included in UMLS as most of them are French terminologies), 2,538,595 concepts, 9,982,113 terms, 10,120,417 relations and 32 managed languages. Overall, the number of distinct UMLS concepts, with at least one French translation in UMLS (MeSH, MedDRA, WHO-ART, ICPC), is 85,405 vs. 392,699 in HeTOP (Fig. 1.39). This crosslingual portal is dedicated to various usages by different types of users: translators, students, teachers, researchers, librarians, physicians, etc. HeTOP allows one to search and browse Health terminologies and ontologies in a second (Grosjean et al., 2012). Crosslinguality allows matrix navigation : among terminologies, but also among languages.

FIGURE 1.40: The 14 fields of each Q-Code in the HeTOP interface of which conceptual links are described in fig.1.16

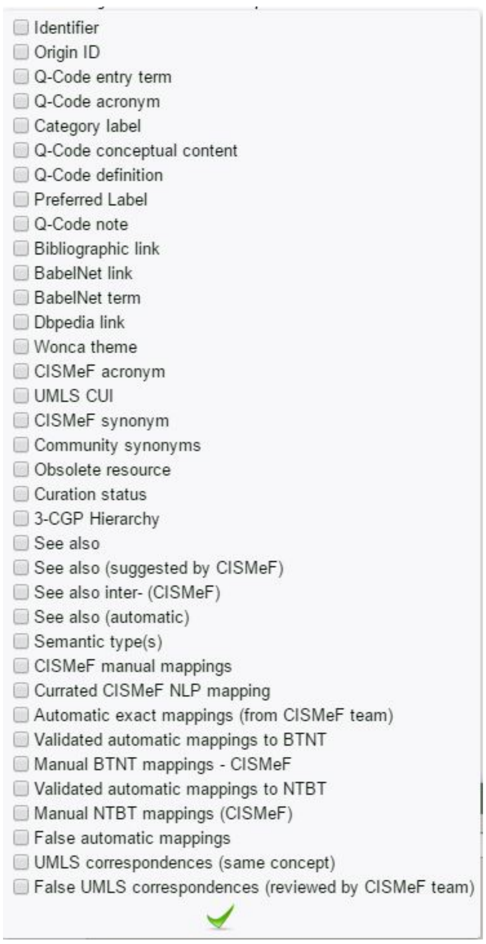
HeTop field	Signification	Remark
Category ID	Alphanumeric identifier of the Q-code	Letter Q followed by one letter (C,D,P,S,T,R,H,O) followed by maximum four numeric digits
Category label	Full title of the category	First letter capital, plural possible, can be compound words
Preferred term (PT)	Normalized title of the category	English, masculine, singular, lower case
Preferred term, other language	Translated (es, pt, nl, vn, tr, fr) version of the PT (more allowed)	Will be used as text word [TW] search term in PubMed equation (in English)
Supplementary entry term(s)	Additional search term(s)	Will be used as text word [TW] search term in PubMed equation(in English)
Category definition	Definition of the title of the category	Reflecting the corporate culture of GP/FM
Category conceptual content	Set of definitions or descriptions from online nomenclature, dictionaries or thesauri	Shows the extension of the concepts, includes the curated MeSH definitions
Automatic HeTop interterminologic relations	Automatic mapping of the PT by the HeTop embedded terminologies	Each proposal could be checked accepted or refused
Terminological features	Broader Than Narrower Term (BTNT) or Narrower Than Broader Term (NTBT)	Establish the hierarchical position of the PT mapping and links it to the HeTop semantic network
Curated MeSH	Accepted MeSH(s) which meaning correspond to the defined content of the PT	Will be used as [MH] in the PubMed equation automatically generated by the HeTop system
Refused links	Manually terminological links judged non-convenient	Will not be used in the search equation
Bibliographic free full text links	URL of citations of free full texts highlighting the content of the Q-Code	Generally chosen in PubMed but also through Google scholar
Babelnet.org link	Link to the URI of the corresponding babelnet.org entry(ies)	Map the Q-Code to an extensive semantic network of multilinguistic knowledge
DBpedia.org link	Link to the URL of the corresponding DBpedia entry(ies).	Map the Q-Code to a major knowledge semantic database. Could be Wikipedia if DBpedia missing

1.3.6.2 Publication of ICPC in Multiple Languages on HeTOP server

Thanks to contacts all around the world, the collaboration with D2IM team allowed for the completion of the on-line publication of ICPC-2 in 19 languages in 2015 (Schuers et al., 2015). Ukrainian and Greek have been added in 2017. Some colleagues went a step further and translated the HeTOP interface allowing health professionals to use it in their native language to access ICPC-2.

The database has recently been updated through the integration of ICPC-2 v6.0 (April 2017), which includes the last verified transcoding of the International Classification of Disease, tenth revision (ICD-10). This World Health Organization hierarchized nomenclature is also edited on HeTOP. In turn, the system allows automated, supervised or manual mappings with ICPC-2 and other terminologies, including MeSH for 20% of ICPC-Codes. Alignments between ICPC-2 rubrics and MeSH are automated and, unlike Q-Codes, have not been checked manually.

FIGURE 1.41: The fields of the HeTOP graphic editing interface can be chosen before to be exported.

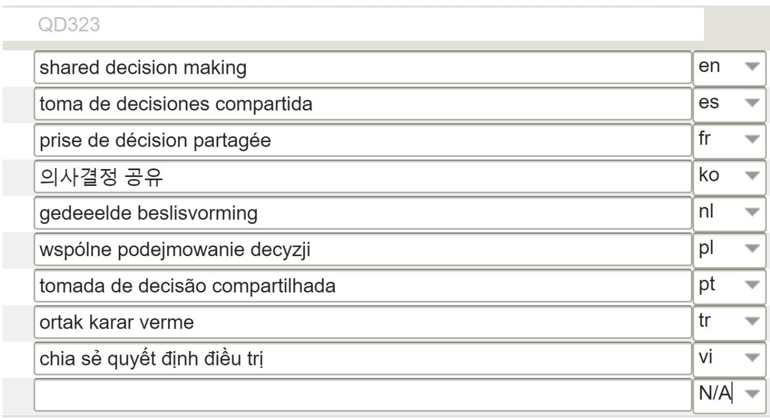


1.3.6.3 Developing the Q-Codes Terminology on HeTOP

Short after, Q-Codes took place as last born of terminologies on the HeTOP server. Adding Q-Codes to ICPC-2, it was then possible to develop a complete terminology adapted to GP/FM and PHC needs. The HeTOP server can be subdivided into fields and relationships as necessary. After careful analysis, 14 fields have been designed and are displayed in the fig.1.40 and their export facilities displayed in Fig. 1.41. The fields of this database are linked to each other along their respective semantic or definitional relationships, as described in fig. 1.16, page 36. Each field can be opened through multiple inputs. Each input can be expressed in any language. Each field can be exported into CSV, EXCEL or OWL format through dedicated web services as shown in th eFig.1.39.

Thus, the taxonomy of Q-Codes (concepts, properties and relations) was formalized and implemented onto HeTOP in Web Ontology Language (OWL). OWL is a relatively recent knowledge-representation language, based on the Resource Description Framework (RDF) standard, which is considered the de-facto language for ontology implementation (Antoniou and Harmelen, 2004) (McGuinness and Harmelen, 2004).

FIGURE 1.42: Detail of DBGUI, the editing interface of HeTOP. The entry field of the Preferred term *Shared decision making QD323* is shown here in 9 languages. Any additional language is allowed.

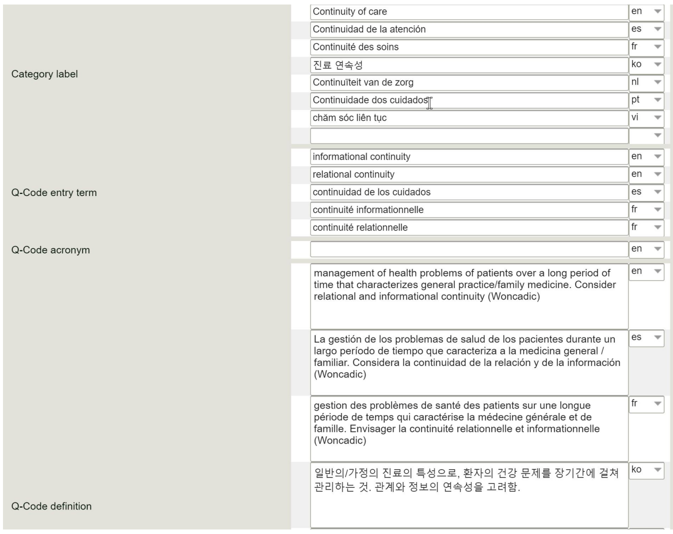


DBGUI, Database Graphic User Interface, allows one to edit sophisticated terminological records on the HeTOP server(Grosjean, 2014). A DBGUI user guide has been edited on the 3CGP web page (`http://3cgp.docpatient.net/tutorials/`). The screen copy presented at the fig 1.43 shows a detailed explanation of the DBGUI interface. The fields are opened along 'needed' input in languages' or along 'needed definitional fields'. The Q-Codes fields can be filled automatically, information can be exported from an Excel file for instance. This has been the default way in which the base was fed with translations of the Preferred Term and Definitions of Q-Codes. The various versions of ICPC-2 have been filled automatically. The

Q-Codes fields have been filled one by one manually. The content of most fields could be easily reviewed.

- Preferred term, category label, entry terms
  Each Q-Code has been identified by a Preferred Term and a Category Label. Preferred term (PF) is for machine use and has to be neutral or male singular and in lower case. Category label, often identical to Preferred term is for human use. Entry terms are synonyms or quasi synonyms. The chosen Preferred term, the added entry terms, the chosen MeSH, the entry terms of the chosen MeSH are used by the Inforoute a webservice of the HeTOP server to build a Pubmed search equation which gives immediate results in the Pubmed interface (Merabti, Lelong, and Darmoni, 2015).

FIGURE 1.43: Detail of DBGUI, the editing interface of HeTOP. The fields Category label, Q-Codes entry terms, and Q-Codes definition are displayed subdivided along the needed language.

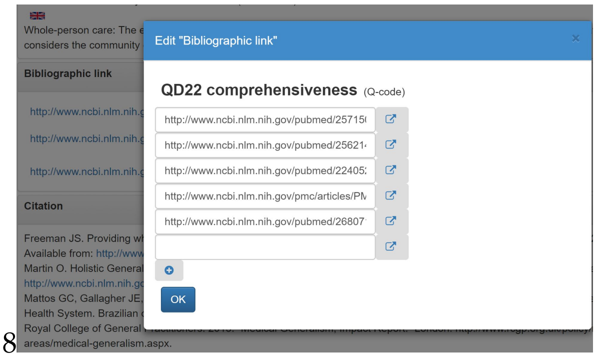


- Definition, conceptual content and translations
  Furthermore, a careful analysis of several available online terminologies and dictionaries (list available in Appendix G, page 223) has allowed for the conceptual content of a concept to be highlighted, including its various appellations, definitions or interpretations. This shows that concepts in medicine exhibit great variability, depending on the authors and their academic or cultural background. The general idea was to provide large semantic content for each concept. The most appropriate conceptual content has been chosen as a definition.

- Translation in numerous languages
  The Definition, the Preferred Term and the Category Label have been translated by volunteers into 10 languages at the beginning, more are

ongoing. The category labels and the definitions have been edited into 6 books, respectively in English, Spanish, French, Dutch, Portuguese and Vietnamese. (See the references of books in the published papers list, page xxxiii).

FIGURE 1.44: The updated graphical interface of HeTOP allow direct change. Here ready to receive a new URL of a bibliographic citation highlighting the concept at stake.



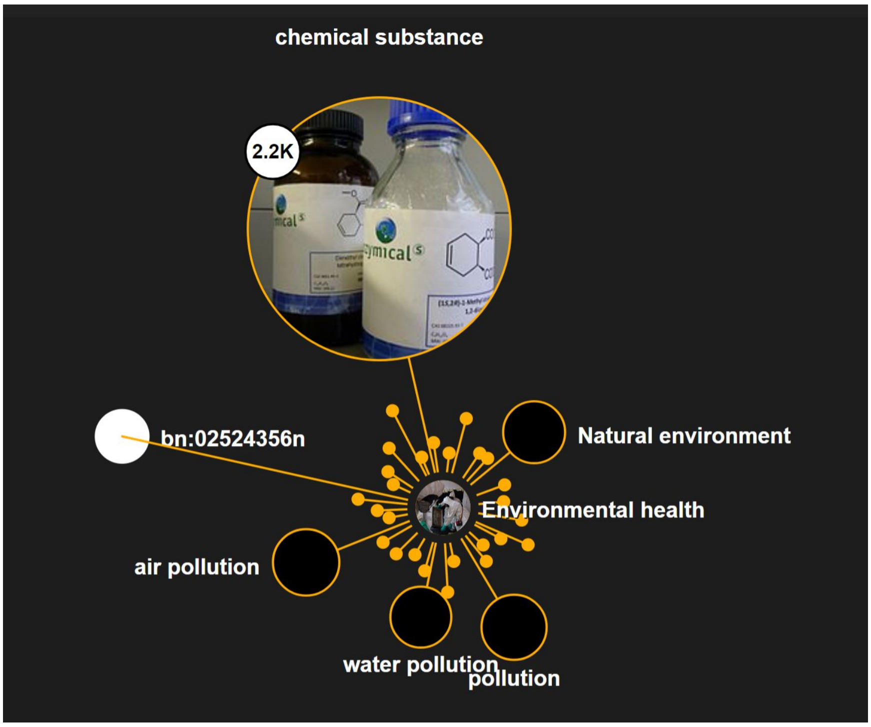
8

- Selected open access bibliography
  Three to five on-line citations of open access published papers were carefully chosen by the author to highlight the conceptual content of the related Q-Codes. The purpose is to highlight the corresponding concept by repositioning it within the context of use. The citations quoted are systematically and deliberately chosen in full free access. The intention behind this is to make it easier for the GPs and students to gain access to publications.

- Mapping to Babelnet & DBpedia
  Correspondences between the professional terms of the Q-Codes are also identified in the on-line semantic data base Babelnet (`http://babelnet.org`) and in DBpedia (`http://wiki.dbpedia.org`), the semantic web expression of Wikipedia. This emphasizes the fact that the family doctor is the natural interface between patient and health-care systems. Lexical knowledge repositories, like Babelnet (Navigli and Ponzetto, 2012) and DBpedia (Yamamoto, Yamaguchi, and Yonezawa, 2013) are essential components not only to human understanding of text, but to performing language-oriented automatic tasks.

  The Babelnet interface displays the semantic relationship between concepts in an exceptional interface (see fig. 1.45). Most terms in Babelnet and DBpedia could be considered as layman's terms or non-professional terms. It's the duty of the family doctor to interpret the professional terms into layman's terms for his/her patient. Preparing a cross mapping between layman's and professional terms is a necessity supported by numerous workers in the field (Vander Stichele et al., 1991; Cardillo, 2011; Cardillo, 2015b). And finally their Internet addresses are also configured into URI formats (see paragraph 1.3.8.2, page 87). This could be of interest in the future population of ontologies.

FIGURE 1.45: Graphics interface of the concept Environmental health on www.babelnet.org multilingual semantic network (partial view).



1.3.7 Mapping to Medical Subjects Headings (MeSH) and links to PubMed through Inforoute

1.3.7.1 The Inforoute facility on CISMeF

CISMeF ([French], acronym for Catalog and Index of French Language Health Resources on the Internet is a quality-controlled health gateway to catalog and index the most important and quality-controlled sources of institutional health information in French (Darmoni and Joubert, 2000) . The objective of CISMeF, which use the underlying HeTOP server, is to assist health professionals and consumers in their search for electronic information available on the Internet. (http://www.chu-rouen.fr/cismef/projet-cismef/english/) CISMeF uses two standard tools for organizing information: a controlled vocabulary which encapsulates the MeSH (Medical Subject Heading) thesaurus from the Medline bibliographic database and several element sets of metadata, including the Dublin Core metadata format. Currently (2017), the number of indexed resources totaled over 116.000.

To index resources, CISMeF uses four different concepts: meta-term, keyword, subheading, and resource type. *Meta-term* are set of descriptors which represent a domain. CISMeF Meta-terms correspond to medical specialties (e.g. cardiology), types of medical procedures (e.g. surgery) or health topics (e.g. diagnosis, therapy), which has semantic links with one or more MeSH terms and subheadings.
There are several types of MeSH to which the mapping is available in Inforoute facility.

- MeSH Descriptors (the most well known)
- MeSH Qualifiers
- MeSH publication types
- MeSH supplementary concepts
- MeSH Concepts,(which are linked to at least one of the other above)

They are over 30,000 MeSH concepts and contrarily to PubMed, CISMeF is using MeSH concepts to index its resources (Darmoni et al., 2012). The Inforoute algorithm has been adapted over several years by the D2IM team;
(i) MeSH synonyms have been systematically added to any query concerning one or more MeSH descriptor or supplementary concepts (Thirion, Robu, and Darmoni, 2009).
(ii) all the synonyms of the Unified Medical Language System (UMLS), sharing the same Concept Uniform Identifier (CUI) have been added. This

FIGURE 1.46: Q-Codes without MeSH mapping in the Medical subject Headings 2016

Q-Code ID	Category label
QD321	Medically unexplained symptom
QD323	Shared decision making
QD44	Quaternary prevention
QD442	Disease mongering
QD443	Overinformation
QD444	Overscreening
QD445	Overdiagnosis
QE1	Personal ethical view
QP25	Acceptability
QR2	Epidemiology of primary care
QR35	Action research
QR4	Research network

link to MeSH and through Inforoute infobutton is of first importance for students and undergraduates who are in training. Yet the selected bibliography opens for them new fields of knowledge. Now they can have, at a glance, a bibliography automatically prepared on a determined concept (see indexing for bibliographic searches (In French) on `http://3cgp.docpatient.net/`). Over two years, this work has been done through the editing interface DBGUI. MeSH mappings to Q-Codes have been available for all but ten concepts in 2016, enumerated in fig. 1.46 (Jamoulle, 2016). When no MeSH is available, Inforoute will use the Preferred Label of the Q-Code as well as the entry terms chosen for each Q-Codes marked with the suffix [TW] (text word) and perform a free text search in PubMed.

1.3.7.2 Mapping 182 Q-Codes to MeSH

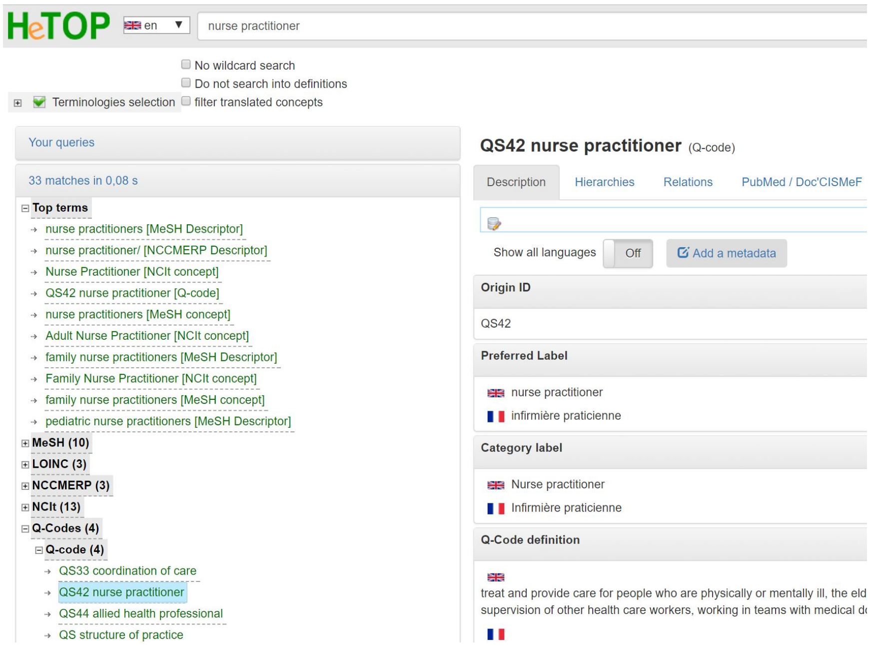
The creation of Q-Codes has raised some concerns about the possible mappings to the Medical subject Headings (MeSH). We gather here thoughts and observations over the correspondences between each Q-Code and the terminological base MeSH. In the HeTOP server, relevant MeSH terms - related to the Q-Codes, were selected and integrated both automatically and manually by the InfoRoute facility, an NLP - enhanced automated query system (Merabti, Lelong, and Darmoni, 2015; Lelong et al., 2016). The PubMed search profile built is relevant to each specific Q-Code. InfoRoute infobutton (see Fig.1.47) is a powerful tool for leveraging Medline querying based on semantic mappings (Thirion, Pereira, and Névéol, 2006).

- Main descriptors
  Automatic proposals of mapping between Q-Codes and MeSH for each of the 182 concepts identified in the on-line portal have been carefully studied (see an example at Fig. 1.48). They have been accepted, refused or modified (i.e. curated) in close collaboration with Melissa Resnick, medical librarian (Houston, USA) who has analyzed carefully and commented the content of each mapping before filtering usable proposals.

  The Inforoute automatic PubMed query builder presented at the Fig.1.47 will use the curated MeSH descriptors [MH] with the corresponding MeSH entry terms [TW] to build the query. Curated means here that the choice has been checked, accepted or rejected by the reviewer as shown in the Fig. 1.48. The curated [MH] will retrieve all the citations already indexed by the NLM Indexers. The results of the automatic PubMed query is shown in Fig. 1.49; The [MH], described as *MeSH descriptor* in the *currated mappings* in the Fig. 1.48 are underlined in yellow.

- Entry terms
  The [TW] terms i.e. terms in text words, will retrieve all the abstracts and titles which contain those free text terms, yet or no yet indexed. Entry terms of Q-Codes are carefully chosen synonyms in use in the medical field for each Q-Code preferred term. If any, the Q-Codes

FIGURE 1.47: The multilingual Inforoute info-button allows access to an automatic query on CISFMef and LiSSa (French knowledge bases) and on PubMed. A query builder is also available. Searched term; reason for encounter (RFE). Various occurrences of RFE in several terminologies are shown on the left. All entries are interactive (www.hetop.eu)



entry terms are added to the query as Text Word [TW]. This will augment the recall and lower the precision.

- NTBT & BTNT
  If accepted, it has been determined whether it is an exact mapping or whether the concept encompasses or subsumes the chosen Q-Code. If the proposal encompasses it, the proposal has been classified under the Narrower Than Broader Term (NTBT) label, meaning that the concept is broader than the chosen theme, encompassing a larger group of terms. If the chosen concept is a detail, or a subsumption of a Q-Code, it has been labeled Broader Than Narrower Term (BTNT). In the example shown in fig. 1.48, the Q-Code is *Nurse Practitioner*. The Nurse‘s role is a broader concept than Nurse Practitioner, so it has been classified NTBT. The Nurse’s patient-relationship concept is the offspring of the Nurse Practitioner concept. Therefore, it will be ranked BTNT.

  In the same example we see that HeTOP works on several terminologies. The Nurse Practitioner concept, consequently, is found in various terminologies, such as: (i) MeSH descriptor ; (ii) MeSH concept ; (iii) NCIt entry (National Cancer Institute terminology); (iv) SNOMED-CT (Systematized Nomenclature of Medicine - Clinical Term) ; (v) NCCMERP entry (National Coordinating Council for Medication

FIGURE 1.48: The various mappings to the concept Nurse practitioner are shown on the HeTop web service. see text.

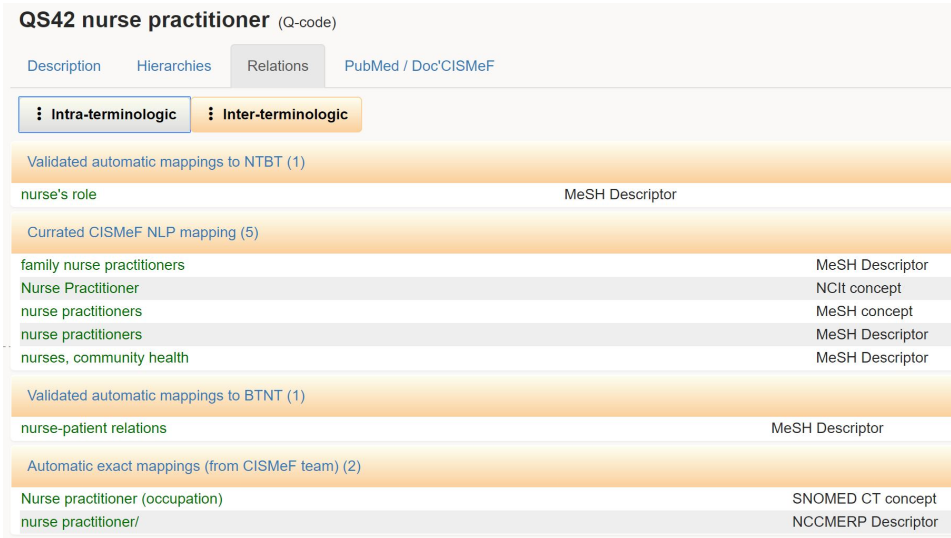


FIGURE 1.49: The PubMed search strategy as output of the Inforoute system is shown here for the Q-Code QS42 Nurse practitioner. The query, automatically build, uses the chosen MeSH (underlined in yellow) and their related [TW] and yields 23,256 citations on July 29, 2017.

(("family nurse practitioners"[MH] OR ("family nurse practitioners"[TW] OR "nurse practitioner, family"[TW] OR "nurse practitioners, family"[TW] OR "family nurse practitioner"[TW])) OR ("nurses, community health"[MH] OR ("nurses, community health"[TW] OR "nurse, visiting"[TW] OR "home nurse"[TW] OR "nurses, visiting"[TW] OR "nurses, home"[TW] OR "nurse, home health"[TW] OR "community health nurse"[TW] OR "nurse, community health"[TW] OR "health visitor"[TW] OR "community health nurses"[TW] OR "nurses, home health"[TW] OR "health visitors"[TW] OR "home health nurse"[TW] OR "home nurses"[TW] OR "nurse, home"[TW] OR "home health nurses"[TW] OR "visiting nurses"[TW] OR "visiting nurse"[TW])) OR ("nurse practitioners"[MH] OR ("nurse practitioners"[TW] OR "nurse practitioner"[TW] OR "practitioner, nurse"[TW] OR "practitioners, nurse"[TW])) OR ("nurse practitioner"[TW])) OR "family nurse practitioners" OR "nurses, community health" OR "nurse practitioners" OR "nurse practitioner"

Error Reporting and Prevention). Although these semantic subtleties are not immediately necessary for the functioning of the Q-Codes database, they will be indispensable in moving towards the ontology stage.

In the following paragraphs, we will address some concerns about MeSH mapping with Q-Codes. This will allow for us to understand the method used to align the Q-Codes concepts with the MeSH.

1.3.7.3 Primary Care and Family Medicine distinction in Q-Codes and MeSH

- The Secondary Care and Primary Care concepts in the MeSH.

The introduction in the National Center for Biotechnology Information (NCBI) on-line MeSH searching facility of the truncated term *hospital** retrieves 260 semantically associated results. The contrast with first line care is evident. One can find 20 MeSH related to first line primary care. Of those 5 are in reality hospital based care for outpatients (see Fig. 1.50). Primary care and general practice appeared more *supplemental* to the basic foundations of the MeSH thesaurus, designed for secondary care description (Mendis and Solangaarachchi, 2005)

The dates of introduction are also interesting. Although *Family Practice* was introduced in 1978, *General Practice*, as parent of *Family Practice* appeared only in 2011 as a *Health Occupation* as a child of the main MeSH *Medicine*. This implies that before 2011, those descriptors were not in use and that related citations are not available. There is no documentation to explain why those descriptors appeared in 2011 (source; O. Bodenreider, NLM – personal mail). It's worth noting that previous indexing of *Primary care*, before 1972, was *Comprehensive health care.* This is quite appealing from epistemological point of view. The MeSH *Ambulatory* is mainly used to deal with hospital issued care which are distributed outside the hospital. It's surprising that the MeSH *Management Information System's* includes *Ambulatory care information system* but not *Primary care information system*, a non-existing entry in the NLM thesaurus.

The MeSH *Outpatients* definition, with bold emphasis on ambulatory, is quite clear in this instance: *Persons who receive ambulatory care at an outpatient department or clinic without room and board being provided*. This formulation allows disambiguating the use of ambulatory as pertaining to the hospital domain and not to Primary care in the MeSH thesaurus.

Thus, we do use the two upper parts of the Fig. 1.50 to address respectively the Primary Health Care and Family Medicine issues. By using the tag [Mesh:noExp] we chose not to include MeSH terms found below the MeSH *Primary Health Care* in the MeSH hierarchy like *Continuity of Patient Care* or *Refusal to treat* which are attribute of GP/FM but not specific of the first level of care.

- The 8 PHC related MeSH.

  The MeSH quoted in the upper part of the Fig. 1.50 have been chosen as alignment in Inforoute for the Q-Code QS1, *Primary care setting*. The particularity of Inforoute is the addition of the respective MeSH entry terms with the suffix [TW] to the chosen MeSH. This enhances the recall but lowers the precision. (Jenuwine and Floyd, 2004) (See also Appendix E), page 209. This search strategy yields 473.427 citations on June 6, 2017, less than 0.02% of the total Medline database.

FIGURE 1.50: First line care related MeSH in PubMed

Types	MeSH	Year Introduced:
Primary Health care (PHC) related MeSH	“Community Health Centers”[MeSH:NoExp]	1979
	“Community Health Services” [MeSH:NoExp]	1967
	“Community Mental Health Centers”[MeSH:NoExp]	1979
	“Community Mental Health Services”[MeSH]	1967
	“Home Care Agencies”[MeSH]	1995
	“Home Care Services” [MeSH:NoExp]	1967
	“Primary Health care” [MeSH:NoExp]	1974
	“Rural Health Services” [MeSH:NoExp]	1996
General Practice : Family Medicine (GP/FM) related MeSH	“Physician, Family” [MeSH]	1974
	“Community medicine” [MeSH]	1977
	“Family Practice” [MeSH]	1978-2010
	“Gatekeeping” [MeSH]	2000
	“General practice” [MeSH]	2011
	“General practitioner” [MeSH]	2011
	“Physician, Primary care [MeSH]	2011
Hospital based care related MeSH	“Ambulatory care”[MeSH]	1968
	“Ambulatoty Care Information Systems”[MeSH]	1991
	“Ambulatory Care Facilities”[MeSH]	1983
	“Outpatients”[MeSH]	1991
	“Outpatients Clinics, Hospital”[MeSH]	1978

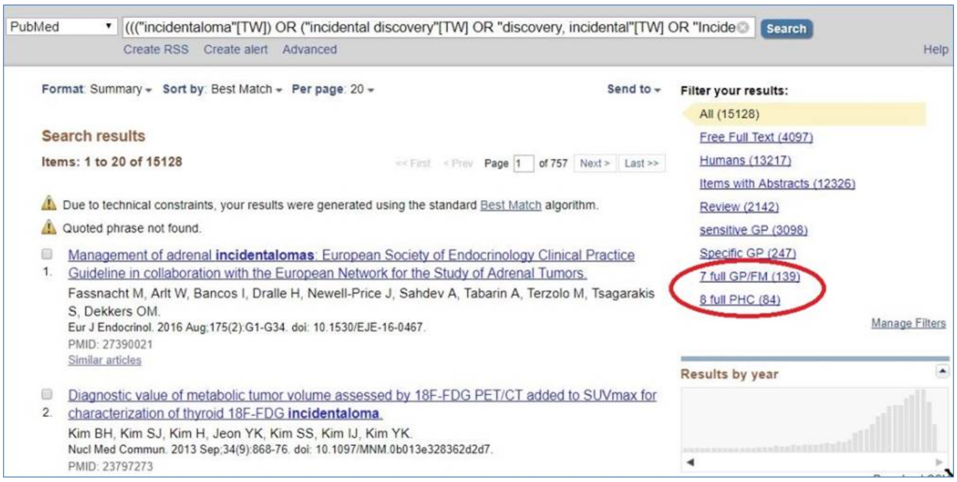
- The 7 GP/FM related MeSH
  The French Meta-term *Médecine générale* [general practice] set of descriptors which represent the domain general practice encompasses these 7 Mesh terms. On September 09, 2017, this Meta-term yields 197.759 citations (see also Appendix E, page 209). We reproduce here the seven MeSH in French with the English translation:

  - médecine générale [General practice][MH]
  - médecine de famille [Family Practice][MH]
  - médecins généralistes [General Practitioners][MH]
  - médecin de famille [Physician, Family][MH]
  - médecine communautaire [community medicine [MH]
  - médecins de premier recours [Physician, Primary care][MH]
  - médecin référent [gatekeeping][MH]

  Those seven MeSH quoted in the middle part of the Fig. 1.50 have also been chosen as mappings for the Q-Code QS41 *Family physician*. The resulting search strategy is documented in the Appendix E, page 209.

  MyNCBI is a PubMed facility that allows one to have a personal page in PubMed and to apply specific filters among other facilities. The fig. 1.51 shows a screen-copy of the PubMed interface with the MyNCBI of the author. The QD324 Incidentaloma results are filtered either by the GP/FM either by the PHC specific filter.

FIGURE 1.51: The search strategies GP/FM and PHC (see Appendix E) have been entered as filters in MyNCBI facilities. The fist search is about *QD324 Incidentaloma* with the corresponding HeTOP query which yields 15.128 citations on July 30, 2017. The red circle shows the results given by the two respective filters GP/FM (139) and PHC filters (84)



1.3.7.4 Other examples of confusion in the MeSH

- The Q-Code *Care Manager* QD3 would certainly fit with the definition of the MeSH *Clinical Governance*. But this definition is entirely related to UK National Health Service which introduces a bias in the search. This is why we have refused the link to the MeSH *Clinical Governance* and added instead *Clinical Guidance* as supplementary entry term for the HeTOP search equation.

- *Action research* is a MeSH entry term for the MeSH *Health Services research* of which the scope note in the MeSH thesaurus is ; *The integration of epidemiologic, sociological, economic, and other analytic sciences in the study of health services*. In GP/FM and PHC, *action research* is usually related to participatory research in which actors and field of research could be transformed by the search. So we discard the relation with the *Health Services research* MeSH.

- Entry terms are sometimes more specific than the main MeSH ex : the MeSH *Delivery of Health Care* ; The concept is concerned with all aspects of providing and distributing health services to a patient population. The MeSH *Delivery of Health Care* is addressing health care as a whole while its numerous entry terms as *Contraceptive, Distribution, or Community-Based Distribution* are very specific to *Primary Care*. This makes *Delivery of Health Care* difficult to use in the realm of GP/FM.

- The MeSH *Referral and consultation* contains *Gatekeepers, Health Service* as entry terms. But referral doesn't include mandatorily the concept of gatekeeper, which is more related to the organization of the first line.

- Scales and assessment tools are numerous in Medline. We have chosen to use only some descriptors, more adapted to GP/FM as *relative value scales; outcome assessment (health care); visual analog scale; questionnaires; psychiatric status rating scales.* Worth to note that Mini Mental State Evaluation (MMSE) and Likert scale, daily tools in primary care, have no entry in MeSH. We have added them to our entry terms in the Q-Codes QR52 Scale.

- The case of *Shared Decision Making* (SDM) in MeSH is also appealing. In Medline, SDM is a MeSH entry term (i.e. a synonym) of the MeSH *Decision making*. The consequences are that searching SDM in PubMed retrieves thousands of citations dealing with *Decision making*, which rarely deal with the decision shared with the patient. Consequently the automatic mapping proposed by HeTOP between SDM and Decision making has to be labeled false. The resulting search strategy will be *"shared decision making"[TW]* which retrieved 4687 citations on June 14, 2017. If the MeSH *Decision making* is added to the first search, the number of citations jumped to 168,794 on June 14, 2017 .

FIGURE 1.52: Some refused mappings between Q-Codes and MeSH.

Q-Code	Preferred term	MeSH mapping refused
QD3	care manager	accountable care organizations managed care programs health planning advance care planning
QP43	patient autonomy	choice behavior
QD22	comprehensiveness	comprehensive health care
QD323	shared decision making	decision making
QS33	coordination of care	delivery of health care, integrated
QR35	action research	health services research
QC2	gender issue	homophobia sexual dysfunctions, psychological sexual dysfunction, physiological
QD443	overinformation	information literacy access to information information seeking behavior consumer health information
QP25	acceptability	patient acceptance of health care
QT42	online knowledge-sharing	publishing
QS32	referral	referral and consultation
QD31	health risk management	risk management
QD42	secondary prevention	secondary prevention
QC55	ritual mutilation	self mutilation
QP61	social network	social networking social networks
QC24	transgender	transsexualism

- *Primary care provider*: this concept is not defined in MeSH although the term exists-though solely in the definition of the MeSH secondary care ; *Specialized health care delivered as a follow-up or referral from a PRIMARY CARE provider*.

- *Comprehensiveness*: this concept has various meanings in the ongoing GP/FM literature. When considering the patient doctor relationships he is equivalent to biopsychosocial, whole person care, holistic care or globality (Allen et al., 2005). From organizational point of view he is used for its integrative content, as *efforts to improve equity in access, community empowerment and participation, social and environmental health determinants, and intersectoral action* (Labonté et al., 2014). The use done by Barbara Starfield is related to organization as she speaks about *comprehensiveness of services within primary care settings* (Starfield, 2010). Only this last meaning is in line with the MeSH *Comprehensive Health Care Providing for the full range of personal health services for diagnosis, treatment, follow-up and rehabilitation of patients.* Consequently, *Comprehensiveness*, in his GP/FM sense, described by Allen et al. (2005) has no place in MeSH.

- The recent (2016) MeSH *medical overuse* has a restricted scope note; *Excessive or unnecessary utilization of health services by patients or physicians*. Nevertheless it has 30 entry terms including *misdiagnosis, overdiagnosis or overtreatment* among others. To generic, it is useless for characterizing the details of the recent move in the domain of too much medicine.

- One could consider the first part of the definition of the MeSH *Medical Futility ; The absence of a useful purpose or useful result in a diagnostic procedure or therapeutic intervention* as a good description of the concept Overtreatment. Nevertheless, when considering its real use in Medline, one sees that this MeSH is exclusively related to terminal care as explained by the second part of the MeSH definition; *The situation of a patient whose condition will not be improved by treatment or instances in which treatment preserves permanent unconsciousness or cannot end dependence on intensive medical care*.

FIGURE 1.53: Some approximate mappings between Q-Codes and MeSH.

QCode ID	Preferred term	Approximate MeSH mapping
QR53	functional status	Patient acuity
QT24	balint group	Sensitivity training groups
QT25	trainer	Mentors
QD22	comprehensiveness	Holistic health
QD42	secondary prevention	Early diagnosis Mass screening
QD446	overtreatment	Medical futility

- Some mapping between Q-Codes concepts and MeSH are seen as approximate (see Fig. 1.53). *Functional status* (Q-Code QR53) is defined in WONCA dictionary as *the ability of a person to perform and adapt to his environment, measured both objectively and subjectively over a stated period of time* (Bentzen N.(ed), 2003). This is rather to be mapped to the MeSH *Patient acuity; An assessment of a patient's illness, its chronicity, severity, and other qualitative aspects*, than to the MeSH *Health Status Indicators; The measurement of the health status for a given population using a variety of indices, including morbidity, mortality, and available health resources*. *Patient acuity* [MH] is a child of *Health outcome indicators*[MH]. But as *Appraisals, Health Risk* is an entry term of *Health Outcome Indicator* it could only be considered as a partial map to Functional Status.

- *Balint group* could be mapped to the MeSH *Sensitivity training group; A group of people who meet in an unstructured setting to learn about themselves, interpersonal relationships, and group processes and about larger social systems.* But *Balint group* are specific to health care provider which is not mentioned in the MeSH scope note of *Sensitivity training group*.

1.3.7.5 The MeSH and the Prevention Concept

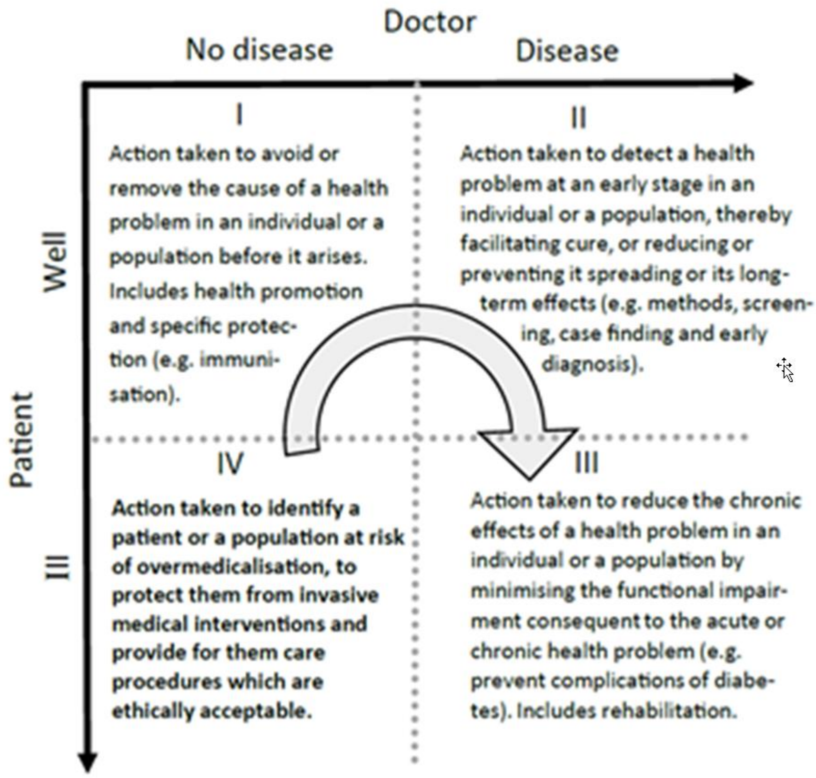
The domain of prevention is one of the most complicated to understand in the MeSH thesaurus. There is no MeSH to describe the general domain of *clinical prevention*. The subheading *prevention and control* will fit for this use; *Used with disease headings for increasing human or animal resistance against disease (e.g., immunization), for control of transmission agents, for prevention and control of environmental hazards, or for prevention and control of social factors leading to disease.* It includes *preventive measures in individual cases*. But subheading are not usable in the Inforoute automatic search facility. Worth to note that we are not confounding *Prevention* and *Health promotion* (Vanmeerbeek et al., 2015) which has a different entry in Q-Codes ; QD23 Health education.

WONCA has his own definition of clinical prevention ; *PREVENTION : action to avoid occurrence or development of a health problem and/or its complications* (Bentzen N.(ed), 2003).

The clinical prevention concept can be divided into four categories. This is more understandable by positioning the four definitions in a double entry table, crossing the patient feelings and the doctor decisions. One can understand quickly that the fourth form has a deep consequence for all the 3 others, in a cybernetic move indicated by the arrow. (see Fig. 1.54). Quality assurance of the whole system is deeply rooted in GP/FM concept of prevention.

Several MeSH are available which addresses the issue of **Primary prevention** in clinical care i.e. before disease and before the patient is aware of any problem. Our chosen search strategy encompasses those three MeSH: *"Primary Prevention"[Mesh] OR "Immunization"[Mesh] OR "Vaccination"[Mesh]*

FIGURE 1.54: The four forms of prevention in GP/FM (Jamoulle & Roland 1995)



including their term entries tagged as as Text Word [TW] (431,023 citations on October 11, 2017).

The domain of **secondary prevention** is probably the most controversial. The cardiologists are using it in its pure chronological sense, i.e. after an event as for instance aspirin after infarcts or stroke. The MeSH definition quoted for *secondary prevention* is; *The prevention of recurrences or exacerbations of a disease or complications of its therapy.* One notes that this is exactly the content of the concept of tertiary prevention in family medicine. The GPs are using secondary prevention when they bet the patient, unaware of any problem, would have a disease they are searching for. This corresponds to the definition of case finding or screening. As we will see in the next point, the MeSH *Secondary prevention* doesn't fit with the concept containing similar words used in family medicine. For searching secondary prevention in Medline we will use only two MeSH as in the following search strategy; *"Early Diagnosis"[Mesh]) OR "Mass Screening"[Mesh]* (142174 citations on June 5, 2017). We have to point the MeSH *Secondary prevention* as false in the HeTOP interface as search term for the Q-Code secondary prevention.

The domain of **tertiary prevention** in Family practice addresses the prevention of complications and the rehabilitation. Consequently and in line whit what has been said in the last paragraph, the chosen equation for searching items dealing with tertiary prevention in Medline would paradoxically be : *(("Secondary Prevention"[Mesh]) OR "Tertiary Prevention"[Mesh]) OR "Rehabilitation"[Mesh]* (275.256 citations on June 5, 2017). This Inforoute born results are not so fruitful. Although retrieving a lot of publications, it's difficult to find one addressing what is Tertiary prevention in Family medicine. This term is not part of the common academic teaching and training language.

The concept of **quaternary prevention**, born in 1986 (Jamoulle, 1986), published in the WONCA dictionary in 2003 (Bentzen N.(ed), 2003), diffused worldwide in the GP/FM world, is not existing in MeSH. The search of the text words with the equation *"quaternary prevention"[TW]* retrieves 52 citations in PubMed (on July 30, 22017). The four WONCA definitions including *quaternary prevention* have been incorporated in DeCS, the PAHO Health Sciences descriptors in 2015. (`https://tinyurl.com/DeCS-P4`).

1.3.7.6 Study of he mappings GP/FM and PHC with MeSH 2016 & 2017

The MeSH increases each year by several thousand units. Of all these new descriptors, only a few relate to family medicine. The Q-Codes are linked to the MeSH quite precisely. It is therefore important to see the new opportunities for mappings between the two terminologies. In addition, each year, the MeSH are systematically translated into French and incorporated into the HeTOP database, with a certain delay compared to the publication of the MeSH in English. It is therefore necessary firstly to choose the new MeSH candidates for mappings with Q-Codes and, on the other hand, to establish these mappings in HeTOP by filling the DBGUI creation interface (see Fig. 1.43, page 71. In 2016 we have made a revision of this list to identify the new MeSH compatible with the realm of GP/FM (Jamoulle, 2016). Of the list of concepts of the Fig 1.3.7.6, one only appears as new MeSH in the list of MesH published in 2017. The concept *Medically Unexplained Symptom*, known under the acronym MUS, heavily discussed by GPs, has been accepted as a MeSH in 2016. The term has already been quoted into the definition of the MeSH *Somatoform disorder* since 1981. The Fig 1.55 shows a sample of some MeSH edited by NLM in 2017 of which *Medically unexplained symptoms*. This work was carried out in 2016 and is underway for the new MeSH of 2017. The new MeSH 2016 corresponding to a general medical field was published in the International Journal of Health Policy and Management (Jamoulle, 2016).

1.3.7.7 Mapping to Health Science Descriptors (DeCS)

The multilingual structured vocabulary DeCS (Health Sciences Descriptors) serves as a unique language in indexing articles from scientific materials. It also searches and retrieves subjects from scientific literature, pulling from information sources available on the Virtual Health Library (VHL) such as MEDLINE and LILACS (Portuguese abbreviation for Literatura

FIGURE 1.55: Some MeSH edited in 2017 by the National Library of Medicine and their Q-Codes mappings

New MeSH 2017	Scope note	Included in
Medically Unexplained Symptoms	Persistent health symptoms which remain unexplained after a complete medical evaluation. A cluster of symptoms that consistently appear together but without a known cause are referred to as a MEDICALLY UNEXPLAINED SYNDROME (MUS).	QD321 Medically Unexplained Symptoms
Mentoring	The art and practice of training, guiding, COUNSELING, and providing support to a less experienced, trained, or knowledgeable person(MeSH 2017)	QT25 trainer
Multiple Chronic Conditions	Two or more concurrent chronic physical, mental, or behavioral health problems in an individual. (MeSH 2017)	QD322 multimorbidity
Teacher Training	A curriculum-centered process of equipping individuals with professional knowledge and skills required to become effective teachers. (MeSH 2017)	QT25 trainer
Vaping	The act of inhaling and exhaling vapors produced from ELECTRONIC CIGARETTES.	QC4 Addict
Work-Life Balance	The optimal arrangement of an individual's on-the-job and private time to facilitate health and personal satisfaction without negatively impacting productivity and professional success. (MeSH 2017)	QD8 Work-Life Balance

Latino-Americana e do Caribe em Ciências da Saúde, Latin American and Caribbean Literature on Health Sciences) (http://lilacs.bvsalud.org/en/). DeCS is one of the most important and comprehensive indexes of scientific/technical health literature in Latin America and the Caribbean (Pellizzon, 2004).

DeCS was developed from/modeled after MeSH (Pereira et al., 2012). It is available in four languages (en, es, pt, fr) under the form of a standardized terminological record online. Expanding upon the MeSH vocabulary, DeCS covers 4583 (May 2017) supplemental descriptors -called here non-MeSH descriptors- classified into 4 categories: 1861 homeopathy, 160 science and health, 2305 public health and 411 about health surveillance. Attentive reading of these supplemental DeCS, with the help of Arthur Treuherz, senior terminologist at BIREME retrieved 131 possible correspondences with Q-Codes. These have been introduced in the Q-Codes data base with their definitions.

1.3.8 Dissemination phase

This phase has been ongoing since this work's commencement in 2013. Various initiatives allow for the worldwide diffusion of information and are detailed in the section Results.

- availability of ICPC-2 in 21 languages and of Q-Codes in 10 languages on the HeTOP server, providing free inscription.
- availability of ICPC-2 and Q-Codes under URI format
- export facilities for the whole database through a dedicated web service (www.hetop.eu/hetop/documentation/ws.html).
- creation of the 3CGP web site (http://3cgp.docpatient.net/)
- publication of the Terminology in GP/FM as books in 6 languages

- emergence of the Q-Codes working group inside the WONCA International Classification Committee
- development of HeTOP and DBGUI user guides
- development of a 3CGP user guide

In this thesis, we see that the results are confused with the methods, highlighting the adage that the path is sometimes more important than the goal. In this part of the methods, we will show how the database we have created can serve first as a multilingual human reading resource for teaching and research, secondly as Uniform Resource Identifiers for machine reading, and thirdly as and exportable knowledge-base fit for the Linked data world.

1.3.8.1 Consulting the data base through the HeTOP web service

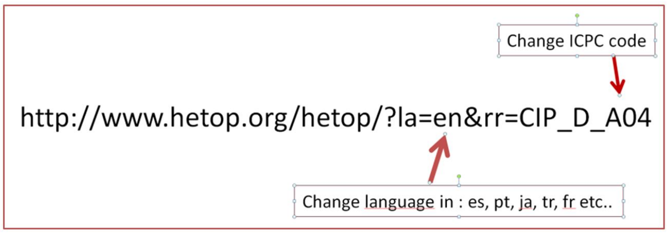
The HeTOP database is freely accessible by means of a simple registration process. The multilingual HeTOP web page (www.hetop.eu) gives access to the interrelated content of more than 70 terminologies. Among those, the most important in GP/FM are the Medical Subject headings (MeSH), the Health Descriptors (DeCS), the Dictionnaire des Résultats de Consultation (DRC) [Dictionary of encounter output], the International Classification of Disease (ICD), the International Classification of Primary Care (ICPC) and the Contextual classification in GP/FM (Q-Codes). We will see in further that this database is a source of knowledge and pedagogic support in several languages. The URL www.hetop.eu/Q gives direct access to the Q-Codes page.

1.3.8.2 Reaching the base through Universal Resource Identifiers

The Department of Information and Medical informatics at Rouen University, which has mastered HeTOP, has given General Practice a wonderful opportunity to highlight its specific fields of knowledge. It has allowed for free access to the same material in URI format (see Fig.1.56). ICPC in 21 languages and Q-Codes in 10 languages (more expected) are available in URI format, rubric by rubric and language by language, thanks to the dedicated work of Julien Grosjean, PhD, Associate Professor of Computer Sciences, and Stefan Darmoni, MD, PhD, Professor of Medical Informatics, at Department of Information and Medical informatics at Rouen University. This will greatly increase the visibility of ICPC and Q-Codes. This will not, however, replace the e-health Norwegian ICPC repository (https://ehelse.no/icpc-2e-english-version).

This is a pedagogic tool aimed at the dissemination of knowledge. URIs will be maintained and could be used in any web based health information system, EMRs, or as teaching material. The chain of character of a URI is stable. Language are expressed under the ISO 639 Codes for the representation of names of languages and ICPC-2 or Q-Codes classes by their

FIGURE 1.56: The URI for the code ICPC-2 A04 (Tiredness) in English



respective codes (see fig. 1.56). The following URIs are giving access to the hierarchies and rubrics of the corresponding classification. Note that each entry give access to a detailed terminological description, mappings to other terminologies and to automatic queries on resources like PubMed.

To change the language; change the ISO 639 for the language ; Ex: =en for English, =pt for Portuguese. For ICPC-2, the are 21 languages allowed and 10 for Q-codes (en, fr, es, pt, tr, vi, ko, nl, ka, de - more in progress). To change the rubric; change the code at the end of the URI.

Some examples ;

- URIs to reach the hierarchy of ICPC and Q-Codes in English.
  - ICPC-2: http://www.hetop.org/hetop/?la=en&rr=CIP_C_ARBO&tab=1
  - ICPC-2 Process: http://www.hetop.org/hetop/?la=en&rr=CIP_C_ARBOPROC&tab=1
  - Q-Codes: http://www.hetop.eu/hetop/Q?la=en&rr=CGP_CO_Q&tab=1
- URIs to reach each rubrics of ICPC and Q-Codes in English.
- Other examples :
  ICPC-2 S91 Psöriazis [Psoriasis] In Turkish:
  http://www.hetop.org/hetop/?la=tr&rr=CIP_D_S91
  ICPC-2 S91 Vảy nến [Psoriasis] in Vietnamese :
  http://www.hetop.org/hetop/?la=vi&rr=CIP_D_S91
  ICPC process code 33 microbiological/immunological test in English:
  http://www.hetop.org/hetop/?la=en&rr=CIP_P_33
  ICPC process code 36 analyse de selles [faeces test] in French:
  http://www.hetop.org/hetop/#la=fr&rr=CIP_P_36
  ICPC-2 S01 皮膚の疼痛、圧痛 [Skin pain] in Japanese :
  http://www.hetop.org/hetop/?la=ja&rr=CIP_D_S01
  Q-Code QC Patient category in English :
  http://www.hetop.eu/hetop/Q?la=en&rr=CGP_QC_QC
  Q-Code QD323 Toma de decisiones compartida [Shared decision making] in Spanish :
  http://www.hetop.eu/hetop/Q?la=es&rr=CGP_QC_QD323

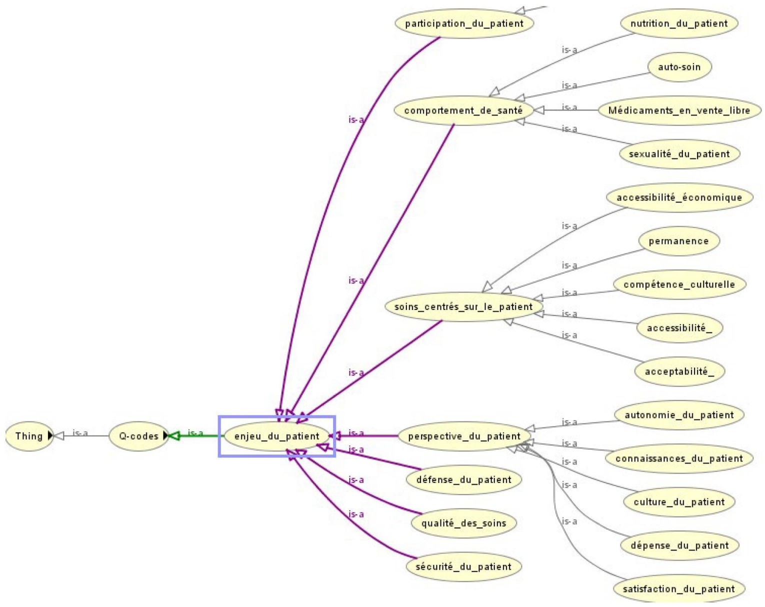
- The QO domain (Other) and its URIs in English:
  QO : Other http://hetop.eu/hetop/Q?la=en&rr=CGP_QC_QO
  QO1 : Unable to code, unclear http://hetop.eu/hetop/Q?la=en&rr=CGP_QC_QO1
  QO2 : Acronym http://hetop.eu/hetop/Q?la=en&rr=CGP_QC_QO2
  QO2 : Out of scope of GP/FM
  http://hetop.eu/hetop/Q?la=en&rr=CGP_QC_QO3
  QO4 : Consider new code
  http://hetop.eu/hetop/Q?la=en&rr=CGP_QC_QO4

1.3.8.3 Exporting the database for further uses

The content of the HeTOP database could be downloaded in CSV (EXCEL), RDF, SKOS or OWL format through a dedicated web service (www.hetop.eu/hetop/documentation/ws.html). For further development in ontology creation, the whole base can be exported into OWL. Those files can also be imported in the WebProtégé server, an ontology editor developed at Sanford (See Fig. 1.57). WebProtégé is a free, open-source ontology editor and framework for building intelligent systems. (http://protege.stanford.edu/) created at Stanford Center for Biomedical Informatics Research at the Stanford University School of Medicine ((Tudorache et al., 2013). Webprotégé is a knowledge acquisition tool for the Web, accessible from any Web browser. Protégé is aimed at developing and maintaining ontologies.

All the information gathered in the HeTOP website can be extracted in OWL and consequently incorporated in WebProtégé as shown in Fig.1.57. Q-Codes has the format of a concept hierarchy and could be considered as lightweight ontology . A concept hierarchy is *a way of defining a conceptualization of an application domain in terms of concepts and relationships expressed in a formal language* (Giunchiglia and Shvaiko, 2003). By introducing the Q-Codes concept hierarchy into WebProtégé, we can consider populating this hierarchy and developing a full ontology.

FIGURE 1.57: The opening of the domain QP (patient issue) in WebProtégé OWLV iz (in French - partial view)



1.4 RESULTS

Most results of this thesis have been published in the medical literature. In the present introduction some unpublished elements are addressed. Most important papers published in medical journals are reproduced in the following chapters or are quoted when appropriate.

1.4.1 Results of preliminary research in the domain of classification, terminology and interoperability

1.4.1.1 In Search of Interoperability

Since 1992, the author has been involved in taxonomic research in the field of GP/FM. At the turn of the century, it became clear that the limits of the pick-list type terminologies created by earlier generations were close to being reached. In turn, a better way of managing clinical information in GP/FM had to be found. Participation in the MERITERM group has

FIGURE 1.58: The Meriterm logo
(http://meriterm.org/)

MERITERM
European research group in Terminology

allowed for a deeper understanding of the mechanisms of interoperability. MERITERM is a consortium devoted to joined research activities over medical terminologies and classification systems. MERITERM stands for Medical Enduser and Reference Interface Terminology. The project aims to identify an extensive, but well delineated, set of core, multi-lingual meanings for words and phrases commonly used in medical communication. In addition, it aims to build an interface terminology for human/machine and human/human interaction (medical specialists, general practitioners, allied health professionals, pharmacists and patients). This core set will be linked to existing linguistic resources, as they are readily available in many languages. It will also be linked to a broad array of international medical classifications. Difficulties that arise when trying to converge concepts present in several terminologist tools (the objective of interoperability) are considerable. By examining the French text of a guideline for Belgian GPs, terminological inadequacies, inconsistencies, and the difficultly of interoperability of concepts have been shown. This kind of analysis reinforces the impression that it is essential to utilize the immense power of computers and a new management of terminology to have truly interoperable data. A study dealing with this topic, published in *Informatics in Primary Care*, is reproduced in Chapter 5, page 147.

1.4.1.2 Towards Semantics

New, powerful tools for data management have emerged within the last decade, that will revolutionize electronic health records and health information systems (Liyanage, Liaw, and De Lusignan, 2012). This emphasizes the role of GP/FM, as General Practitioners are the first, main and most common point of contact with patients. They are the source and endpoint in the circle of information generated by judicious medical documentation and wise secondary use of medical data. The paper published in *Merit Research Journal of Medicine and Medical Sciences* is an introduction to health terminologies, ontologies, semantic data, and linked open data. All are expressions used by computer scientists, preparing themselves for The Semantic Web within the realm of health care data. This paper is reproduced in Chapter 6, page 159.

1.4.1.3 From Terminologies to Ontologies

The in-depth study of interoperability's difficult conditions and new capabilities of information management in semantic technologies have opened the door to another universe. Although GP/FM is highly complex and often intertwines with biosciences/anthropology- while simultaneously covering all medical terminology, the creation of an ontology for GP/FM was no longer impossible. It was necessary to patiently assemble the pieces from existing ones. Freedom from language through the use of concepts as a central element allowed for surprisingly quick dissemination in the GP/FM field. Therefore, ICPC patient-centered classification, available in 21 languages, should serve as an essential basis for a future ontologies. By associating a classification for professional context under Q-Codes, an additional step is realized. The ICPC set and Q-Codes, now known as Core Content Classification in Family Medicine (3CGP), expressed on the HeTOP site and available in URI format, is the beginning of what could potentially be a universal ontology for GP/FM. In this ongoing state, it is called a lightweight ontology, which forms a solid basis for future development. This vision is shared by Professor Gustavo Gusso, Head of the Department of Primary Care at Sao Paulo University. He assisted in sharing the importance of this ontology through a paper published in the *The World Book of Family Medicine – Iberoamericana Edition 2016* (Gusso and Jamoulle, 2016).

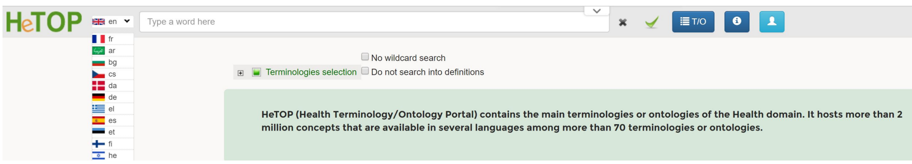
1.4.2 New Tools for GP/FM

The tools developed to carry out this research can be considered results, and they are presented here as such. The HeTOP server and its DBGUI editor, the Terminologies in 10 languages, the provision of ICPC-2 and the Q-Codes in the form of Uniform Resource Identifiers (URIs), the 3CGP web site, the Q-Code working group are technical, communicational or human realizations aimed at the achievement of this endeavor.

1.4.2.1 On-line classifications & terminology

ICPC-2 and Q-Codes are among the terminologies published on the HeTOP server. The provision of ICPC-2 into 21 languages (at this date) by the Rouen team was the subject of a Poster presented at the French Congress of General Practice in 2016 (Schuers et al., 2015). For now, at the end of 2017, ICPPC is edited in 21 languages.
The process of developing the Q-Codes into 10 languages is detailed in the

FIGURE 1.59: The HeTOP multilingual terminology server http://www.hetop.eu showing some of the available languages



Methods section. The process has been the subject of several publications. It has already been said that the purpose of this thesis was articulated in three domains; Family Medicine, computational linguistics and documentary sciences. This explains why the creation of Q-Codes was published in three newspapers. It was published from three points of view, corresponding first to computer sciences, then to family medicine, and finally to documentology- or more precisely, grey literature.

The paper presented in Manchester before a panel of computer scientists at MIE 2017 is reproduced in Chapter 8, page 179 (Jamoulle et al., 2017a). The publication for GPs will be published in the European Journal of General Practice. Finally, the last paper is ready to be published in the proceedings of the Congress of Grey Literature in Rome, October 2017, which have a more linguistic orientation with a focus on documentary sciences. Its abstract is reproduced on chapter 10, page 189 (Jamoulle et al., 2017d).

1.4.2.2 The 3CGP Website

Dissemination through an Internet pedagogical web page is encompassed in a website dedicated to this work. The domain name *docpatient.net* has been the property of the author for over 20 years. A sub-domain has also been opened. At this address, one can find all materials needed to understand and apply the Q-Codes/ ICPC including copy desks in various languages. (http://3cgp.docpatient.net/)

FIGURE 1.60: Upper ribbon of the website http://3cgp.docpatient.net/, showing the different rubrics

1.4.2.3 Q-Codes Working Group inside the WONCA International Classification Committee (WICC)

During the WONCA International Classification Committee in Turku, Finland in 2016, the work on Q-Codes was presented. The WICC accepted the task of creating a Q-Codes working group to continue and expand this work. A dedicated Google group is also available. Information on Q-Codes is now also available on the WICC webpages (http://www.ph3c.org).

1.4.2.4 The Multilingual Terminology Books

The translation of the Q-Codes terms and their definitions by colleagues gave rise to a presentation in the form of vocabulary lists. It has become known as *Terminology in Family Medicine*. Each definition has been extracted from existing terminologies or dictionaries. The Dictionary of General Practice edited by Bentzen N.(ed) (2003) on behalf of WONCA provided a good deal of it, and the other definitions were provided from MeSH, WHO Glossary, and other sources. The list of sources of the definitions is edited in the books of these terminologies and reproduced at Appendix G, page 223.
The list of Q-Codes (url http://3cgp.docpatient.net/tabular/) and terminologies (url http://3cgp.docpatient.net/the-book/) can be found on-line in English , French, Dutch, Spanish, Portuguese, Turkish, Korean, Vietnamese, German and Georgian. In addition, the terminologies in English, French, Dutch, Spanish and Portuguese have been published in book form. They are printed in Belgium, and they can be ordered on-line via: http://goo.gl/oWAJHi. The same book, written in Vietnamese, has been published in Vietnam and is available in full in pdf format at: http://orbi.ulg.ac.be/handle/2268/202214

1.4.2.5 HeTOP Server and DBGUI User Guides

Three on-line Power-points, which allow for deeper comprehension of these tools are available at: http://3cgp.docpatient.net/tutorials/. The first two are demonstrations of the HeTOP and Q-Codes for the users. The third has been developed in case some volunteers would be keen to participate in editing the Q-Codes base.

- Exploring HeTOP and ICPC-2 on HeTOP
- Exploring Q-Code on HeTOP
- Editing with DBGUI on HeTOP

1.4.2.6 The 3CGP User Guide

The Q-Codes user guide is edited on-line, in French, as a part of a guide for final work for trainees (see further) and in English at http://3cgp.docpatient.net/tabular/. The guide for trainees tackles the work stages successively: choice of the problem, definition of objectives, bibliographic research, ethical aspects, action plan, and drafting plan. It deals also with the choice of objectives and the appropriate method of investigation, motivation for the thesis work, planning of the follow-up, evaluation criteria, and quality control of the work (Vanmeerbeek, Felgueroso-Bueno, and Lafontaine, 2016). The indexing guide sets out the purposes of on-line indexing: visibility of works of the discipline (grey literature), knowledge networks, and quality control. Coding is based on the ICPC-2 for clinical aspects and Q-Codes for contextual aspects of the work.

1.4.2.7 3CGP in URI format

ICPC and Q-Codes are now available in URI format. The Uniform Resource Identifier (URI) is explained in the Methods and the various access granted. (see 1.3.8.2, page 87). A URI allows one to tag a word or text by the corresponding Internet address. This is the simplest way to demonstrate the strength of this technology. But URIs are also the fundamental basis for the development of the semantic web, linking data to other strands of data. In turn, they are the basis of distributed data.

1.4.3 Analyzing data with 3CGP

The tools for encoding with ICPC-2 have been available for quite a while. However, building the taxonomy for Q-Codes has been an arduous journey, but it was ultimately met with success. The hard-copy versions of Q-Codes were numbered from 0.0 to 2.5. The on-line version, available on the HeTOP server, can be considered as the electronic version 2.5. (see versioning at page 52).

The presentations accepted in congresses about Q-Codes construction, were systematically published on the ORBI documentary database from the University of Liège. They can, therefore, be viewed online. Here, the results of the conferences from Paris 2007, Covilha 2013, Clermont Ferrand 2013 and Lille 2014 are presented. Those sources are described in Fig. 1.19.

In fact, the genesis of the Q-Codes classification began well before 2007. In 2005, at the annual WICC meeting, a presentation entitled *Attempt to classify main descriptors of GP / FM job. Proposal for a metaclinical classification*" (Jamoulle, 2005) was presented. To separate the conceptual world of the patient (Schrans et al., 2016), from that of the physician, the term contextual professional aspects is used now.

At the end of 2017, the Congress of the Brazilian Society of Family and Community Medicine to be held in Curutiba on December 7, 2017, resulted in 1746 abstracts encoded by participants at the initiative of the Scientific Committee (Daniel Knupp, Paulo Poli and Gustavo Gusso). In Portugal, at the initiative of Professor Luiz Miguel Santiago of the General Practice Clinic of the Faculty of Medicine of the University of Coimbra, 300 master thesis are presently coded by Dr. Ariana Tavares.

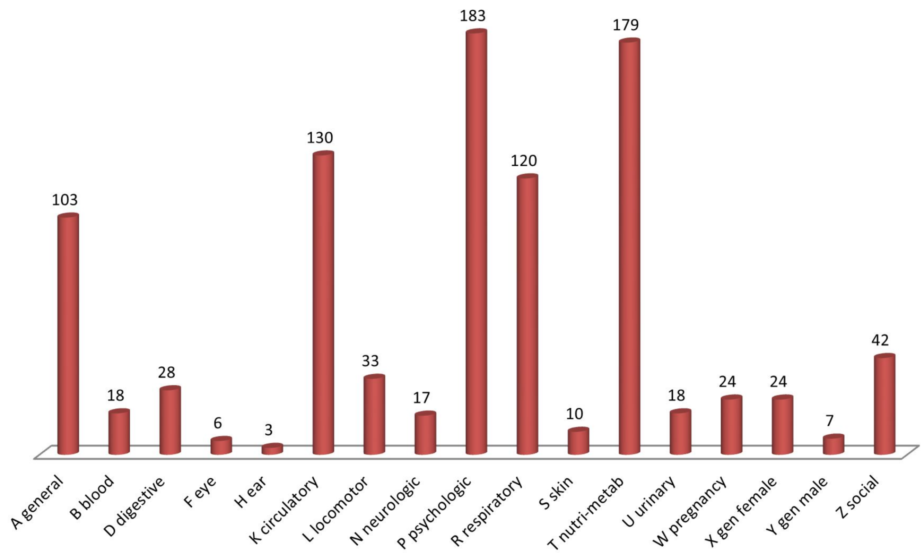
1.4.3.1 Analysis of Data from WONCA Europe 2007 Conference

The work to identify contextual or managerial concepts in abstracts presented during conferences began in 2007. The 998 abstracts of the 2007 WONCA Europe congress have been carefully analyzed, in which main themes were identified by the author, classified using ICPC for identification of clinical issues, and applied to a self-ordered authority list of concepts (Q-Codes) for non-clinical or managerial issues. Q-Codes version 2.0 has been used jointly with ICPC-2. In fact, version 2.0 (see Appendix A, page 191), has been elaborated upon, during the coding process in a inductive/deductive process, as explained above.

- Main findings in WONCA 2007 communications

  The two figures below show the general distribution of ICPC-2 codes and Q-Codes. ICPC-2 users are accustomed to this histogram presentation. The number of cases is in the y-axis; the 17 chapters of ICPC-2 are in the x-axis. Usually, ICPC-2 is used in clinical epidemiology to record the activity of family physicians. Then, the distribution of health problems in the consulting population is considered. Here, the distribution of abstracts grouped into similar categories is examined. In the first case, it is a statistical approach to what GPs do. In the second, it is an anaysis of what they discuss. Let's look at the patient issues that the GPs discussed. In Fig. 1.61 it is seen that psychological problems (183 cases in P), followed closely by nutritional problems (179 cases in T) were often discussed. In fact, the distribution of codes into chapters show that GPs mainly discuss depression and diabetes. For the rest, the common topic of discussion was clinical studies. One point attracts a lot of attention. The number of social problems (42

FIGURE 1.61: Distribution of ICPC-2 codes in the 998 abstracts of the 2007 WONCA Conference



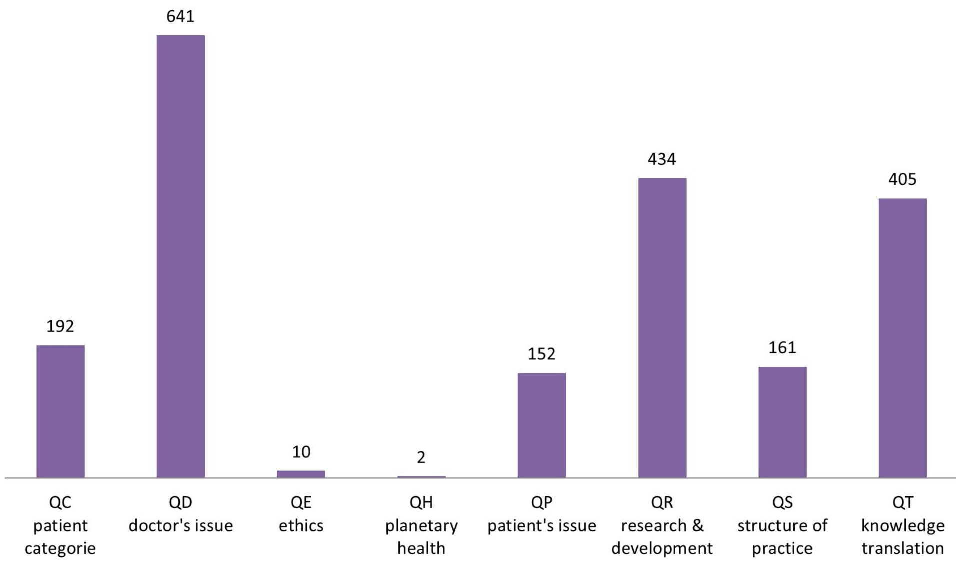
cases in Z) discussed is very important when compared to what is encountered in the data-retrieval coming from clinical activities with patients. It can, therefore, be said that when GPs met in Paris in 2007, they often discussed P & Z problem- i.e., psycho-social issues and T problem (nutritional & metabolic issues), mainly diabetes. Note that Chapter Z of ICPC-2 incorporates items that are considered part of the psychological field, such as the loss of a loved one or intra-family violence.

Let us now consider the results of the Q-Code Ver.2.0 coding (see Fig. 1.62). The number of cases is in y-axis. In x-axis we find the 8 domains of Q-Codes. It must be clear for the reader that a code is attributed when the corresponding theme is addressed in an abstract. For example, saying that ten codes have been found in a conference implies that ethical themes can be found in ten different abstracts.

The reader should read the codes used in Appendix A, page 191 to understand the distribution of the Q-codes in the histogram. The field QD, Doctor's issue, contains areas that deal mostly with decisions and professional competence of the physicians -i.e., communication, prevention, evaluation of a health problem or medico-legal problems. We see that GPs often discuss their area of expertise with 641 codes in the QD (Doctor's issue). Patients' perspectives (QP, 152 codes) and patient categories (QC, 192 codes) are less frequently discussed.

The QR (research & development) domains codes and the Training and Teaching domain, renamed Knowledge Translation (QT) respectively document 434 and 405 codes. It is a WONCA conference, and

FIGURE 1.62: Distribution of Q-Codes in the 998 Abstracts of WONCA Conference 2007



the academic aspect is thus strongly marked by these two strong prevalences. It is known that WONCA brings together a very large number of university departments of GP/FM. Thus, we see that the latter is very active, as shown by the work presented in the field of research and pedagogy.

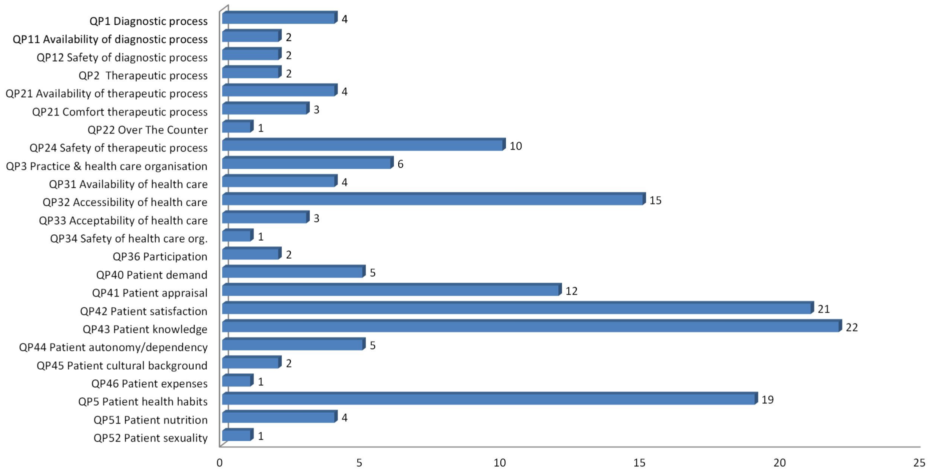
The domain QS structure of practice (161 codes) contains the items related to the overall organization of care and thus to the determinants of primary health care such as infrastructure, practice relationships or coordination.

The lack of codes attributed to ethical issues (QE), only 10 codes, and the virtual absence of codes in the field of planetary health (QH), only 2 codes, is challenging. We have only analyzed here the abstracts provided by individual participants, not the special session's abstracts that were not accessible at the time of analysis. Thus, it can be inferred that abstracts about ethical issues are carried out by doctors anxious to discuss it and not by specific working groups. We note that this trend is also found in the other conferences discussed later on.

The QP domain contains (see Fig.1.63): accessibility, availability, acceptability, or security-related items. But it is obvious that the whole set concerns both doctors and patients. The choice of the class QP for question relative to those concepts answers the question of who is most concerned with the category described. The distribution of the abstract coding of the QP Patient's issues shown in the Fig. 1.63

reveals an interest of doctors in what patients know. Patient's knowledge: 21 codes, patient satisfaction: 21 codes, patient assessment: 15 codes and their health habits: 19 codes. Accessibility to health care: 15 codes and safety of therapeutic processes: 10 codes are also well documented.

FIGURE 1.63: Distribution of 152 Q-Codes QP Patient's issue in WONCA Conference 2007. On 998 abstracts



The almost total absence of abstracts related to the health of the planet in Fig. 1.62 is astonishing. In 2007, the issue of global warming is not regarded as a major influence and subject worthy of respect. Yet, the importance of the environment as a determinant of health is well known, but GPS do not talk about it with the level of urgency it demands.

- About a Particular Code, the QC32 Migrants.

  The analysis of details in knowledge acquisition can be rich (Fig.1.64). During the WONCA 2007 conference in Paris, twenty three participants exchanged their experiences with migrants. The concurrent codings show the prevalence of poor health conditions and the influence of social problems, which are reflected in the analysis of the communications. Three abstracts address process issues (Imaging, Immunization, Blood analysis). Precision of information could potentially be poor, as we have no precise indication in the abstracts expressing what kind of psychological problems are discussed. In this case, coding at the level of chapter of ICPC could be necessary. Discussion about patients, who have many psychological problems without precise description, were coded P (for ICPC-2 chapter P, psychological). Nutritional problems, addressed as a whole, were coded T (for ICPC-2 chapter T). Accepting clusters like T89/T90 was sometimes necessary, as the information about the conditions was often unclear in the abstract (T89 insulin dependent, T90 non insulin dependent).

FIGURE 1.64: 23 abstracts with the Q-Code QC32-migrant out of 998 abstracts of WONCA 2007. Comorbidity is shown by the ICPC-2 codes identified.

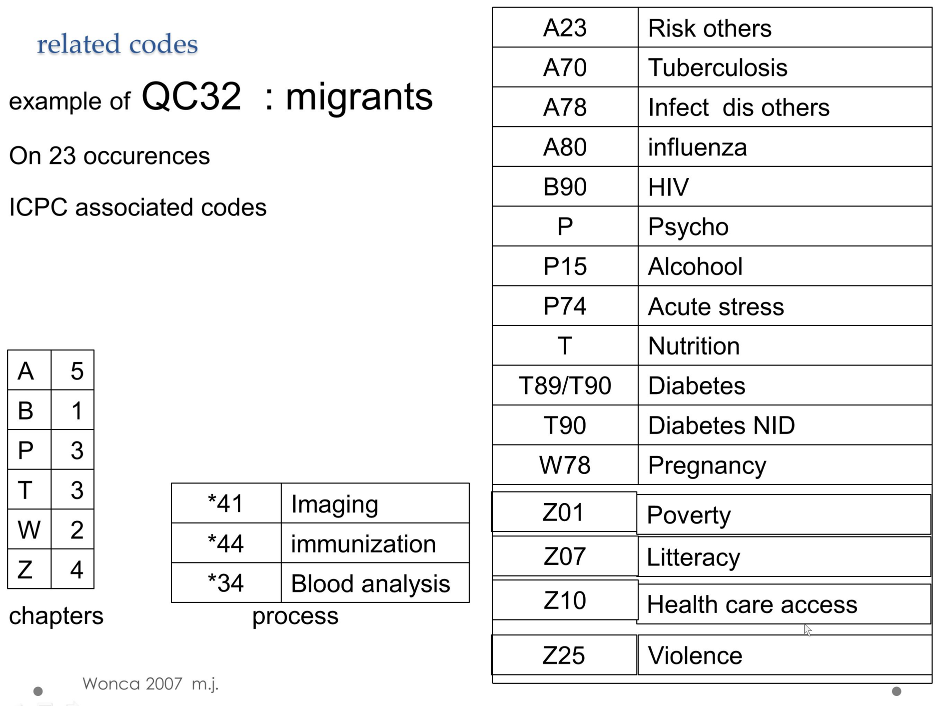
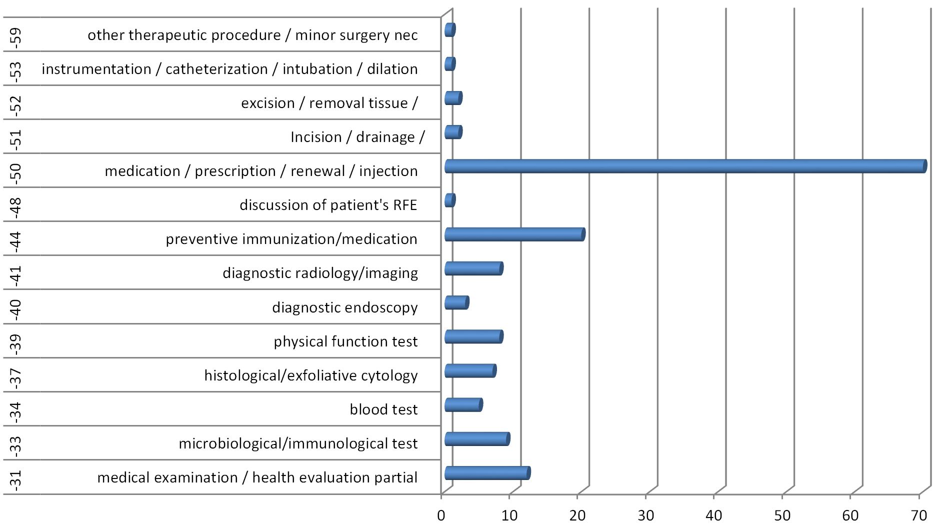


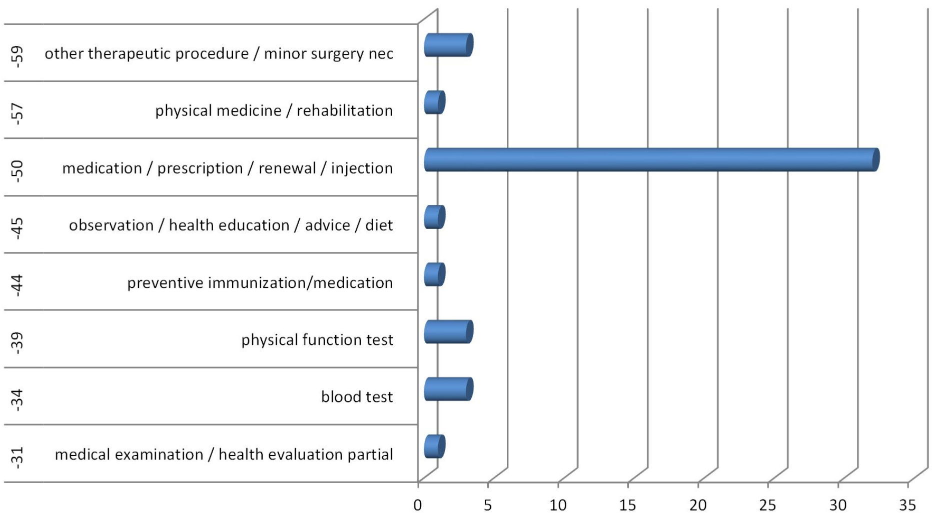
FIGURE 1.65: On 872 abstracts from the WONCA Conference, Paris, 2007. Distribution of abstract addressing procedures. (codes ICPC-2 Process)



1.4.3.2 Comparing Processes Between Abstracts from WONCA 2007, Portugal 2013 & Lille 2014

When in Covilhã, invited by APMGF to the Portuguese 18th National Conference of Family Medicine, September 28, 2013, the author was granted the opportunity to read the *Livro de Resumos* (book of abstracts) and analyze 128 abstracts written by young, enthusiastic GPs. These abstracts were then indexed with Q-Codes version 2.0 and ICPC. Analysis of these abstracts was

FIGURE 1.66: On 128 abstracts from the Covilha meeting, Portugal, 2013. Distribution of abstracts addressing procedures (codes ICPC-2 Process)



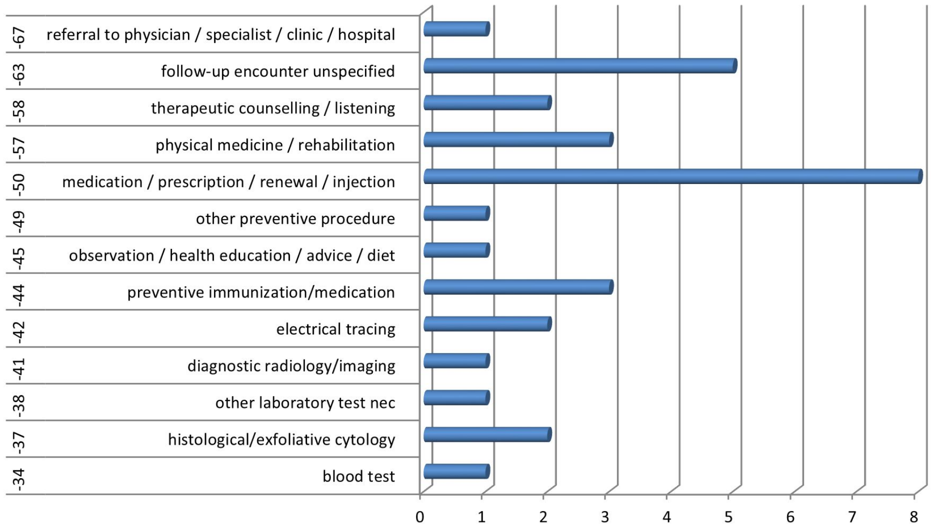
presented to the general audience as a wrap presentation at the end of the congress. This experience was related in the Portuguese Journal of General Practice (RPCG), and this paper is reproduced at the chapter 3, page 137.

A shared particularity of the doctor assisting in all three conferences analyzed is the discussions about drugs. In ICPC-2, drug prescription is a process identified by the code -50. GP commitment to the discussion of drugs is shown in the three following tables.

The data shows that, during the WONCA conference in Paris (Fig. 1.65), the most discussed procedures were over drugs-i.e., drugs prescription, drugs indications, drugs related problems. This issue is present in 70 abstracts on 872 (16%) analyzed. In 128 abstracts presented in Covilha 2013, (Fig. 1.66) 42 (32%) deal with process issues. 32 of which are drug related. This means that 25% of the 128 abstracts addressed drug issues.

The Lille congress (Fig. 1.67) offers a similar picture, though French GPs tend to discuss less patient related issues like procedures; 33 (15%) abstracts addressed processes of which eight (3%) abstracts are indexed with 50 drug related problems, five by follow-up encounters (code -63), three for vaccinations (code -44) and two for Pap-smears (code -37). Thus, processes are

FIGURE 1.67: On 211 accepted abstracts of the CNGE Lille congress 2014. Distribution of abstracts addressing procedures (codes ICPC-2 Process)



discussed in varied ways (WONCA 16%, Covilhã 32%, Lille 3%). Nevertheless, drugs are always at the top ranking of process discussion. The low proportion of processes discussed in Lille could be explained by the nature of the conference, which was dedicated to Training and Teaching.

This discussion on drugs could be considered as a common GP/FM task. Yet, the marked exclusivity of one procedure over the other shows a certain selection bias. The procedures used in GP/FM are numerous and varied, and much more research could be done on various kinds of processes. A qualitative study might be appropriate in understanding the participant's fixation on drug discussions.

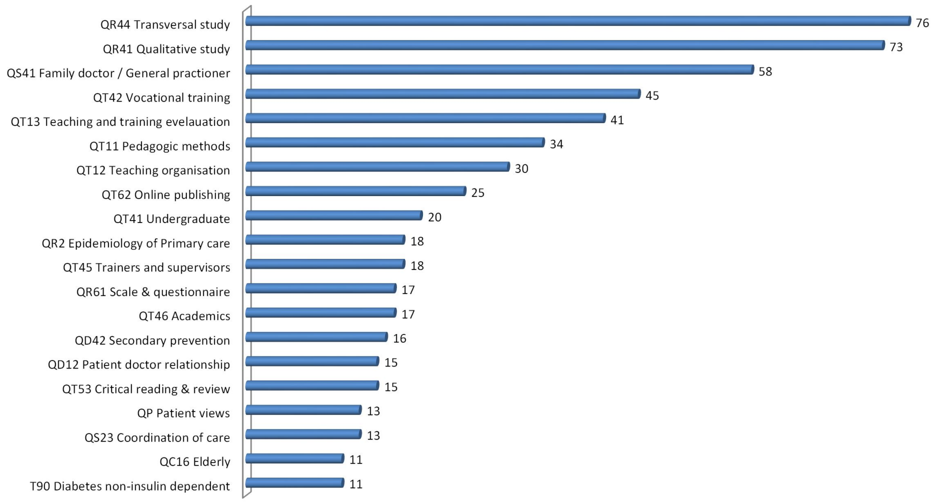
1.4.3.3 Analysis of Data from CNGE 2013 and 2014

The submitted abstracts from the CNGE conference held in Clermont Ferrand (France) in 2013 and of the CNGE congress held in Lille (France) in 2014 were analyzed before the Lille Congress 2014. The results were presented as keynote, introducing the conference.

- Top Twenty Descriptors at the Lille Congress 2014.
  The figure 1.68 gives the distribution of the codes of abstracts presented in Lille 2014. The Top 20 descriptors used in the 211 abstracts show that the first patient clinical issue (diabetes) comes in 20th place. We learn that in research methods, the first is the usual observational transversal study (QR44), followed by qualitative research methods (QR41). Unsurprisingly, the general practitioner as a theme of study

is in third place (QS41). Vocational training (QT42) and undergraduate (QT41) are in the top ten. In this congress, it's all about research and teaching. QR and QT have the first rank.

FIGURE 1.68: Top 20 descriptors Lille 2014

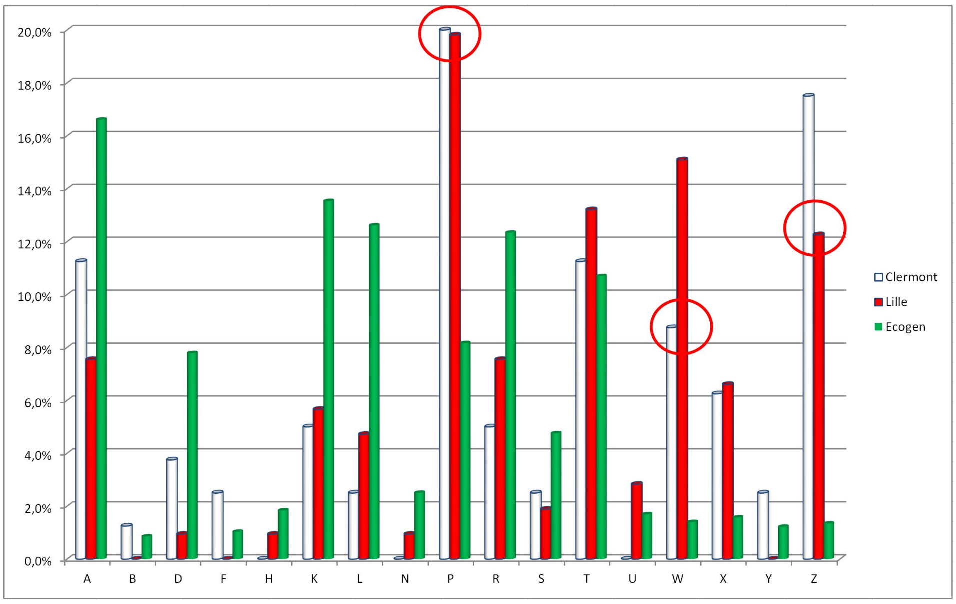


- Comparing Discussed Themes and Daily Encountered Problems
  A comparison is done between the distribution of clinical problems coded by ICPC-2 during CNGE congresses 2013-2014 and during clinical encounters (data produced in France in the ECOGEN project in 2013, courtesy of Prof. Letrilliart). Thus, discussed themes during congresses (CNGE) against found problems in the day to day work (ECOGEN) are compared.

In the Fig. 1.69 the white columns are the ICPC-2 codes from the Clermont conference of the CNGE (France) in 2013; the black columns (red for on-line readers) are the ICPC-2 codes found in the abstracts of the Lille congress of the CNGE (France) in 2014; the dark grey columns (green for on-line readers) are the data collected from the ECOGEN project, France, 2013. The ECOGEN project gathered data, for three months (12/2011 to 04/2012), from 45.642 clinical encounters between 20,613 patients and 128 GPs throughout France (Gelly et al., 2014). The participating GPs are assumed to be from the same background as trainers in GP/FM are in both case the participating doctors, members of CNGE.

Thus, the two first are conference data and the third is clinical data. The first two are subjects discussed by GPs, and the second is their daily experience. First, let's say that differences exist between the 2013 and 2014 conferences of the French Society of teachers. Curiously, in 2014, no case of B (hematology), D (Digestive), F (eye) and Y (genital male) were addressed while the chapters are always present in Clermont 2013. The chapter T (nutrition), W (pregnancy) and X (genital female) are in the same range for 2013 and 2014.

FIGURE 1.69: Comparison of Data from Conferences and Data from Practice; ICPC-2 coding of two CNGE congresses (2013 & 2014) and from clinical encounters in GP practice (2014). Discrepancies between data issued from practice (green) and data discussed in conferences (white and red) are pointed with a red circle. (ECOGEN data, courtesy of Pr.Letrilliard)



Comparing problems discussed versus problems encountered illicits quite exciting results. First of all, the general distribution is quite different. Only the chapter T has the same amount of quotes in the three sources. This is mainly due to the T89 & T90 (Diabetes). Someone used to examine such ICPC-2 curves classed by chapter would say that the Ecogen clinically based data reflect the general landscape of GP/FM with more A (general), a lot of K (circulatory), R (respiratory), L (locomotor), T (nutrition), some P (psychology) very few W (pregnancy), X (genital female) and very few Z (social) (green or dark grey in the Fig. 1.69).

The general picture is different when one examines the abstracts of the communications in congresses in 2013 and 2014 in the same country and you make an attempt to answer to the question ; what are they discussing about?

The answer is, without discussion, that those physicians discuss the psychological and social problems of their patients first. Much more than retrieved in the data issued from their daily work. But, conversely, they are discussing much less Digestive (D), Circulatory (K), Locomotor (L) or Respiratory (R) problems than retrieved in their daily work. We are in France, it is a teacher's congress for which Balint (Balint and Valabrega, 1960) and Freud's legacy (Westen, 1998) are very important. The psychoanalytic orientation of many French practitioners may have a role to play. But this difference is still remarkable and worth exploring by other qualitative research. Remarkably, when one examines the refused abstracts, we found the same proportions of P, W and Z refused. So it can't be a reviewer bias.

Its difficult to explain the propensity to discuss pregnancy and family planning problems (W chapter) in both congress with so low numbers of encounters dealing with this issue when in clinical settings. This is also worth of another study.

1.4.4 Use of 3CGP in indexing systems for humans

All the applications described below are published on the site `http://3cgp.docpatient.net`.

1.4.4.1 The universe of GP/FM at a glance

This extract of a mail received in 2017 from a Brazilian doctor in the second year of his GP/FM internship, who intended to visit Europe, highlight the interest of Q-Codes in the teaching process and as table of content for GP/FM.

> *I already heard about the Q-Codes but now I could understand how it works. My domains of interest are: QC32, QD12, QD15, QD26, QD323, QD44, QP21, QR7, QS11, QS12, QS33, QT11 and QT12. I know it's a lot of issues but I hope I can learn a little bit in this internship.* (Rebolho R, personal communication)

For the novice, here is an explanation of this varied interests of this trainee in GP/FM, which could be accurately communicated through his use of Q-Codes:

> QC32; refugee, QD12; doctor-patient relationship, QD15; motivational interviewing, QD26; palliative care, QD323; shared decision making, QD44 ; quaternary prevention, QP21; accessibility, QR7; economics, primary health care, QS11; management of practice, QS12; economy of practice, QS33; coordination of care, QT11; pedagogic method, QT12; teaching organization.

1.4.4.2 Bibliographic and indexing tools

- Teaching bibliographic retrieval to undergraduate and vocational training.
  As a GP/FM trainer, the author has had several students during the

last two years who have been interested in using Q-Codes as a bibliographic tool to prepare their master thesis. During the consultation Q-Codes are also a good way to draw the trainees' attention on the contextual knowledge that they must have to carry out a consultation. Few of them knew the concepts of accessibility, even less cultural accessibility, organizing referral or coordination of care. The interested reader will be able to read how using Q-Codes to highlight context of a consultation available here;
http://3cgp.docpatient.net/teaching-with-q-codes.

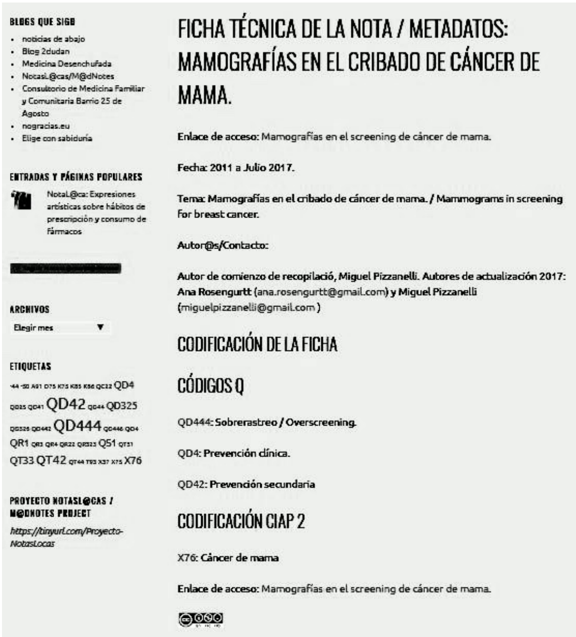
- Analyzing content of a book about GP/FM
Q-Codes have been used to analyze the content of a book in GP/FM. On the occasion of 20 years of WONCA Europe, an editorial committee selected 100 of the best abstracts edited in the past 25 years and has asked each author to make a chapter for the WONCA Book published in 2015 at the WONCA Europe congress. The 100 chapters of the WONCA Book 2015 were analyzed using 3CGP by the author who was involved in the editorial committee of the book. The results show that the editorial board has favored themes that affect the creation and the organization of knowledge into the profession. Clinical items (7 ICPC codes on 150) or affecting directly the patients issues (25 QP codes on 150) are poorly represented. Through this particular lens, it appears that the WONCA editors have chosen more academically than clinically oriented subjects. See the full poster at http://orbi.ulg.ac.be/handle/2268/18577.

- Indexing Tunisian GP literature in PubMed.
At the Tunisian general practitioners' congress, in Tunis in 2015, a study of the image of Tunisian family medicine was carried out through the publications cited in PubMed. On September 25, 2015, the MeSH "Tunisia" gives 6022 indexed publications. By filtering these 6022 publications with the search strategy constructed for the Q-Codes QS1 (see Appendix E, page 209) we have obtained 58 publications dealing with PHC including 45 with abstracts. From 1985 to 2015 there are 37 abstracts in French, 8 in English. These abstracts were analyzed with ICPC-2 and Q-Codes. They are 12 without ICPC-2 codes i.e. not related to clinical issue and 33 contains at least one ICPC-2 code. The general picture is dominated by cross-sectional studies on diseases and management of practice. A short qualitative study was carried out on these abstracts. The whole set of abstract shows a profession decried and minored and which lacks means for its development (Jamoulle, 2015b).

- Indexing virtual repositories in GP/FM in Uruguay.
  - The virtual bibliographic repository of the Quaternary prevention Special Interest group is maintained using Q-Codes as indexing system: see https://goo.gl/VTnJjH (Pizzanelli, 2015)

– The virtual bibliographic repository of the Notas@Locas collaborative project dedicated to gather exchange between South American GPs is maintained using 3CGP as indexing system (Fig. 1.70). See https://notaslocasmadnotes.wordpress.com, https://tinyurl.com/Proyecto-NotasLocas and Twitter @RedMadNotes.

FIGURE 1.70: Example of use of 3CGP (Q-Codes & CIAP-2 [ICPC-2]) for indexing of grey literature.About mammographies. (In Spanish).



- indexing of second opinion requests from the Brazilian Telehealth system (ongoing work)

  This approach is now used by researchers on data from Pernambuco, Brazil, for initially manually indexing a sample of 550 questions; with an ultimate goal of semi-automated indexing of larger data sets, measured in the tens of thousands of question-answer pairs. These question-answer pairs originate from the Brazilian Telehealth system, representing communication between rural health care providers and nurses and doctors in the urban Telehealth centers. (Resnick et al., 2013).

1.4.4.3 Indexing master theses in GP/FM

- Coding new thesis in GP/FM. French speaking universities. Belgium.

  Q-Codes could be used as base for bibliographic search for master theses. The implementation of the indexing of the final work of the French-speaking family physicians of the Walloon Region (Belgium) has made a significant step forward in achieving the main objective of this thesis, namely to manage knowledge in family medicine.

The work of doctors in general and family medicine training is often of high quality, requires considerable investment from the authors and sometimes represents little-explored research areas.The experiment will begin in 2018. The system developed to index master thesis is reused by our Brazilian and Portuguese colleagues.

- Use of Q-Codes for indexing of theses, Coimbra, Portugal

  Authors : Ariana Tavares, student in medicine, Susana Medeiros, MD, Luiz Miguel Santiago MD, PhD. Especialista em MGF, USF Topázio, Professor Associado Convidado da UBI, Coimbra, Portugal. (Personal communication. September 25, 2017).

  In The Coimbra's Faculty of Medicine, the master theses are being indexed by the Q-Codes classification. It is an amount of about 300. The work is Ongoing. The Fig. 1.71 shown the coding process for some titles (in Portuguese).

FIGURE 1.71: Coding process of the Coimbra thesis. The titles of the thesis are shown with chosen Q-Codes and ICPC Codes. (work in progress)(Working data, courtesy of Ariana Tavares. October 2, 2017)

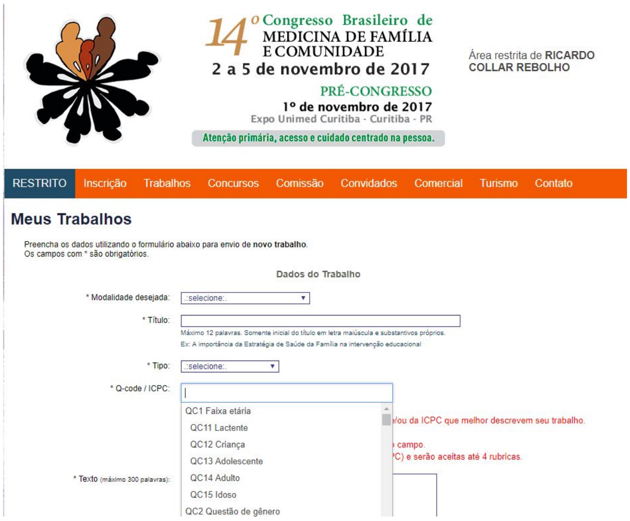
Title of the theses	ICPC1	ICPC2	ICPC3	Q-Code1	Q-Code2	Q-Code3	Year	Gende
Atendimento jovem - consulta diferenciada do				QC13	QD11	QS1	2008	F
Realização pessoal, sexualidade, depressão na	P76			QC13	QP52	QP43?	2008	M
A adesão anti-hipertensiva e a sua preponderância no controlo da doença	K86	Z11		Z11	QD31		2008	F
Redução do risco cardio-vascular global através da continuidade de cuidados assistenciais prestados a uma população hipertensa pelo	K	K86		QD31	QD25	QS41	2008	F
Ruídos de comunicação, as interrupções nas consultas – a visão do utente				QD11	QP25	QP41	2009	F
Quando suspeitar de mononucleose infecciosa – um caso clínico atípico	A75			QR36	QD32		2009	F
Os rastreios: cancro do colo do útero, cancro da mama e cancro do cólon e recto	X75	X76	D75			QD42	2009	M
Novas abordagens no rastreio do cancro da	X76					QD42	2009	F
Avaliação de qualidade do rastreio do cancro do colo do útero nas utentes do Centro de	X75				QS1	QD42	2009	F
Adesão à vacina contra o cancro do colo do	X75				QP43?	QD41	2009	F
Opções terapêuticas no tratamento da dor crónica na população idosa	A01			QC15	QD32		2009	F

1.4.4.4 3CGP and the organisation of congresses.

- Indexing the congress of SBMFC, Curutiba, December 2017.
  Authors : Daniel Knupp, Paulo Poli Neto, Gustavo Gusso (Brazilian Society of Family and Community Medicine (SBMFC) with the support of Melissa Resnick and Frank Shamenek (USA) for the statistical analysis (Personal communication. October 2, 2017).
  The 3CGP coding system (see Fig. 1.72) is in use at the deposit page of the 14th Congress of the Brazilian Society of Family and Community

Medicine (SBMFC). The participants have to chose at least two codes and maximum 4 codes of ICPC-2 and of Q-Codes. This experiment is underway.

FIGURE 1.72: 3CGP as coding system for abstracts, filled out by the contributors to the 14th SBMFC congress in Curutiba, Brazil. Q-Codes in Portuguese. http://www.cbmfc2017.com.br/trabalhos/



The 3CGP indexing system was chosen to index work presented by authors at the time of the on-line repository. First data was available at the beginning of October 2017. They are 1746 abstracts accepted of which 384 oral presentations and 1362 posters. On 1746 abstracts 755 ICPC codes have been used and 3.424 Q-Codes. On the 384 oral presentation 997 Q-Codes and 186 ICPC codes have been assigned.

First results are shown in the Fig. 1.73 & 1.74. The general frame of the distribution of 755 ICPC codes (on 1,746 abstracts) shows the interest of GPs to discuss P & Z problem, ie psychological and social problem as was the case in all the congresses analyzed so far. Comparison with Fig.1.69, page 105 shows the same interest for P& Z expressed in all congresses. the relative scarcity of ethical themes (QE) and Environmental problems (QH) is striking and similar to what was found in previous congresses. They are 44 QO codes (Other).This means that the communicators would not have found the adequate subject for their communication in 2,5% (44 on 1746) of the cases. This will be analyzed in depth. This analysis will be presented to the participants of the congress.

- Organization of congress in GP/FM. France
  An application for managing scientific congresses, especially general practice congresses is highly desirable. This implies that the user can consult the schedules of a conference, the subject, the number of the room, etc. In addition to provide practical information about

FIGURE 1.73: Distribution of 755 ICPC codes in 1746 abstracts accepted (384 oral presentations and 1,362 posters) (SBMFC – Curutiba – 2017) Importance of A (prevention), P (psychological), T (metabolic and nutritional) & Z (social problems) in the themes discussed.
(Data: courtesy of Daniel Knupp, SBMFC)

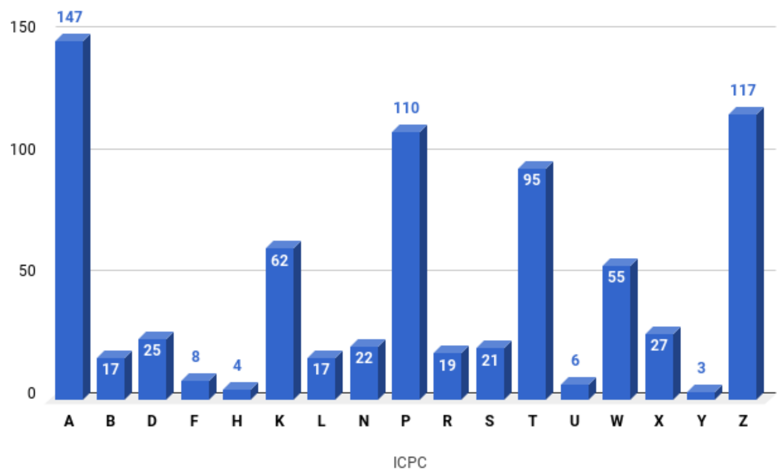
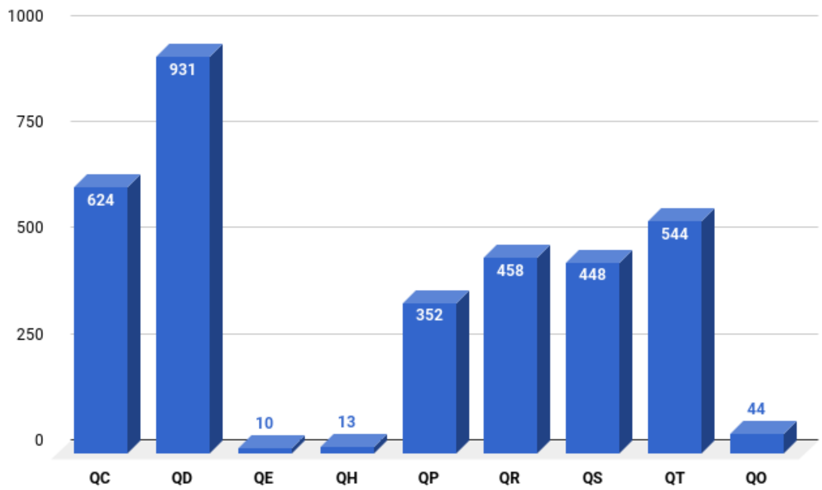


FIGURE 1.74: Distribution of 3,424 Q-Codes in 1746 abstracts accepted (384 oral presentations and 1362 posters) (SBMFC – Curutiba – 2017). Scarcity of ethical themes (QE) and Environmental problems (QH).
(Data: courtesy of Daniel Knupp, SBMFC)



each conference, the application must be able to reference authors, abstracts, and various documents based on keywords or on 3CGP (ICPC-2 & Q-Codes). The Fig. 1.75 shows a proposal for an advanced search application (in French) including 3CGP.

The goal is also to provide search functionalities to the user so that he can access the content of his choice in a quick and simplified way. The trainees of Graduate School in Engineering ESAIP (https://www.esaip.org/en/) have carried out an end-of-study work on this topic under the direction of Mr Chistian Simon of SILK Informatique (http://www.silk-info.com/). The trainees produced a report (in French) that shows the steps necessary to carry out the development project of an ICPC-2 and Q-Code based general practitioner congress application. The full report is available in French at https://goo.gl/w92MGr (Missie et al., 2015).

FIGURE 1.75: 3CGP as coding system for abstracts could be used as a query system in congresses. Proof of concept. Done by the trainees of the School of Engineering, Angers, France. (In French)

1.4.5 Use of 3CGP by computers

The use of 3CGP by machines is still in development. The knowledge base specific to family medicine, assembled in HeTOP, is ready for future technological development in both natural language processing and for the semantic web. Here we present the first promising steps of a still long road.

1.4.5.1 Multi-Label Text Classification of Medical Abstracts, 2015

Author : Sébastien Rigaux. University of Liège, Belgium.

The objective of the work of Sebastien Rigaux (Rigaux, 2015) was to find an automated method capable of analyzing the contents of non-clinical general practice articles and of predicting the corresponding Q-Codes categories. The main difficulties arise from the small amount of sample data available, the large number of categories to be identified, and the high specificity of the scope of Q-Codes making categories difficult to discern.

The first step was to compare the efficiency of the methods of vectorization and standardization (Bag-Of-Words, Term Frequency, Term Frequency-Inverse Document Frequency ...) by combining them with those of automatic natural language processing (root- lemmatization and filtered lemmatization according to the nature of the corpus words), and with Feature Selection and Feature Extraction (Chi-Squared, Bi-Normal Separation, Mutual Information, Cube Mutual Information, Singular Value Decomposition). This step has shown that filtered lemmatization, TF-IDF and SVD bring real gains of accuracy to classifiers.

The second step was to analyze the effectiveness of these various techniques, depending on the Machine Learning algorithms used to see if methods such as Bernoulli Naïve Bayes, Multinomial Naïve Bayes or Stochastic Gradient Descent can improve the classification performance obtained at Support Vector Machine support. As a result, only SGD equalizes and even surpasses SVM.

The third step was to compare the previous results, obtained from the French corpus, with similar classifiers applied to an Anglophone corpus. This made it possible to verify that the effectiveness of the different techniques studied is globally independent of the language used. However, the gains in precision differ somewhat for certain techniques such as SVD, TF-IDF. Similarly, it has been observed that the English corpus is less sensitive to methods of automatic natural language processing.

Once all these analyzes and comparisons were made, two Q-Codes classifiers could be proposed. The first, for the Francophone corpus, uses SVD while the second uses SVM, TF-IDF and SVD. The two classifiers also use filtered lemmatization, and obtain a modest F1-score of 0.452 and 0.344 respectively. Full work available in French on; `https://tinyurl.com/yc5ej2bw`

1.4.5.2 Automatic coding with ECTMV3, France

Author : Chloé CABOT, D2IM, University of Rouen, France

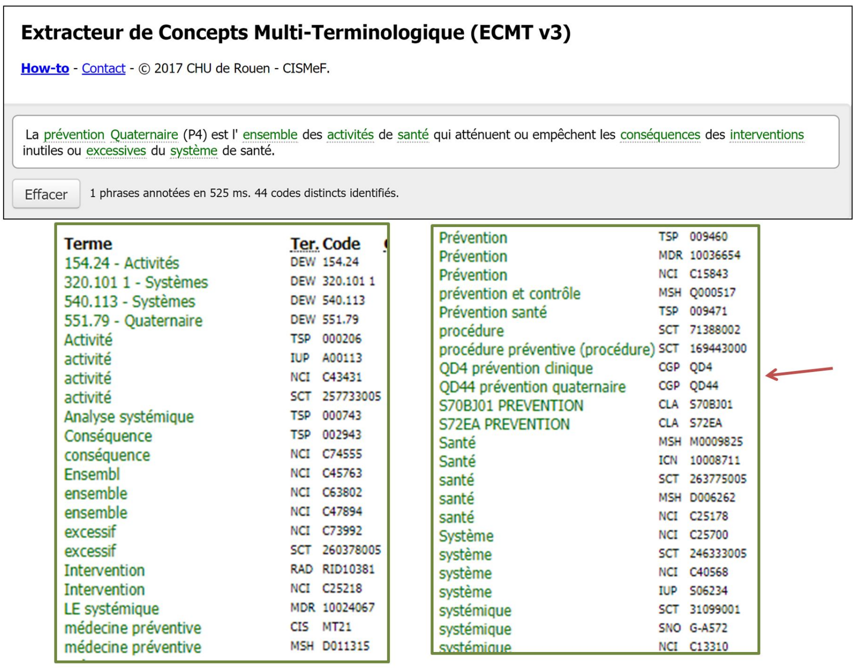
ECMT (Extracting Concepts with Multiple Terminologies) is an annotator and a web-based service developed at D2IM, Rouen. It aims to fully automatically identify clinically relevant entities in medical texts in French with several types of documents: abstracts titles, documents about marketed drugs and death certificates (Cabot, 2016).

The extraction is performed at the phrase level of the text. ECMT has also a user-friendly interface accessible after authentication.

(`http://ecmt.chu-rouen.fr/`). The pertinent terms are retained based on the HeTOP resources. The Fig. 1.76 gives an example of processing the phrase : *La prévention Quaternaire (P4) est l'ensemble des activités de santé qui atténuent ou empêchent les conséquences des interventions inutiles ou excessives du système de santé. [Quaternary Prevention (P4) is the set of health activities that mitigate or prevent the consequences of unnecessary or excessive health system interventions.]* ECMT extracts the terms in French from terminologies in HeTOP like Medical Subject Heading (MeSH) or National Cancer Institute terminology (NCI) as well as from Q-Codes terminology (CGP).

ECMTV3 has been used to analyze the coverage of 32 terminologies available in French on HeTOP on the French medical corpus LiSSa (http://www.lissa.fr/)(Cabot et al., 2017b). LISSa is a search engine which uses the same facilities as CISMEf and references 1,067,382 scientific articles in French in the field of Health (on August 17, 2017). Adapting ECMT to index French congresses in GP/FM with ICPC-2 and Q-Codes could be an interesting challenge.

FIGURE 1.76: Automatic extraction of concepts by ECMT v3. by numerous terminologies such as MeSH (MSH), National Cancer Institute (NCI), MedDRA (MDR), SNOMED (SNO) etc. The red arrow shows the automated identification of concepts in Q-Codes (CGP); QD4 Prevention and QD44 Quaternary prevention.(in French)



1.4.5.3 Q-Codes in WebProtégé

Protégé is a free, open-source ontology editor and framework for building intelligent systems; (see https://protege.stanford.edu/. We have already described this editor in the Method section at the paragraph 1.3.8.3 and Fig: 1.57, page 89. Protégé's plug-in architecture can be adapted to build both simple and complex ontology-based applications. Developers can integrate the output of Protégé with rule systems or other problem solvers to construct a wide range of intelligent systems.

1.4.5.4 Using Q-Codes in an e-learning program, Vietnam

Author: Thành Liêm Võ, Phạm Ngọc Thạch Medical University, Ho Chi Minh City, Vietnam. (Personal communication, 2016)

The family medicine unit of the Phạm Ngọc Thạch Medical University, Ho Chi Minh City, Vietnam (http://www.pnt.edu.vn/vi/) has incorporated the Q-Codes in Vietnamese in an e-learning system for medical students. For the first time we have a glossary of Family medicine (FM) terminology who helps us to understand and standardize the complex concepts of this discipline: between the countries and between the regions in Vietnam. It reduces the variety of these interpretations. For now, we are preparing new curriculum of FM ; 3 – month – training program for doctors which will be applied at national level, and the new 1-month program for FM nursing will be developed. In this direction, Vietnamese version of Q-Codes has been used as a source of reference in the working of Belgo-Vietnamese FM expert committee in Vietnam. The Q-Codes are important to introduce the new concept of FM through these definitions and references. Q-Codes will be used as the reference resource for future research in FM. For now, Q-Codes have been integrated in format of glossary in 3 month FM training at Phạm Ngọc Thạch Medical University. In the below screen copy the term *dự phòng* means *clinical prevention* is mapped to the Q-Codes QD4. The corresponding definition appears as a pop-up when pointing the arrow on the term.

FIGURE 1.77: Language Processing in e-learning in Vietnam. The term *dự phòng*, meaning Prevention and corresponding to the Q-Codes QD4 Clinical prevention, is automatically tagged and linked to the definition. (In Vietnamese)

1.5 DISCUSSION

1.5.1 Summary of the research questions

We have defined three specific aims and resulting research questions. In the following paragraph we will attempt to show that we have answered to our research questions and address some limitations.

1.5.2 Identification of knowledge produced by GPs

1.5.2.1 Can we identify the knowledge gathered?

The association of the International Classification of Primary Care in its three components; Symptoms, Procedures and Diagnostics with the Q-Codes forms an indexing system of 1252 items. This system therefore covers clinical and contextual elements specific to general and family medicine. Thus, we have a system that allows us to identify patients' symptoms and complaints, diagnosis or disease hypotheses, processes used by physicians, either by themselves or by third parties, and finally the context of application given by Q-Codes. Using the respective acronyms one can write:

ICPC + ICPC Process + Q-Codes = 3CGP

Although Q-Codes are not exclusively medical, but also managerial, they represent a form of controlled medical, multipurpose vocabulary that is subject to further additions. As stated by Cimino (1998) *the unit of symbolic processing is the concept - an embodiment of a particular meaning*. Q-Codes are a medical subject authority list, with a comprehensive series of mutually exclusive concepts. According to guidelines set by Cimino, we have tried to gather a set of *non-redundant, shareable, multipurpose, high-quality permanent concepts, in a mono-hierarchic organization, identified by a set of definitions and linked to existing terminologies*.This set allows to identify knowledge as expressed by GPs from several countries.

1.5.2.2 Can an ontology for GP/FM be devised?

This study proposes a system of Knowledge Management (KM) in GP/FM. By complementing the clinical classification ICPC-2 with a new professional contextual one, the Q-Codes. The Core Content Classification (3CGP) could potentially fill a major gap in knowledge management of GP/FM. Conceived as a multilingual lightweight ontology fit for new Internet technologies, natural language processing and the Semantic Web, 3CGP gives the opportunity to unravel GP/FM productivity and establish GP/FM as a professional discipline worthy of respect.

Though this project took years of work, it acts only as a base from which future researchers may expand. As it was designed according to terminological and ontological principles, is available in web ontology language (OWL) and ready for use with Linked Data. The set of ICPC-2 and Q-Codes is a light-weight ontology; however, because it adapts Natural Language

Processing and is exportable to a Web Ontology editor it could serve as the basis for the development of a real ontology in GP/FM. But the path to a real ontology is still a long time in the making.

1.5.2.3 Can the detection of terms be automated ?

To find an automated method capable of analyzing the contents of non-clinical general practice articles and of predicting the corresponding Q-Codes categories is reachable but need additional work. We have prepared the track for automated textual analysis and the semantic web. Automatic annotators are still available of which one is ready for Q-Codes research in French. Nevertheless much work is needed to reach an automation of text analysis specific to GP/FM and multilingual concepts recognition. First of all we need to populate the lightweight ontology obtained and design a real ontology in GP/FM.

1.5.3 Towards a draft of the table of content in GP/FM

1.5.3.1 Is it possible to produce a table of content in GP/FM?

The aim of this research was to identify main themes discussed in abstracts for family physicians during conferences. Clinical concepts discussed in the conference abstracts could be retrieved the through International Classification of Primary Care (ICPC). The modes of GP/FM organization have also been identified in these abstracts, and they are organized under a new taxonomy called Q-Codes, which is an add-on to ICPC-2.

To the best of our knowledge, there is nothing similar available that has been developed for both human and machine use. There is also not anything of this measure that demonstrates the complexity of GP/FM. Due to the overlap GP/FM with the first line of health service, this tool could also be useful in Primary Care. All doctors and health managers, for whom proximity and health management are of utmost importance, could potentially reuse Q-Codes for their needs. Once the Q-Codes had continued to evolve into a true ontology, with more precise specifications, the combination of the Q-Codes with the ICPC classification, resulting in 3CGP, could be considered as an extended table of content for GP/FM teaching and training.

1.5.3.2 What knowledge representation formalism in GP/FM ?

The Q-Codes have been developed as a lightweight ontology with complex relations and properties. The same methodology has been in use for ICPC (Cardillo, 2011). We are thus developing a light weight ontology, the simplest form of an ontology, under the name of 3CGP.

We have chosen the lightweight ontology as a representation formalism, as a first step to prepare an ontology in GP/FM, an objective out of reach for us. Nevertheless all the knowledge gathered could be exported in various computer languages, for human (CSV, Excel) or machine use (OWL) and thus further development.

1.5.4 To propose an experimental indexing system fit for GP/FM

1.5.4.1 Is it possible to produce a new indexing system?

We have questioned whether the visibility of the productions of family doctors has so far been hampered by insufficient indexing. The knowledge management (KM) tools, developed for identification of GP/FM grey literature, have attracted the interest of GPs around the world. Clarification of concepts for pedagogical or bibliographic purposes, indexing of dissertations or conferences, and semantic analyses of on-line discussions of group of family physicians are all demonstrations of the need for KM in the GP/FM field. We have developed a new taxonomy as an initial stepping stone to address KM in this field, We hope this will contribute to unlock the potential of the huge amount of knowledge gathered by GPs.

1.5.4.2 Does this allow for visibility of GP/FM grey literature?

We have allowed ourselves to ask if the visibility of the productions of family doctors has so far been largely ignored. The knowledge management tools, developed for identification of GP/FM grey literature, have attracted the interest of GPs around the world. Identification of concepts for pedagogical or bibliographic purposes, indexing of theses or conferences, and for usually lost online discussions of group of family physicians are all demonstrations of the need for KM in the GP/FM field. There has been a definite need exhibited, and this work has proposed an initial stepping stone to address it. We are confident our work will disentangle a huge amount of knowledge gathered by GPs.

1.5.4.3 Could this produce a link between knowledge and networks of national and international doctors/researchers?

English has always been the fall-back language of GP/FM. However, family doctors speak to their patients and with one another in their own language, which leads to confusion in translation and various vocabularies. Having tools that facilitate the use of multiple languages, while simultaneously communicating the same concept without variation in context or understanding is incredibly important and useful/necessary for GPs. Having a tool that accommodates so many specific mother-tongues may explain the enthusiasm of so many international colleagues that wished to participate in this multilingual edition.

First steps are ongoing to develop an online corpus of knowledge for GP/FM linked by an unique indexing system through the indexing of various knowledge bases specific to GP/FM worldwide. Texts identified by a common indexing system can now benefit from the same metadata system. These can be expressed according to semantic web standards. Everything is ready for a consistent use of the Internet as a distributed database in the service of family medicine. The Linked Open Data universe opens in front of us.

1.5.5 Similar studies

1.5.5.1 In the footsteps of FAMLI

In the 1980s, WONCA created a bibliography committee, led by Dr. Jan McWinney- a Canadian professor of General Medicine. From 1980-1992, this committee edited the Family Medicine Literature Index (FAMLI), an index that contains 12 volumes of literature specific to GP/FM (CFPC, 1980-1992). In turn, Dr. McWinney played a key role in establishing GP/FM as a respected profession, among other by illustrating the importance of GP/FM literature documentation. We have followed in the committee's footsteps by applying Internet techniques to this type of indexing. By using new, innovative technologies, we have made a readily available tool that is adapted to GP/FM management of specific information fields/compatible with the MeSH. FAMLI's index highlighted both the need an importance of GP/FM contribution to scientific literature.

1.5.5.2 Attempt to identify GP/FM fields

Many authors have attempted to identify, either implicitly or explicitly, the content of GP/FM. However, these attempts have often taken a top-down approach, resulting in incomplete fields of knowledge. We shall not recite here the innumerable GP/FM treatises and textbooks, which have been edited by experts from all countries, that do not give sufficient explanation for various elements of GP/FM. Though it lacks important factors, the top-down approach often prevails in the studies of experts who have looked at definitions of GP/FM.

Other researchers have developed bibliographic search filters (Brown et al., 2014; Gill et al., 2014; Pols et al., 2015). Unfortunately, these tools often confuse GP/FM and Primary Care services. Australian Primary Health Care Research & Information Service (PHCRIS) (`http://www.phcris.org.au/`) does this as well, though it does offer an extensive and outstanding information service for PHC working people. In Austria, Jelercic et al. (2010) created a very original and fruitful approach to identifying the specific MeSH of GP/FM. However, our research both expands and substantiates missing pieces that many researchers fail to address.

From a bottom-up approach, treating GPs as experts in their field of knowledge, Kruschinski et al. (2010) analyzed 614 abstracts presented at European General Practice Research Network (EGPRN) conferences. Kruschininski then established a classification for their areas of interest. Miller et al. (2005b) and Miller et al. (2005a) analyzed content of the first and second volumes of the Annals of Family Medicine in a move to characterize primary care research and to identify opportunities for new directions in categorizing GP/FM concepts. Both offer a quite interesting classifications of research methods. However, they fail to differentiate between GP/FM and PHC. In the Canadian Triple C (Core Competencies Curriculum) for GP/FM, numerous experts have addressed the scope of training and evaluation for family medicine residency (Oandasan, Saucier, and (eds), 2013). In an extensive document, they have made an attempt to exhaustively describe content of GP/FM and PHC. But again, it is a proposal of what must be and not an analysis of what is, speaking of hypothetical rather than actual realities.

There are also countless publications over both defining and applying ICPC-2 to the clinical description of GP/FM and PHC. In an original move, Schrans et al. (2016) analyzed contextual patient information and its influences over the clinical decision-making process in PHC. Person-related information (PeRI) is described in 33 PeRI codes, grouped in seven mutually exclusive code-families. This new contextual patient oriented classification is intended mainly for research purposes and for the attributes of patient-centered care. We could examine whether this approach is complementary to our.

We have expanded upon this by analyzing occupational contextual information, which forms the basis of GP/FM practice and management. We have also expanded upon this by using Q-Codes to edit information in an online database, compatible with The Semantic Web.

1.5.6 Illustration of the operational relevance of the findings

A team of researchers from the Foundation for Research and Technology Hellas (FORTH), Institute of Computer Science have developed a Semantic Web based concept annotator for medical sciences (Sfakianaki et al., 2015). It is a Natural Language Processing engine which can translate free text into targeted queries, automatically transforming a clinical research question into a request description. The machine will answer accordingly. (`https://www.ics.forth.gr/`). A set of examples clinical questions have been given, the first of which is reproduced here :

> *John has lung cancer and has been treated with carboplatin which is known for toxicology adverse effects. I would like to find literature and reference related to such events for the specific drug.*

The machine is able to recognize *lung cancer, treated, carboplatin* and *adverse effect* and link them to knowledge bases. We, as electronic medical record users, are waiting for this kind of machine to be readily available for daily use in computers. Nevertheless, our sentence as GPs may be structured differently. For example, a question may say something along the lines of:

> *John, a Maghrebian patient, has lung cancer and has been treated with carboplatin, which is known for toxicological, adverse effects. He has*

> *been very sick and is no longer willing to follow treatment. He is depressed and expresses fear that spirits have invaded his soul. He has visited me as his family doctor to explain the situation. I would like to find literature about patient knowledge, North African cultural background, compliance, coordination of care, motivational interviewing and the role of the family doctor in managing patient denial.*

The purpose of this citation is not to dismiss the outstanding work already done by FORTH. However, by adapting the example to GP/FM reality, we are willing to point out that GP/FM has a different field of knowledge and different needs that need to be dealt with. GP/FM does need marvelous machines like concept annotators, but these machines must be adapted to the different needs of the GP/FM field. This would not be possible without participating in a detailed conceptual analysis of the GP/FM profession, designing taxonomies as Knowledge Management tools, and taking the first step towards creating a GP/FM ontology. The present work wishes to show this , to propose a solution, and to take a step towards addressing this need.

1.5.7 Study Limitations

1.5.8 Researcher biases

The Q-Codes form the initial building blocks of classification of professional and managerial aspects in the GP/FM field. However, this approach has been filled with the personal experience of the author, which may lead to unintentional biases. As an empirical document, one has tried to change, fill in the gaps and modify this emerging taxonomy, using content analysis of GP/FM publications, experience of peers, critiques, and application to real work. Finding the MeSH's corresponding descriptors, and multiple searching and indexing exercises on published documents have also been a good way to verify the applicability of the taxonomy. Another medical expert, not involved in family medicine world may have classified those items differently. We argue that we rely on Grounded theory for qualitative research. This is partially true as the domain has not to be fully discovered and is interpreted along the personal knowledge of the author. But as stated by Corbin and Strauss (1990); *Those who use grounded theory procedure accept responsibility of their interpretative role [. . .] researchers assume the responsibility of interpreting what is observed, heard or read.* We didn't try to give an explanation to GP/FM phenomenon but, in a first move, to retrieve and interpret its specificity.

1.5.8.1 A single-researcher study

An important issue to address is that there was a seven-year hiatus in this research, shown by the dates of the conference abstracts analyzed. This was due to an extended illness of the main author. Despite this hiatus, research was eventually able to move forward. Any negative effects resulting from this hiatus may be offset by the fact that only one researcher analyzed the abstracts. Bradley, Curry, and Devers (2007), qualitative data analysis experts, argue that *a single researcher conducting all the coding is both sufficient and preferred.[...]. In such cases, the researcher is the instrument; data collection and analysis are so intertwined that they should be integrated in a single person who is the choreographer of his/her own dance[...] However, bias of said researcher could have influence over the collection of data and its analysis. Therefore, disclosure of the researcher's biases and philosophical approaches is essential.*

1.5.8.2 An extended team of co-researchers

The author may be considered an expert in the field of GP/FM. Intensive collaboration with experts from other fields-i.e, terminologists, computational linguists, documentary scientists, information system experts, knowledge engineers, medical informatics specialists, was the crucial ingredient in the creation of this lightweight ontology. Translators can also be considered as co-researchers. The search is about organizing knowledge around concepts and no longer around the terms of a particular language. English serves only as a working language. The sophisticated interface available on-line allows each translator to see the result of his/her work and compare it to other languages. At first, only the list of preferred terms was translated into several languages. It was the Vietnamese translators who proposed the first to translate the definitions to ensure the conceptual correction of their terms translation proposals. The translators also made it possible to refine certain concepts and adapt some definitions. It is in this sense that we can say that this research has become a network effort.

1.5.8.3 Potentially Eurocentric

Another limitation of this study is that the data is Eurocentric. This is due to the fact that the conference abstracts analyzed present work done mainly by European GPs. Thus, the Q-Codes concepts might not be fully representative of other geographical areas. This, in turn, may limit worldwide usability. Nevertheless, the fact that the Q-Codes have been translated into four non-European languages (Turkish, Vietnamese, Georgian and Korean) implies that the translators have found points of connection to their own culture in the proposed concepts. However, this illustrates that the globalization of GP/FM concepts are strongly influenced by the Occidental, Anglo-Saxon origins of GP/FM (Simon, 2009; Gutierrez and Scheid, 2002).

1.5.8.4 Validity and Reliability

Another potential issue is the validity of the identification process and the list of concepts generated. Validity is concerned with *whether a variable measures what it is supposed to measure* (Bollen 1984 cited by Adcock and Collier, 2001). Here, we deal with the identification of concepts in texts. Adcock and Collier (2001) also state: *Because background concepts routinely include a* *variety of meanings, the formation of systematized concepts often involves choosing among them*. They distinguish between a *consensual concept* and a *contested concept*. It is supposed that a text about *gender violence* will be identified with the corresponding concept *gender violence* by a reader. But, for more ambiguous terms, like *continuity* which is often confused with *permanence*, or more contestable concepts like *disease mongering* and *deprescription* which some colleagues may have no knowledge of, how does one proceed? This ambiguity may have played a role in the execution of this project.

Identifying concepts works off the supposition that the same universe of discourse or world of reference is shared. We have already demonstrated that definitions of GP/FM and PHC are often overlapping, which shows that many worlds of reference come into play. (Jamoulle et al., 2017c). Moreover, some definitions are more complementary to one another than others. Thus, measuring validity- moreover inter-observer validity- will be a difficult task. For even in the same country, two GPs don't always share the same world of reference. Therefore, despite the fact that we are convinced that the bottom-up approach is the preferable method, there is no denying that, by defining a hierarchy of concepts and choosing the best definition through this approach, we are subject to the influence of personal values and beliefs.

There are as many definitions of *validity* in qualitative research as there are authors. *Face validity*, in quantitative research, is defined as *the extent to which a test is subjectively viewed as covering the concept it purports to measure.* (Holden, 2010). Noble and Smith (2015) propose a new terminology and criterium to evaluate the credibility of research findings. Usual terms used in quantitative research such as *validity*, *reliability* or *generalisability* are replaced with *Truth value, Consistency* and *Applicability*;

- *Truth value (validity); Recognizes that multiple realities exist; the researchers outline personal experiences and viewpoints that may have resulted in methodological bias*
- *Consistency (reliability); Establishes that, ultimately an independent researcher should be able arrive at similar or comparable findings.*
- *Applicability (generalisability) ; Gives consideration to whether findings can be applied to other contexts, settings or groups.*

Johnson (1997) describes many types of validity in qualitative research of which those two:

- *Descriptive Validity : a factual accuracy of the account as reported by the researcher*
- *Interpretive Validity : accuracy in portraying the meaning attached by participants to what is being studied.*

The *Face Validity* quoted byHolden, 2010 could thus be related to the *Truth Value* of Noble and Smith (2015) or to the *Descriptive Validity* of Johnson (1997).

Evaluating *Truth Value - Face Validity - Descriptive Validity* is recognizing that the interpretation bias, the particular way in which researcher view reality, corresponds to the reality in his/her colleague's world of reference. Many participants offered to translate and contribute to the development of the tool. This made a good argument in favor of the *Truth Value* of these findings. On the other hand, we have seen that the tool could be applied to very different situations in different countries in different languages. These two last points can bear witness to good *Truth Value* but also to good *Applicability*.

Evaluating *Consistency - Reliability - Interpretive Validity* is refering to whether these Q-Codes could be tested. It was imperative that the Q-Codes could be evaluated through extensive use of GP/FM grey literature indexing before being considered a valid construct. One measure used for testing was inter-indexer reliability. But, according to Funk and Reid (1983), who have studied the PubMed data-base for consistency in indexing, the quality of indexing cannot be directly measured, as there is no right or wrong way to index an article or abstract. In turn, the issue of holding abstracts to ambiguous standards of correctness is a potential downfall.

Testing inter-indexer reliability is ongoing. Belgian French speaking doctors in vocational training will identify concepts in their master's theses using 3CGP. At the department of family medicine, Coimbra university, Portugal, the master theses in GP/FM are currently being indexed by 3CGP. Brazilian GPs presenting a communication to the next congress of the Brazilian Society of Family and Community Medicine (SBMFC) have coded, using 3CGP as well, a large amount of abstracts. This material (384 oral communications and 1316 posters abstracts) all coded by 3CGP, is waiting to be analyzed. In Uruguay, the Q-Codes are in use to index online discussions about GP/FM sensitive matters in the Not@sLoc@s experimental web pages. The very fact that the Belgian, Portuguese, Brazilian and Uruguayan professionals in charge of GP/FM have attempted to use this new taxonomy and test its consistency implies that they recognize the interest of 3CGP, adding to its *Truth Value*. It also provides a measure within which the value/world-wide usability of the Q-Codes can be better tested in a multilingual world.

But the last word will go to Funk and Reid (1983) who emphasize that *it seems that, [in indexing] realistically, the upper level of consistency for any type of intellectual choice falls far below 100%*. Therefore, no matter how much effort that may be put in by all involved, there will always be a lack of universal consistency, which will naturally contribute negatively to this research.

1.5.8.5 Taxonomic Limitations

Looking for Concepts in Texts
Information retrieval is *finding material of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers)* (Braslavski et al., 2015). Extracting information from texts could be done through keywords extraction. Those keywords could be managed following some predetermined ordering principle. Ordering a set of concepts over a particular field is intrinsic to humans. The most used form of classification is hierarchical, in which a top concept has child which become parents of sub-concepts that are then organized into subcategories. There are innumerable forms of classification, but we have chosen the simplest one. Following the alphabetical model development for ICPC-2, we used the letter *Q* to fit within the identification logic of this classification. A second letter has been chosen to identify one of eight main domains that may cover the GP/FM universe of discourse . The choice of the domain is up to the coding researcher, who may apply his or her own bias. This applied bias must be taken into account when moving forward. Nevertheless, if numerous researchers agree on the division presented, the ordering principle may become a standard.

Proposing 8 Domains
We are not proposing a standard; however, we are proposing a searcher bias in need of discussion. The main aim of this research is facilitate the management of information produced by family doctors and to prepare it for further computerized development/reuse. The current version of this program is named Q-Codes, in honor of its creator Professor Lamberts, but it is still only a preliminary version (2.5). It will obviously evolve, and names will most likely change. But the need to manage GP/FM information in a structured and standardized way must remain a substantial facet of research. The future of the profession is at stake.

It is important to recognize that Q-Codes have been created from a limited number of abstracts (only 2300 communications). If a concept was not present in the read abstracts, it will have no place in the Q-codes. This emphasizes that the current program is limited to a small number of abstracts within the GP/FM field. Q-Codes would need to integrate much more information to be considered a fully applicable program to the field of GP/FM. Further conferences will contribute new concepts to this, while simultaneously helping GP/FM to evolve. We hope that the structure of this proposed classification will remain enough strong to support the introduction of new items, but it must be taken into account that as more information is added, the basis could potentially not be strong enough to accommodate all.

The Rag-Bag
The rag-bag (QO for *other*) no longer follows the definition of *other* proposed by the International Classification of Disease. The first ever list of disease

classifications recorded a hundred lines of diseases, and all that did not fit was categorized as *Other* (Bowker and Star, 1999). Here, the rag bag was to be used in some particular situations (see paragraph 1.3.5.1, page 63) of which the most important is QO4 (*to be considered for a new code*). Here, the QO4 code is our attempt to accommodate future research and evolutions, allowing for the integration of new concepts into the taxonomy. By creating the rag-bag, we hope to provide a preliminary solution that facilitates the continuous growth of Q-Codes.

Coding Techniques

One issue with the Q-codes lightweight ontology involves the unique identifiers. Cimino (1998) notes that when building an ontology, there is an *irresistible temptation to make the unique identifier a hierarchical code which reflects the concept's position in the hierarchy.* He also describes the advantages of using unique identifiers, saying that:

- *the hierarchical relations of the terms can be understood more easily by humans;* and
- *it is easier to query a database for members of a class.*

However, there are inherent disadvantages to using unique identifiers. The first issue, which we have encountered here, is that the coding system runs out of room to grow (Cimino, 1996; Cimino, 1998). This can be due to limited depth, limited breadth, or both of the unique identifier. For instance, when the code has a limited number of positions (digits), the depth of the hierarchy is limited. A possible method for dealing with limited breadth of the Q-Codes is assigning unique identifiers that allow tree addresses which can be of arbitrary length and breadth (Cimino, 1996). This can be done through using numbers containing decimals, similar to MeSH. But, when the positions in the code are limited to a fixed number of digits, the breadth of the hierarchy is limited. For instance, at each level, the Q-Codes can only support 9 positions, which poses a restriction. In addition, once a code is assigned, reclassifying the concept will break the hierarchical coding scheme Cimino (1998). There are two strategies that can be used to fix this issue:

- *i). Assign unique identifiers independent of the position of the concept in the hierarchy*, or
- *ii). Assign unique identifiers that allow tree addresses which can be of arbitrary length and breadth* (Cimino, 1996).

Some definitions of the Q-codes were born after intensive exchanges with several authors by mail. For example, the Medically Unexplained Symptoms (MUS) category. One could argue that MUS is a clinical theme and consequently finds its place in the clinical classification ICPC-2. However, this is not the reality. No patient claims to have MUS, and no doctor will label a problem as MUS, outside of a handful of researchers. The corresponding clinical issue would be labeled, instead, as *somatoform disease,non* *disease disease*, or something equally applicable. There is indeed an unlimited list of denominations for these ambiguous situations, which are often uninterpretable by the providers. Each Q-Code supposes the acting doctor uses a specific frame of reference for his/her own activity. In this sense, Q-Codes are an essential part of the analysis for Quality Assurance programs (Roland, Prevost, and Jamoulle, 2001). Further ontological research is needed to determine whether the two main rules of taxonomic thinking have been respected: completeness (all identified) and exclusivity (a place for each concept) (Ittoo and Bouma, 2013b).

The mnemotechnical and alphanumeric approach to the allocation of codes in the Q-Codes taxonomy (developed in analogy with ICPC) has well-known limitations. A more machine-oriented structure could be conceived in future versions of Q-Codes. The value of which, will be determined by the quality and speed of its maintenance and updating processes. This will be handled by a working group of the WONCA International Classification Committee (WICC) (`www.ph3c.org/Q`).

1.5.8.6 Scarcity of Ethical and Environmental Fields

Another issue imposed on this study was the lack of data for Ethical and Environmental fields. This inhibits us from fully editing and accommodating more work within these domains.

Unfortunately, there has been no opportunity for fruitful contact with the WONCA Ethics Special Interest group. The number of communications/abstracts over ethics has been very scarce in the conferences analyzed. This could explain the relative poverty of the QE domain. It is worth noting that Ethical Issues have often been addressed in collective sessions of some conferences, with no abstracts available. This could also explain the few number of abstracts dealing with ethical issues available for analysis.

The citations related to the environment were even less than those related to Ethics. This may be explained by the fact that a large part of the abstracts analyzed date from 2007, when the question of the planetary health was only discussed by a small groups of doctors. The abstracts analyzed in 2013 and 2014 are mainly devoted to teaching. A source selection bias can therefore explain the relative poverty of the QH domain. Unfortunately, the abstracts obtained from the Brazilian doctors attending the 2017 SBMFC congress are as poor as the WONCA 2007 and the CNGE 213 and 2014 ones in Environmental issues. This is raising the question of teaching and training the Environmental problems in the schools of medicine in a global way.

After the publication in the Lancet of the Manifesto *Safeguarding human health in the Anthropocene epoch* by the Lancet Commission on planetary health (Whitmee et al., 2015), fruitful contact through mail with the Environment group of WONCA (`https://tinyurl.com/WONCA-QH`) (Blashki et al., 2014) has facilitated the change from QH (Hazard) to QH (Planetary Health).

In the QD domain (Doctor's Issue), the prevention categories have been influenced heavily by the quaternary prevention groups the main author has contributed to. The two papers *The words of prevention I and II* published in the Brazilian Journal of Family and Community Medicine (Jamoulle, 2015a; Jamoulle et al., 2015c) show this influence and explain the presence of the subcategories under the QD44 *Quaternary Prevention* category.

1.5.8.7 Advantages and limits of the Semantic web

- Is There A Place in the Linked Data Universe for ICPC-2 and Q-Codes?

  The Q-Codes bases, like all the terminologies edited on HeTOP server, are fit for the Semantic web. As we have seen in the introduction (see paragraph 1.1.6.9, page 29), Semantic Web technologies promote common data formats and exchange protocols on the Web, like the Resource Description Framework (RDF), the Web Ontology Language (OWL-2) and the extraction language SPARQL. Q-Codes and ICPC-2 could find their place in the Open Data universe.

- Interface Between Life Sciences and Public Domain Vocabulary?

  Another issue posed, and potentially solved, against this study is the interface between Life Sciences and the Public Domain Vocabulary. Scientific articles often do not translate into the vernacular of the public, which posed a specific challenge to this work. The life science domain is extremely developed in the LOD, and it allows for explicit links to be made between datasets using shared semantics from standard ontologies and vocabularies. This, in turn, facilitates a greater degree of data integration (Jupp et al., 2014). The public domain is represented by general ontologies in the Linked Open Data cloud, especially by Dbpedia, YAGO or Babelnet. DBpedia (`http://dbpedia.org/ontology/`) extracts knowledge from 111 different language editions of Wikipedia but in a structured form, represented as an ontology and not as textual articles (Lehmann et al., 2015). YAGO (Yet Another Great Ontology) (`https://tinyurl.com/m9uo24v`) is a knowledge base developed at the Max Planck Institute for Computer Science in Saarbrücken in collaboration with Telecom ParisTech University (Demner-Fushman et al., 2013). BabelNet 3.7 (`http://babelnet.org/`) covers 271 languages and about 14 million concepts and named entities (Navigli and Ponzetto, 2012). Therefore, a potential solution was found. This is why, when possible, each Q-Code is linked to at least one DBpedia entry and one Babelnet entry. As Q-Codes are also linked to PubMed by the MeSH, the database offers a unique opportunity to connect life sciences, language, and vernacular language (a domain specific to GP/FM practices).

- Addressing the issue of quality in the Semantic web.

  As said by Doorn (2017) *Just opening up research data is not enough: the data should also be of sufficient quality*. The quality of data could be assessed using the FAIR Guiding Principles. Data has to **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable (Wilkinson et al., 2016).

Ermilov et al. (2016) discuss several issues dealing with semantic data. The first of which deals with evaluating the quality of information, the coverage a certain dataset provides, and the possible violation of privacy rules and links between datasets. Some of these problems have been already addressed by Benjamins et al. (2002). Following the contemplations of these authors, the main challenges to be addressed were (i) the availability of content, (ii) ontology availability (iii) scalability, (iv) multilinguality, (v) visualization to reduce information overload, and (vi) stability of Semantic Web languages.

Recently, Mihindukulasooriya, García-Castro, and Gómez-Pérez (2017) published a paper presenting LD Sniffer, a web-based open source tool for performing quality assessment over the accessibility of Linked Data. The perception of difficulties with quality and accessibility of semantic data must, therefore, push us to ensure that high quality data is published. We must also foresee that the content on our database will be continously regulated/edited. Those quality requirements are at the heart of the Linked Open Vocabulary, which this study will continue to strive to improve.

- The supposed simplicity of the language.

  Semantic difficulties may arise from the supposed simplicity of the language. The relation *is_a* is not a simple relation. Aristotelian logic, which decomposes the proposition into subject and predicate (Younes, 2016), is not sufficient in rendering reality. Following Wittgenstein, the relation *is_a* has at least three semantic interpretations. As stated by Wittgenstein (1922) in Tractatus Logico Philospophicus (TLP 3.323): *In the language of everyday life it very often happens that the same word signifies in two different ways – and therefore belongs to two different symbols – or that two words, which signify in different ways, are apparently applied in the same way in the proposition. Thus the word "is" appears as the copula, as the sign of equality, and as the expression of existence [......] In the proposition "Green is green" – where the first word is a proper name as the last an adjective – these words have not merely different meanings but they are different symbols.*

  Language could be more complex than its use in Health Information system. As stated by Elish and Boyd (2017); *Because computational systems require precise definitions and mathematically sound logics, sociocultural phenomena that are typically nuanced and fuzzy are rendered in coarse ways when implemented into code*. Again, the last word will be given to Wittgenstein (TLP 4.002): *Language disguises the thought. So that from the external form of the clothes one cannot infer the form of the thought.*

1.5.9 Suggestions for Further Research

There are many opportunities for further research for Q-Codes.

- We suggest that the indexing of grey literature in GP/FM remains ongoing and continues to be studied by both the original researchers and future researchers. Constant updating of the Q-Codes is needed for missed and emergent themes

- Reproductibility of the coding process have to be assessed and links to MeSH tested.

- The possibility of associating with 3CGP the recently developed PeRI code ((Schrans et al., 2016) relevant to the patient's context should be evaluated.

- Another suggestion for research would be to map Q-codes, already mapped to Medical Subject Headings (MeSH), to South American Health Descriptors (DeCS). Research in the domain of automatic annotation of NLP systems needs to be analyzed by future researchers, finding ways to make it easier to index and retrieve GP/FM-related abstracts in English and other languages.

- Also, as Q-codes cover managerial issues, it is believed that they could be adapted to other medical specialties. Thus, Q-Codes, as an extension to other classifications such as SNOMED-CT or ICD-11, should be researched so that the indexing of conference abstracts for other medical specialties may be integrated.

- Working on the immense reservoir of grey literature in GP/FM could be the next step. The idea is to give it visibility and utility and to contribute to the construction of networks of researchers in PHC and GP/FM. Resources could stay where they are, in conference settings, in universities, on the net, but will be managed following the principles of distributed data and semantic web technologies. An interface will be designed, allowing to find the data where they are and deferencing it at a glance. This will allow GP/FM to enter the world of Linked open data (see LOD `http://lod-cloud.net/`).

- Finally, the feasibility of creating a full ontology with 3CGP, using Protégé, could potentially be investigated by future researchers as well as the integration with upper level ontologies.

1.6 CONCLUSION

While exploring the core content of GP/FM is not a new idea, we have tried to represent GP/FM fields of knowledge through the development of a simple taxonomy, the Q-Codes. We have integrated the Q-Codes with the well known classification tool, ICPC-2. It is hoped that organizers of GP/FM conferences will ask authors or reviewers to classify the abstracts with the combined ICPC-2/Q-Codes system, also known as 3CGP. This, in turn, can support discussion of future fields of knowledge in GP/FM. Finally, and most importantly, the use of 3CGP could lead to improved access to information in GP/FM grey literature.

GP/FM is crossed by many currents of thought and is constantly confronted with the new applications of science, information technology and first line of health-care management. It is an evolutionary profession, as evidenced by the very many working groups of the World Association of Family Doctors (WONCA), and it needs a system which acclimates and accommodates these various adaptations. The concepts covered in 3CGP do not cover all the areas addressed by these groups. However, if time allows, we hope for continuation of this work. Perhaps in the future, with the updates of various researchers and doctors, the knowledge management system proposed will be perfected.

The Q-Codes are not a disembodied realization. The Q-Codes have been built as an identification tool for realistic GP/FM use. Here in Europe, particularly in Belgium and France, GP/FM is often presented as an unimportant sub-specialty, and, in turn, primary care is dismissed. This leads to a large gap in information knowledge, as GPs are usually the first to meet with various patients. So, the Q-Codes were built as a claim for specific, dedicated fields of knowledge within a realm of complexity. This kind of work can't be achieved by a professional terminologist or machine alone. Like Bowker and Star (1999) and Eco and McEwen (2001), we feel that categorization is also a political act, an appropriation or interpretation of a field of knowledge and duties. Family doctors, the main proxy for patients, should therefore take the lead in managing information in health care, as they are the ones that deal most often with various patient needs.

Chapter 2

Published Papers Supporting This Thesis

The idea of developing a taxonomy of knowledge discussed by family doctors had come to the author at the turn of the century. The organization of the papers and posters of GP/FM congresses was not effective, and too much time was spent from one poster to another, in order to identify those that corresponded to a particular field of interest. In 2006, French colleagues have given access to abstracts that would be presented before the 2007 WONCA Europe Conference. This is where it all began. But, the presentation of the analysis of 998 abstracts of the same congress made in Paris 2007 did not attract much attention. With this, the author has been touched by illness, and the project put aside. Five years later, in 2013, it was the enthusiasm of the young Portuguese doctors who revived the old flame, and the author decided to resume this study.

Attending the Portuguese Association of General and Family Medicine 2013 annual conference (APMGF – Associação Portuguesa de Medicina Geral e Familiar) in Covilha, Portugal, was an opportunity to witness the youth and enthusiasm of attendees, as well as the great quality of their work. The 128 submitted abstracts were peer-reviewed by the conference's organizers and printed in a leaflet for the attendees. Codification of those 128 abstracts was done the same day according to the ICPC-2 structure as well as the Q-Codes system, experimented six years earlier in Paris at the WONCA 2007 conference using ICPC-2 & Q-Codes version 0.2 (see Appendix A, page 191 and Appendix B, page 195). The results of the analysis were presented to the audience immediately, very well received by participants and published as a letter to the editor in the Portuguese Journal of General medicine in 2013. That letter is reproduced here (see chapter 3, page 137).

The 1980s saw the beginning of the computerized medical record. It had been necessary to learn how to move from the written form to the computerized one. However, computer tools were not used to their full capacity. Most transformations were, at best, automated writing and tailored accounting sheets. We went from the pencil to the keyboard and from the accounting sheets to the accounting software. Everything went faster, but inefficiently. The International Classification of Primary Care ICPC-2, born in 1987, and its grid structure were remarkably adapted to data processing software. However, this did not go beyond comparing digits at a very high speed. There was no question of tackling the meaning of the information produced. The computer could not understand what was calculated. Users were left with the discomfort of having to check the words corresponding to the concepts their brains dictated to them. We have assisted in the development of multiple, huge terminologies for health information systems, leading to multiple and huge problems in their interoperability. The chapter 4, page 147 and 5 chapters, page 147, deal with the delicate issue of terminological relationships and the limits of semantic interoperability.

At the beginning of the new century, the power of computers, the immensity of the Internet and the ideas of some genius, especially Sir Tim Berners-Lee, precipitated a new revolution. Berners-Lee, Hendler, and Lassila (2001) stated; *The Semantic Web will enable machines to COMPREHEND semantic documents and data, not only human speech and writings*. The semantic revolution was on its way. Using the computer to reason was becoming possible. To do this, it was necessary to put in place the appropriate tools. This is discussed in chapter 6, page 159, after an exploration of the fascinating universe of the semantic web.

Work done with ICPC were essential to the development of the tools we are preparing. We will not reproduce here the numerous papers and books written on this subject, which can be found in the list of publications.(See Published papers, page xxxi). Thanks to the collaboration of many general practitioners from all over the world, ICPC-2 has been edited in 21 languages on the HeTOP cross-lingual terminology portal (`www.hetop.eu`). Those languages are mostly European (e.g. Spanish, Portuguese) but also in Turkish, Japanese, Mandarin or Vietnamese. The HeTOP interface has been translated in more than 10 languages, including by WONCA colleagues (in Turkish, Vietnamese and Romanian, others ongoing). Accessing a health web site in his/her native language is of utmost importance for researchers. This translation work was the subject of a poster which is not reproduced in this thesis but which is available on-line (Schuers et al., 2015). The HeTOP portal was the subject of another poster, presented in Montevideo and which won the prize for the best poster (Jamoulle et al., 2015a) . The appendix F, page 219, gives readers access to the Uniform Resources Identifiers (URIs) of ICPC-2 and Q-Codes on the HeTOP server of the University of Rouen, France.

One of the points discussed in detail in the first chapter was the relationship between the Q-Codes and the Medical subject Headings. Four papers have been published on this subject. The issue of the relationship between MeSH and prevention has been the subject of two publications in the Brazilian Journal of Family and Community Medicine (Jamoulle, 2015a; Jamoulle et al., 2015c). Each year brings its share of new MeSH and some of them

find an alignment with the concepts of Q-Codes. The new MeSH published in 2016 have been carefully studied and inclusion of some MeSH 2016 into our database discussed and published in the International Journal of Health Policy and Management. (Jamoulle, 2016).
The document reproduced in chapter 7, page 169 analyzes the definitions in GP/FM versus PHC. This document makes it possible to distinguish General Practice/Family Medicine (GP/FM) from Primary Health care (PHC). One is the workforce and the other is the specific health service. These two universes are linked by the same approach to health, but they must be searched for by separate bibliographic filters, one filter for the GP/FM and another for PHC. These filters are available in appendix E, page 209. We see in the light of the classification proposed by Aristotle that the family physician occupies an ever-increasing place in the world of health.

This publication has been highlighted by the editor of the Royal Journal of British Practitioner Open. Its front page was edited, adding comments together with the image of the world cloud of GP/FM. The following quote extracted from the publication makes the intent of the paper clearer (`http://bjgpopen.org/` on September 15, 2017).

> ***What's in a word? Defining general practice, family medicine, & primary care.***
> *Would you regard the terms general practice, family medicine, and primary care as synonymous? Research from Jamoulle and colleagues demonstrates that while there are areas of overlap they differ greatly in content. Terminological analysis shows that continuity of care is common to both sets but there are differences in the terms used to depict workforce and structure between general practice/family medicine and primary health care. Clinicians and researchers should be mindful of these in literature searches. Also, the research highlights important discrepancies that should influence organisations like WONCA and the WHO as they seek to understand commonalities and develop future mutual collaborations*(On-line editorial, BJGP, September 15, 2017).

We have said in the introduction to this thesis that three disciplines are dealt with: computer science, the realm of family physicians and language sciences. In the paper 8, page 179, the article of the communication made at the congress of Manchester of the association Medical Informatics Europe 2017 is reproduced. The public was mainly composed of computer scientists and knowledge managers. The next is the paper submitted to the European Journal of General Practice, the official journal of WONCA Europe, whose readers are essentially General Practitioners (See chapter 9, page 187). The tools are presented in the appropriate terms for this audience. Finally, we reproduce the abstract of the communication accepted at the congress of Grey Literature in Rome in October 2017, whose attendants are mainly medical librarians and computational linguists (See chapter 10, page 189). The complete document will be available on-line in due time.

We must recall that the Q-Codes story and lists have been published in 6 languages (English - French - Dutch - Spanish - Portuguese - Turkish - Korean - Vietnamese) in book format of which the English version takes the name *General Practice / Family Medicine Multilingual Terminology - English version.* Other language files will be available on-line.

All the processes, including all the paper published for this thesis, are available on http://3cgp.docpatient.net together with a full 3CGP user guide.

Chapter 3

Using the International Classification for Primary Care (ICPC) and the Core Content Classification for General Practice (3CGP) to classify conference abstracts

Jamoulle M. Using the International Classification for Primary Care (ICPC) and the Core Content Classification for General Practice (3CGP) to classify conference abstracts. Letter. Port J Gen Pract. 2013;29(5).
http://orbi.ulg.ac.be/handle/2268/1716012

Rev Port Clin Geral 2013;29:334-8

USING THE INTERNATIONAL CLASSIFICATION FOR PRIMARY CARE (ICPC) AND THE CORE CONTENT CLASSIFICATION FOR GENERAL PRACTICE (3CGP) TO CLASSIFY CONFERENCE ABSTRACTS

UTILIZAÇÃO DA CLASSIFICAÇÃO INTERNACIONAL DE CUIDADOS PRIMÁRIOS (ICPC) E DA CLASSIFICAÇÃO DO CONTEÚDO NUCLEAR DE CLÍNICA GERAL (3CGP) PARA CLASSIFICAR RESUMOS DE CONGRESSOS

Family medicine is like the Danaïd's barrel, a bottomless pit of knowledge. Each year, thousands of general practitioners (GPs) hypothesize, develop research, gather data, elaborate reports and present their work. Much of this knowledge will be lost or remain hidden because only a few of these studies will be published in medical journals.[1] Books of abstracts are not readily accessible or, if published, are difficult to search. The WONCA Europe web site has gathered abstracts from 1995 to the present but there is no indexation system to retrieve specific subjects.

There have been several attempts to index the WONCA Europe and World conference abstracts with the International Classification of Primary Care (ICPC). This did not work. ICPC only address clinical situations and is unfit for non-clinical issues.

In the 1980s, Professor Henk Lamberts from Amsterdam University developed a classification called Q codes (as Q is not used in ICPC) to index, jointly with ICPC, the main publications of medical journals. In 2006, reusing the Q codes, I developed a classification of non-clinical issues addressed by GPs called 3CGP[2] allowing indexation of WONCA congress communications. 3CGP stands for Core Content Classification in General Practice/Family Medicine, and contains 164 rubrics divided in 8 domains, subdivided in categories and subcategories. To develop it, I have personally indexed the 1000 abstracts of the Paris congress in 2007 and presented my work during this conference.[3] This work has lain dormant for 6 years and opened one eye in June 2013 in Belém, Brazil, during the Sociedade Brasileira de Medicina de Família e Comunidade conference. The organizers of the 2016 WONCA World conference in Rio de Janeiro are looking for an abstract indexation system and have also expressed interest in 3CGP.

In the Portuguese mountains in Covilhã the idea woke up for good. I was invited to the Portuguese 18th national conference of family medicine and read with interest the "Livro de Resumos" [book of abstracts]. The 128 abstracts of very interesting work done by so many young and enthusiastic GPs have been indexed with 3CGP and ICPC. I present here the main results of this work. (Full data available on request.)

203 ICPC codes were used to classify 119 abstracts. Nine abstracts were not codable by ICPC (Figure 1). 36 codes were in component 1 (symptoms – complaints), 123 in component 7 (diagnosis) and 44 were process codes, of which 30 were related to drug prescription (-50) and only one about referral (-67). Chapters P, T, W and

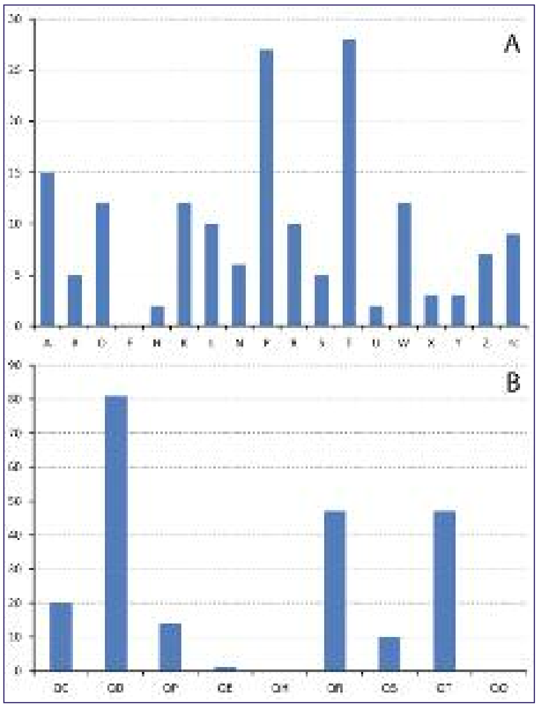


Figure 1. Content analysis of the 128 abstracts presented at the 18th national conference of family medicine with ICPC-2 (panel A) and 3CGP (panel B). nc - not codable with ICPC; QC - patient category; QD – doctors' issues; QP – patients' issues; QE - ethical issues; QH - hazard; QR - research; QS - structure of practice; QT - knowledge management; QO - others.

Z were overrepresented. There were 8 communications about depression (P76), 4 on dementia (P70) and five on tobacco issues (P17). The overwhelming predomi-nance of the T chapter is due to studies on the combi-nation of diabetes, obesity and lipid issues, always at-tractive for young doctors. Less expected were 8 com-munications about pregnancy and 7 addressing social issues.

With 3CGP one sees 36 communications describing disease (QD32 health issue management), 6 concerning children (QC11), 8 about aged people (QC14), 6 about relationships with secondary care (QS2). First place goes to Teaching (39 QT53 critical reading) and Research (25 QR2 epidemiology). 4 dealt with primary prevention (QD41), 12 with secondary (QD42), 4 with tertiary (QD43) and 8 with quaternary prevention (QD44). Only one paper addressed ethical issues.

These two tools show that GPs presenting in Covilhã preferred diseases (QD32 and component 7), screening for them (QD42), drugs (-50) and the so-called meta-bolic syndrome (diabetes, obesity and lipid disorders) but were also deeply interested in mental health, social problems and pregnancy. The influence of teachers and vocational training is evident with many communica-tions about epidemiological research (QR2) and criti-cal reading (QT53).

Much remains to be done before 3CGP becomes a professional tool allowing participants to search their preferred domain in a conference program, but yet one can have a look at abstracts from a different angle. Just like the paper presented in Covilhã, which shows the value of of "quebra-cabeças" [jigsaw puzzles] for preventing dementia,[4] 3CGP allows me to activate my neurons by following the interests of the young generation of dedicated doctors.

Marc Jamoulle*

*Family practitioner and researcher in Primary Care IRSS UCL, member of the Wonca International Classification Committee, member of MERITERM group (meriterm.org)

REFERENCES

1. Hummers-pradier E, Stöcker J, Fischer T. Which Abstracts Do Get Published? – Output Of German Gp Research 1999-2003 [Internet]. WONCA Europe Conference; 2007 Oct; Paris, France. Available from: http://www.woncaeurope.org/content/which-abstracts-do-get-published-%E2%80%93-output-german-gp-research-1999-2003
2. Jamoulle M. Core Content Classification of General Practice / Family Medicine (3CGP) ver 0.2 Oct 2007 [Internet]. Available from: http://docpatient.net/mj/wonca2007/3CGPFMdeskcopy.pdf
3 Jamoulle M, Hullers-Pradier E, Duhot D, Letrilliart L, Gay B. Towards an e-archive for Wonca documents. Workshop. Wonca Europe Conference; 2007 Oct. Paris, France. Available from: http://docpatient.net/mj/wonca2007/243.htm
4. Maltez R, Sá MJ, Gonçalves R. Terceira ronda: a idade em que os quebra-cabeças, quebram demência? Livro de resumos. C0353. 18º Congr. Nac. Med. Geral e Fam. 2013.

Chapter 4

Publishing a Multilingual Medical Terminology According to Terminology Standards and Linked Data Principles

Warnier, M., Roumier, J., Jamoulle, M., Cardillo, E., Vander Stichele, R., & Romary, L. (2012). Publishing a Multilingual Medical Terminology According to Terminology Standards and Linked Data Principles. SABRE Conference 2012.
http://hdl.handle.net/2268/171545

SABRE Conference 2012.

Title:

Publishing a Multilingual Medical Terminology According to Terminology Standards and Linked Data Principles.

Authors:

Maxime WARNIER[1], Joseph ROUMIER[2], Marc JAMOULLE[3], Elena CARDILLO[4], Robert VANDER STICHELE[5], LAURENT ROMARY[6]

[1]UCL-CENTAL, Belgium, student in Master degree, MaximeWarnier@gmail.com
[2]CETIC & Heymans Institute, Belgium,
Joseph.Roumier@cetic.be [3]IRSS-UCL, Belgium,
[3]Family Physician marc.@Jamoulle.com
[4]eHealth-FBK, Italy, Cardillo@fbk.eu
[5]Inria & HUB, France, Laurent.Romary@inria.fr
[6]Heymans Institute, University of Ghent, Belgium, Robert.VanderStichele@ugent.be

Abstract:

The article gives an overview of the results obtained by converting a focused medical resource into RDF triples and linking it with other reference resources.

Keywords: primary care, multilingual terminology, ontology, Linked Data.

The data source is a *multilingual medical database* manually created by Mr Marc Jamoulle, a Belgian general practitioner with a long experience in classification and terminologies for general practitioners. It consists of one hundred and seventy-three French terms identified in a guideline concerning heart failure [1], intended for family physicians and published by the Société Scientifique de Médecine Générale (SSMG, Belgium). This resource is a first step towards the creation of a Medical Reference Terminology [2]. Due to the readership of the publication, those terms are often distinct from the ones used in the common language, as well as from those – even more technical – used by the specialized cardiologists; this situation clearly proofs that, because of the diversity of the terms, interoperability is sometimes hard to preserve. All of them have also been translated into English and the corresponding concepts have been retrieved in four widely recognized and used international classifications (UMLS[1], SNOMED-CT[2], ICD-10[3] and ICPC-2[4]). This allowed to collect their lexical representations and the corresponding internal codes (if available), along with definitions and other useful information.

1 http://www.nlm.nih.gov/research/umls/
2 http://www.ihtsdo.org/snomed-ct/
3 http://www.who.int/classifications/icd/en/
4 http://www.who.int/classifications/icd/adaptations/icpc2/en/index.html
5 http://jena.apache.org/

This preliminary work was part of the Meriterm project.
The major part of the information present in the database was converted into nearly
13.000 RDF triples (serialized in an RDF/XML document) using the *Jena*[5] API for Java. This task was performed in a fully automated manner. The resulting resource can be found at the following URI: http://meriterm.org/heartfailure/heartfailure.rdf [6].

The resulting resource is a focused terminology that contains well-defined concepts, linked to reference resources of the field, with the associated term(s) in French and English (Dutch and Italian are next to be added). It was therefore mandatory to efficiently handle its multilingual nature. Moreover, as already mentioned, it was necessary to distinguish between the various words or phrases coexisting in a same language[7] and to specify the status of each of them (standardized, admitted or preferred).

This leads to a model based on the Terminological Markup Framework (TMF) [3], standardized by the ISO committee (ISO16642 [4]). This model will be described more in details in a further publication. The main components are called *Terminological Entry* (TE), *Language Section* (LS) and *Term Section* (TS). Thus, concepts (at a higher level) and terms are kept separated, but are strongly linked, since each TE can contain an unlimited amount of LS (currently two, for French and English) and, in the same way, each LS can have an undefined number of TS (one for each lexical representation of the concept). Obviously, in order to provide useful information, they must be described using several properties. Those properties link the components to data categories (definition, domain expert...), which, in turn, may be linked to other categories (e.g. a definition can be associated to its source). Similarly to Lemon, the data categories chosen come from ISOcat (ISO12620 [5]). The whole OWL vocabulary is declared at http://meriterm.org/heartfailure/vocabulary.owl.

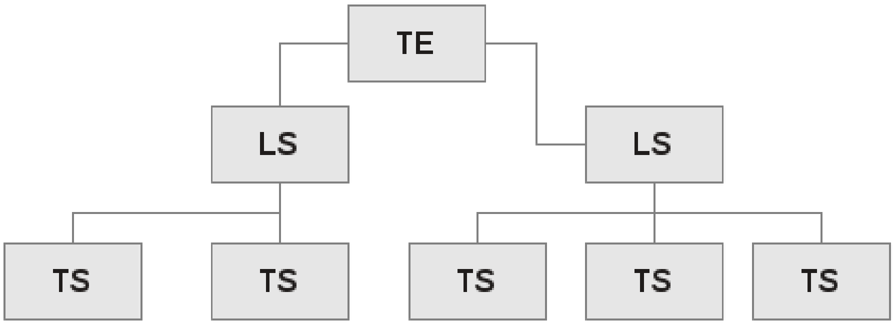
For convenience, direct access to every component is also provided by alternative, more readable (and elegant) URIs[8].

6 Using cURL: `curl -L -H "Accept: application/rdf+xml" http://meriterm.org/heartfailure/heartfailure.rdf`

7 For example, the diseases commonly referred to as *cancers* in common English are known as *malignant neoplasms* by the specialists.

8 Solution inspired by the explanations given by OpenLink Software (http://virtuoso.openlinksw.com/whitepapers/VirtDeploying LinkedDataGuide_Introduction.html#mozToc Id502192).

For instance, accessing http://meriterm.org/heartfailure/25/en/preferred will return an RDF/XML document containing only the few triples that are related to http://meriterm.org/heartfailure/heartfailure.rdf#25_TS_EN_PREF (that is, the *preferred term* of the English language section of the entry whose identifier is 25).[9] On top of it, this spares some bandwidth and processing resources, which is relevant, especially for mobile devices.



A very simplified view of the model

Likewise most available semantic linguistic resources, concepts (i.e. terminological entries) are represented as classes (with the standardized lexical representation as a label annotation), while the lexical representations (i.e. instances of Term Section) are represented as individuals.

Whenever possible, the former are declared *equivalent classes* to classes found in ontologies present on the NCBO BioPortal[10] and in the Data Hub, which have a high visibility (ICD10 is supervised by the World Health Organization and SNOMED- CT is maintained by the International Health Terminology Standards Development Organisation). Although this choice may not be exclusive, *equivalentClass* property were preferred over *sameAs* property, even if the *sameAs* property is more common, because an identity link may induce undesired affirmations and the task of the reasoners could become much harder [6]. However, a given concept does not always perfectly match the ones (that can be defined as *unionOf* concepts) found in the targeted ontologies or terminologies. This can be considered a problem as the formalism used might be too strong.

Finally, the dataset can be queried using SPARQL; for this purpose, a SPARQL endpoint is provided at http://meriterm.org:8081/openrdf-**sesame/repositories/meriterm**.

9 Or http://meriterm.org/heartfailure/25/fr/ for its French language section, or simply http://meriterm.org/heartfailure/25/ for the terminological entry itself.
10 http://bioportal.bioontology.org/

Here are some examples of queries

1. Find all the different lexical representations for a concept (e.g. http://meriterm.org/heartfailure/heartfailure.rdf#1) :

```
PREFIX voc:
<http://meriterm.org/heartfailure/vocabulary.owl#>
PREFIX hf:
<http://meriterm.org/heartfailure/heartfailure.rdf#>

SELECT

DISTINCT
```

2. Find all the concepts that are present in UMLS, but not in ICPC2:

```
PREFIX voc:
<http://meriterm.org/heartfailure/vocabulary.owl#
>  PREFIX hf:
<http://meriterm.org/heartfailure/heartfailure.rd
f#>

SELECT

?concept

?code_umls

WHERE {
        ?concept voc:hasClassificationCode ?classification_code_umls .
        ?classification_code_umls voc:hasOriginatingDatabaseName hf:UMLS .
        ?classification_code_umls voc:hasValue ?code_umls

        OPTIONAL {
                ?concept voc:hasClassificationCode ?classification_code_icpc .
                ?classification_code_icpc voc:hasOriginatingDatabaseName hf:ICPC2
        }

        FILTER(! bound(?classification_code_icpc))
}
```

Aknowledgement:

Part of the work was done within the Meriterm project, that started in 2011 between the Centre d'Expertise en Technologies de l'Information et de la Communication, the Heymans Institute of Pharmacology, the University of Ghent, the Université de Louvain and the Fondazione Bruno Kessler, and aims at creating a framework for publishing and editing medical multilingual linguistic resources using semantic web concepts and tools as well as existing metadata, terminology and lexicon standards.

References:

[1] P. Van Royen, P. Chevalier, G. Dekeulenaer, M. Goossens, P. Koeck, M. Vanhalewyn, P. Van den Heuvel. Recommandation de Bonne Pratique. Insuffisance cardiaque, 2011. http://www.ssmg.be/images/ssmg/files/Recommandations_de_bonne_pratique/RBP_insuffisance_cardiaque.pdf

[2] J. Roumier, R. Vander Stichele, L. Romary, and E. Cardillo, "Approach to the Creation of a Multilingual, Medical Interface Terminology," Nov. 2011. http://hal.inria.fr/hal-00646223_v1/

[3] L. Romary, "An abstract model for the representation of multilingual terminological data: TMF - Terminological Markup Framework", Proc. TAMA 2001 – http://hal.inria.fr/inria-00100405

[4] ISO 16642:2003, Computer applications in terminology – Terminological markup framework (TMF), 2003.

[5] ISO 12620:1999, Computer applications in terminology – Data categories, 1999.

[6] H. Halpin, I. Herman, P. J. Hayes. When owl:sameAs isn't the Same: An Analysis of Identity Links on the Semantic Web, 2010.

Chapter 5

Terminological analysis of a Belgian Heart Failure Guideline in French and mapping to international terminologies and classifications

Jamoulle M, Vander Stichele RH, Cardillo E, Roumier J, Warnier M. Mapping French terms in a Belgian guideline on heart failure to international classifications and nomenclatures: the devil is in the detail. Informatics in Primary Care. 2014;4(189–198):189–98.
http://hdl.handle.net/2268/171599

Research article

Mapping French terms in a Belgian guideline on heart failure to international classifications and nomenclatures: the devil is in the detail

Marc Jamoulle
Department of General Practice, Liège University, Liège, Belgium

Elena Cardillo
Fondazione Bruno Kessler, Trento, Italy

Joseph Roumier
Centre d'Excellence en Technologies de l'Information et de la Communication, Charleroi, Belgium

Maxime Warnier
University of Toulouse, Toulouse, France

Robert Vander Stichele
University of Ghent, Ghent, Belgium

Jamoulle M, Cardillo E, Roumier J, Warnier M, Stichele RV. Mapping French terms in a Belgian guideline on heart failure to international classifications and nomenclatures: the devil is in the detail. *Inform Prim Care*. 2014;21(4):189–198.

http://dx.doi.org/10.14236/jhi.v21i4.66

Author address for correspondence:
Marc Jamoulle
Department of General Practice
Liège University
Place du 20 Août
4000 Liège, Belgium
Email: marc.jamoulle@doct.ulg.be

Accepted September 2014

ABSTRACT

Introduction With growing sophistication of eHealth platforms, medical information is increasingly shared across patients, health care providers, institutions and across borders. This implies more stringent demands on the quality of data entry at the point-of-care. Non-native English-speaking general practitioners (GPs) experience difficulties in interacting with international classification systems and nomenclatures to facilitate the secondary use of their data and to ensure semantic interoperability.

Aim To identify words and phrases pertaining to the heart failure domain and to explore the difficulties in mapping to corresponding concepts in ICPC-2, ICD-10, SNOMED-CT and UMLS.

Methods The medical concepts in a Belgian guideline for GPs in its French version were extracted manually and coded first in ICPC-2, then ICD-10 by a physician, an expert in classification systems. In addition, mappings were sought with SNOMED-CT and UMLS concepts, using the UMLS SNOMED-CT browser.

Results We identified 143 words and phrases, of which 128 referred to a single concept (1-to-1 mapping), while 15 referred to two or more concepts (1-to-*n* mapping to ICPC rubrics or to the other nomenclatures). In the guideline, words or phrases were often too general for specific mapping to a code or term. Marked discrepancy between semantic tags and types was found.

Conclusion This article shows the variability of the various international classifications and nomenclatures, the need for structured guidelines with more attention to precise wording and the need for classification expertise embedded in sophisticated terminological resources. End users need support to perform their clinical work in their own language, while still assuring standardised and semantic interoperable medical registration. Collaboration between computational linguists, knowledge engineers, health informaticians and domain experts is needed.

INTRODUCTION

Speakers of different languages have different ways in representing the world; such differences lie not only in terms, but also in the concepts themselves.[1] The medical language is no exception. The translation of English medical terminological resources to other languages could be difficult.

The general practitioner (GP) is an important node of the information flow in health care. He is at the crossing point between patient and health care worlds and is also a focal point in the life cycle of clinical information. The family doctor has to manage lay terms as well as professional terms in the native language of the patient.[2] Moreover, as patients are considering the Internet and social networking as information sources[3] for health care, mappings to multilingual lay terms are inescapable.[4]

For GPs, guidelines offer coherent, comprehensive and consensual recommendations which at least, if designed rigorously, minimise the potential harms to patients.[5]

In this article, we report the terminological analysis of a guideline for GPs in Belgium (a West European bilingual country), produced in 2012, according to explicit guideline development methods.[6] The topic of the guideline was 'heart failure', a prevalent chronic disease. The guideline was published in French and Dutch, with identical structure, content and text size. The analysis of the words and phrases used in this guideline provided the material for the MERITERM consortium (meriterm.org), a multidisciplinary research group who has proposed a hybrid methodology for the creation of a multilingual reference terminology for the primary care domain in Europe, by combining the application of the International Standards Organisation's (ISO's) standards for multilingual terminologies and the use of Semantic Web technologies for connecting this resource to linguistic resources as well as to medical nomenclatures, classifications and thesauri.[7,8] The long-term objective is to build an end-user terminology that supports not only coding activities for medical registration, but also for information retrieval, quality assurance and creation of information through epidemiological research[9] considering multilingualism and semantic interoperability.[10]

Ontologies and semantic integration technologies are seen as tools to manage knowledge in health care, but require further research and development as well as experimental work.[11] Ontology development should be preceded by thorough descriptive terminological analysis of the real-life language of the persons who will ultimately use the ontology.[12] By analysing the terminological content of a clinical guideline and making an attempt to link the extracted words and phrases to a broad range of relevant nomenclatures and classifications, we provide a glimpse of the complexity of language, machine terminology, medical semantics and ontologies.

The aim of this article was to explore the semantic difficulties in mapping French terms from a general practice guideline on heart failure to relevant concepts in English-based international nomenclatures and classifications.

Our research question was: what are the barriers to the process of mapping clinical terms in clinical guidelines in a language other than English (i.e. French) to concepts available in English language classifications and nomenclatures?

Method

The identification of the clinical terms relevant for patient care within the text of the heart failure guideline was carried out manually by a French-speaking family doctor with expertise in medical classification and terminology.

The selected words were then classified according to an ad hoc grid (disease, function, objective finding, process general, risk, symptom and therapeutics process). Words and phrases related to 'objective findings' and 'therapeutic process' were excluded to remove technical names of medical procedures and medications. Acronyms and names of clinical scales and their domain values were also excluded, as these were taken from standardised terminologies from the beginning. Included French words and phrases were entered into a database (a multilingual reference terminology) and considered as preferred term of a concept, with an informal definition, to indicate the intended sense, in case of polysemy (words having more than one sense). This was necessary to assure correct mapping to concepts in international classifications, but also to further identify the correct preferred terms in other languages[13] and in the Dutch version of the guideline.

Here, report mapping of the selected concepts pertaining to French words and phrases, on the one hand, and concepts in international nomenclatures and classifications on the other hand. After three revisions, all the results were arranged in a database showing the mapping between the terms in French and the concepts in the four international classifications and nomenclatures (e.g. Box 1).

The first step was to semantically map the selected terms to ICPC-2,[14] a classification specific to general practice and primary care. In a second step, the identified ICPC-2 codes were mapped to ICD-10,[15] the main classification for encoding diseases in secondary care. For each mapped concept in ICPC-2 and ICD-10, we retrieved the term and its code. Furthermore, mappings to SNOMED-CT,[16] the nomenclature and reference terminology for coding in electronic health records, and to UMLS,[17] the main terminology used for indexing knowledge in medicine, were sought. For SNOMED-CT, we retrieved the corresponding concept, the code and the fully specified name, including the 'semantic tag'.[18] For UMLS, we retrieved the corresponding term, the code, the 'semantic type, and the definition (if available).[19] For SNOMED-CT, it was not possible to retrieve a definition, as a formal definition of concepts is not provided in SNOMED-CT.

We used specific online browsers for each of the four classifications and nomenclatures described above to extract the mappings (see Box 2).

Codes and semantic values of the different terms and their meanings were then subjected to a comprehensive critical study.[20,21]

Box 1 Column labels and row content of the database with as example: the term cardiac insufficiency and its correspondences in ICPC-2, ICD-10, SNOMED-CT and UMLS

Column label (short)	Column label (long)	Row content
Term (French)	Term used in French for selected concept in French guideline	Décompensation cardiaque
Term (English)	English translation of the term	Cardiac insufficiency
Type	Type of the concept (diagnosis, symptoms and risk)	Diagnosis
ICPC code	Code of the corresponding rubric or chapter in ICPC-2	K77
ICPC-2 rubric	Label of the corresponding rubric in ICPC-2	Heart failure
ICD-10 code	Code of the corresponding concept in ICD-10	i50
ICD-10 label	Label of the corresponding category in ICD-10	Heart failure
SNOMED-CT ID	Corresponding identifier of the corresponding concept in SNOMED-CT	825890014
SNOMED-CT FSN	Fully specified name in SNOMED-CT (with its semantic tag)	Heart failure (disorder)
UMLS CUI	Concept unique identifier (CUI) of the corresponding concept in UMLS	C0018801
UMLS concept name	Concept name in UMLS (with its semantic type)	Heart failure
UMLS Def_Source	Source of the definition in UMLS if any	CSP/PT
UMLS definition	Definition of the concept if any in UMLS	Inability of the heart to pump blood at an adequate rate to fill tissue metabolic requirements or the ability to do so only at an elevated filling pressure.

Box 2 Four classifications and terminologies browsers available online

- **ICPC-2e (en) v4.2beta browser** accessible through the Web page of the World Health Organisation (WHO) collaborating centre for the Family of International Classifications in the Netherlands. **http://icpc.who-fic.nl/browser.aspx** (web site of the WICC ICPC update group), which allows also to find the correct transcoding code to ICD-10.
- **ICD online browser** WHO. ICD-10 Version: 2010. International Statistical Classification of Diseases and Related Health Problems 10th Revision. accessible through: **http://apps.who.int/classifications/icd10/browse/2010/en**
- **SNOMED CT browser** of the UMLS terminology services, a service of the U.S. National Library of Medicine | National Institutes of Health **https://uts.nlm.nih.gov///snomedctBrowser.html** allowing to find the corresponding meaning in SNOMED-CT 2011.
- **Metathesaurus browser** of the UMLS terminology services, a service of the U.S. National Library of Medicine | National Institutes of Health **https://uts.nlm.nih.gov///metathesaurus.html** allowing to find the correspondence between the SNOMED-CT concept and the CUI of UMLS (linked to the SNOMED-CT browser).

Results

In total, 283 words and phrases were selected from the French version of the Belgian guideline. Eight entries were excluded because they were acronyms (e.g. 'AIT Attaque Ischémique Transitoire' [TIA for transient ischaemic attack]) or names and domain values of clinical scales (e.g. New York Heart Classification with four domain values). We excluded 132 words and phrases pertaining to objective findings and therapeutic process. We retained 143 words and phrases.

Out of these 143 words and phrases, 128 referred to a single concept (1-to-1 mapping); an additional 15 words and phrases were so general that eight of them needed to be represented by two more specific concepts and six by three concepts, to be able to make a bridge to ICPC or other nomenclatures in a 1–*n* mapping. The final reference terminology database had 160 lines. (The file is available on http://meriterm.org/K77.xlsx.)

Another eight of the 128 selected terms were so general and unspecific that it was not possible to find any correspondence in the ICPC-2 classification, while for an additional eight terms only a partial correspondence with ICPC-2 was found at the level of the chapter. For these very generic terms, equivalents could be found in the SNOMED-CT and UMLS, but with discrepancies in the choice of semantic tags or types (Table 1).

In addition, five of the 15 terms which could not be represented with a single concept pertained to general but related cardiovascular concepts (slightly different in the label, but without consistent difference in meaning): cardiomyopathy, coronaropathy, coronary ischaemia, myocardial ischaemia and coronary heart disease (Table 2).

Six general concepts needed to be split over two distinct or more concepts in ICPC, based on insulin dependent/non-insulin dependent; acute/chronic or male/female distinctions (Table 3).

Table 1 **Terms so general that they could not or only partially be mapped to the ICPC-2 or ICD 10 classifications – the dash indicates inability to map to concepts in the system**

Term (French)	Term (English)	ICPC-2 code	ICD-10 code	ICD-10 label	SNOMED-CT FSN (semantic tag)	UMLS concept name (semantic type)
Affection thoracique	Thoracic disease	–	–	–	Disorder of thorax (disorder)	Thoracic Diseases (disease or syndrome)
Besoin spirituel	Spiritual need	–	–	–	Spiritual need of patient (observable entity)	Spiritual need (finding)
Cancer	Cancer	–	C00–C97	Malignant neoplasms	Malignant neoplastic disease (disorder)	Malignant Neoplasms (neoplasic process)
Fonctions cognitives	Cognitive functions	–	–	–	Cognitive functions (observable entity)	Cognitive functions (mental process)
Filtration glomerulaire	Glomerular filtration	–	–	–	Glomerular filtration, function (observable entity)	Glomerular filtration (organ or tissue function)
Résistance vasculaire systémique	Systemic vascular resistance	–	–	–	Systemic vascular resistance (observable entity)	Total peripheral resistance (physiologic function)
Risque	Risk	–	–	–	Risk of (contextual qualifier) (qualifier value)	Risk (qualitative concept)
Tolérance à l'effort	Exercise tolerance	–	–	–	Exercise tolerance (observable entity)	Exercise Tolerance (clinical attribute)
Anémie	Anaemia	B	D64.9	Anaemia, unspecified	Anaemia (disorder)	Anaemia (disease or syndrome)
Maladie du foie	Liver disease	D	K70–K77	Diseases of liver	Disorder of liver (disorder)	Liver diseases (disease or syndrome)
Maladie de cœur	Heart disease	K	I51.9	Heart disease, unspecified	Heart disease (disorder)	Heart disease (disease or syndrome)
Maladies vasculaires périphériques	Peripheral vascular diseases	K	I70–I79	Diseases of arteries, arterioles and capillaries	Peripheral vascular disease (disorder)	Peripheral vascular diseases (disease or syndrome)
Fracture	Fracture	L	–	–	Fracture of bone (disorder)	Fracture (disease or syndrome)
Maladie du poumon	Lung disease	R	–	–	Disorder of lung (disorder)	Lung diseases (disease or syndrome)
Maladie de la thyroïde	Thyroid disease	T	E00–E07	Disorders of thyroid gland	Disorder of thyroid gland (disorder)	Thyroid diseases (disease or syndrome)
Maladie du rein	Kidney disease	U	–	–	Kidney disease (disorder)	Kidney diseases (disease or syndrome)

Table 2 Some very generic terms found in the French guidelines and their mappings to ICPC-2 with due correspondences

Term (French)	Term (English)	ICPC-2 code	ICPC-2 rubric	ICD-10 code	SNOMED-CT FSN (semantic tag)	UMLS concept name (semantic type)
Cardiomyopathie	Cardiomyopathy	K84	Heart disease other	I42	Cardiomyopathy (disorder)	Cardiomyopathies (disease or syndrome)
		K73	Congenital anomaly cardiovascular	I42.4	Congenital hypertrophy of cardiac ventricle (disorder)	Congenital hypertrophy of cardiac ventricle (congenital abnormality / disease or syndrome)
		K76	Ischaemic heart disease with or without angina	I25.5	Ischaemic congestive cardiomyopathy (disorder)	Ischaemic congestive cardiomyopathy (disease or syndrome)
Coronaropathie	Coronaropathy	K74	Ischaemic heart disease without angina	I25	Coronary arteritis (disorder)	Coronary arteritis (disease or syndrome)
		K75	Acute myocardial infarction	I25	Acute myocardial infarction (disorder)	Acute myocardial infarction (disease or syndrome)
		K76	Ischaemic heart disease with or without angina	I25	Chronic ischaemic heart disease (disorder)	Chronic myocardial ischaemia (disease or syndrome)
Ischémie coronaire	Coronary ischaemia	K74	Ischaemic heart disease without angina	I25	Coronary arteritis (disorder)	Coronary arteritis (disease or syndrome)
		K76	Acute myocardial infarction	I25	Chronic ischaemic heart disease (disorder)	Chronic myocardial ischaemia (disease or syndrome)
Ischémie myocardique	Myocardic ischaemia	K74	Ischaemic heart disease without angina	I25.9	Myocardial ischaemia (disorder)	Myocardial Ischaemia (disease or syndrome)
		K75	Acute myocardial infarction	I25	Acute myocardial infarction (disorder)	Acute myocardial infarction (disease or syndrome)
Pathologie de l'artère coronaire	Coronary artery pathology	K74	Ischaemic heart disease without angina	I25	Coronary arteritis (disorder)	Coronary arteritis (disease or syndrome)
		K75	Acute myocardial infarction	I25	Acute myocardial infarction (disorder)	Acute myocardial infarction (disease or syndrome)
		K76	Ischaemic heart disease with or without angina	I25	Chronic ischaemic heart disease (disorder)	Chronic myocardila ischaemia (disease or syndrome)

Table 3 Six terms of the French guideline causing problems of specification – the dash indicates inability to bridge to concepts in the system

Term (French)	Term (English)	ICPC-2 Code	ICPC-2 Rubric	ICD-10 Code	ICD 10 Label	SNOMED-CT FSN (semantic tag)	UMLS Concept Name (semantic type)
Diabète	Diabetes	T89	Diabetes insulin dependent	E10	Insulin-dependent diabetes mellitus	Diabetes mellitus (disorder)	Diabetes mellitus (disease or syndrome)
		T90	Diabetes non-insulin dependent	E11	Non-insulin-dependent diabetes mellitus	Diabetes mellitus (disorder)	Diabetes mellitus (disease or syndrome)
Consommation d'alcool excessive	Excessive alcohol consumption	P15	Chronic alcohol abuse	F10.1	Harmful use of alcohol	Alcohol abuse (disorder)	Alcohol abuse (mental or behavioural dysfunction)
		P16	Acute alcohol abuse	F10.0	Acute alcohol intoxication	Alcohol intoxication (disorder)	Acute alcoholic intoxication (mental or behavioural dysfunction)
Maladies sexuellement transmissibles	Sexually transmitted diseases	Y	Male Genital	A50–A64	Infections with a predominantly sexual mode of transmission	Sexually transmitted infectious disease (disorder)	Sexually transmitted diseases (disease or syndrome)
		X	Female Genital	A50–A64	Infections with a predominantly sexual mode of transmission	Sexually transmitted infectious disease (disorder)	Sexually transmitted diseases (disease or syndrome)
Problèmes sexuels	Sexual problems	P08	Sexual fulfilment reduced	F52.1	Sexual aversion and lack of sexual enjoyment	_	Sexual fulfilment reduced (sign or symptom)
		P	Psychological	_	_	History of – sexual problem – female (situation)	H/O: sexual problem – female (finding)
		P	Psychological	_	_	History of – male sex function problem (situation)	H/O: male sex function problem (finding)
		P	Psychological	_	_	Abnormal sexual function (finding)	Sexual dysfunction (finding)
		Y07	Impotence not otherwise specified	N48.4	Impotence of organic origin	Impotence (disorder)	Erectile dysfunction (disease or syndrome)
Syphilis	Syphilis	Y70	Syphilis male	A52.0	Cardiovascular syphilis	Syphilis (disorder)	Syphilis (disease or syndrome)
		X70	Syphilis female	A52.0	Cardiovascular syphilis	Syphilis (disorder)	Syphilis (disease or syndrome)
Trouble du rythme cardiaque	Heart rhythm disorders	K84	Heart disease other	I45.9	Conduction disorder and unspecified	Conduction disorder of the heart (disorder)	Conduction disorder of the heart (disease or syndrome) – Cardiac arrhythmia (pathologic function)

Table 4 **Number of terms of the French guidelines not mapped to the classifications or nomenclatures**

Not in UMLS (1)
Drug induced ankle oedema

Not in SNOMED (2)
Drug induced ankle oedema
Sexual fulfilment reduced

Not in ICPC (8)
Spiritual need
Cancer
Functional capacity
Cognitive functions
Glomerular filtration
Systemic vascular resistance
Risk
Exercise tolerance

Not in ICD (24)
Thoracic disease
Spiritual need
Functional capacity
Cognitive functions
Coronary angiography
Cardiac echography
Glomerular filtration
Fracture
Weight measure
Drug induced ankle oedema
Heart valve disease
Lung disease
Referral to cardiologist
Intercourse
Systemic vascular resistance
Magnetic resonance
Risk
Chest X-ray
Cognitive status
Nutritional status
Fluid balance
Exercise tolerance
Isotopic ventriculography

As shown in Tables 1–4, eight terms could not be mapped to ICPC-2, 24 to ICD-10, 1 to UMLS and 2 to SNOMED-CT, and only one term, 'drug induced ankle oedema', did not map to three classifications/nomenclatures (UMLS, SNOMED-CT and ICD-10).

While looking for semantic mappings, we noticed that heterogeneity is the rule. Even if SNOMED-CT disorders (semantic tags) are often related to UMLS disease or syndrome (semantic type) and UMLS finding and sign, and symptoms recover more or less what is called symptoms in ICPC, no such semantic tag as symptom exists in SNOMED-CT.

We did observe several discrepancies in the semantic values of the two classifications and the two nomenclatures. T06 (loss of appetite) is a symptom in ICPC-2 as well as in ICD-10 (classified in Chapter R), a finding in SNOMED-CT and a disease or syndrome in UMLS (as indicated by the specific browser). The absolute distribution of the retrieved semantic meanings seems to correspond roughly at first glance: we found 80 UMLS diseases or syndrome, 96 SNOMED-CT disorders, 89 ICD10 disease and 76 ICPC-2 diagnoses (Table 5).

However, 41 entries pertaining to ICPC-2 symptoms have corresponding entries in SNOMED-CT (except one) as: 27 findings, 11 disorders and two observable entities. The same ICPC-2 entries are distributed in eight different semantic types in UMLS: 22 sign and symptoms, seven findings and five others (Figure 1).

Hence, observed barriers to mapping French words and phrases were problems of polysemy, homonymy and the use of overly general concepts (hypernimy) which could either not be represented or needed to be related to several more specific concepts (hyponymy) in the different classifications and nomenclatures. In addition, we observed differences in semantic tags and types between the different systems, illustrating differences in world of reference.

This expert-driven review of the meaning of words and phrases from a single guideline in a specific field of general practice, and the mapping of these French words and phrases to concepts in international classifications and nomenclatures proved to be a labour-intensive process.

DISCUSSION

By examining the terminological content of a French guideline for GPs about heart failure, we were able to identify 128 words and phrases that could be related to a single concept, with a 1-to-1 mapping to international classifications, and 15 that needed to be related to two or more concepts, to be able to map (1-to-*n* mapping). In total, 160 concepts were mapped to (1) two international medical classifications (ICPC-2 and ICD-10) and (2) two international nomenclatures SNOMED-CT (for medical registration) and UMLS (for indexing and information retrieval).

A barrier observed was the very low specificity of words and phrases used in the narrative version of the guideline. A first step was to choose the intended implicit sense of each observed word and phrase for mapping to the international classifications. Often several quasi-synonyms were used for very general concepts that needed to be represented in the classifications or nomenclatures by more specific concepts.

In this article, we attempted the mapping to classifications for all observed terms, but it is clear that authors of guidelines will have to make choices in the fuzzy vocabularies and carefully select well-defined concepts at the right level of abstraction and also select the most accurate representation of the concept by a preferred term in their own language. In further terminological work, additional meanings of selected terms in the given language and synonyms for the concept should be dealt with in linguistic resources capable of dealing with the richness and fuzziness of natural language(s).

The attempt to map to several international classifications and nomenclatures was supported using the ICPC classification from general practice as a grid for access to the more granular ICD classification, and ultimately to the two much more specific nomenclatures. The use of the SNOMED-CT to UMLS mapping was useful to retrieve international definitions of concepts (not available in SNOMED-CT). For a number of concepts, this approach was not working, and ICPC (and even ICD) needed to be bypassed.

A limitation of this article was that the extraction of words and phrases, the choice of intended sense and the mapping

Table 5 Distribution of the Semantic values in the terms of the French guideline

UMLS concepts (semantic types)	163	SNOMED-CT (semantic tags)	155	ICD-10 equivalent category	114	ICPC-2 equivalent category	129
Disease or syndrome	80	Disorder	96	Disease	89	Diagnosis	76
Finding	19	Finding	28	–	–	–	–
Sign or symptom	21	–	–	Symptoms, signs & findings (Chapter R)	24	Symptom/complaint	45
Pathologic function	10	Observable entity	17	–	–	–	–
Diagnostic procedure	5	Procedure	6	Factors influencing health status (Chapter Z)	1	Process	8
Mental or behavioural dysfunction	5	Situation	2	–	–	–	–
Mental process	3	Qualifier value	2	–	–	–	–
Organism attribute	2	Event	2	–	–	–	–
Organism function	2	Substance	1	–	–	–	–
Clinical attribute	2	Physical object	1	–	–	–	–
Congenital abnormality	1	–	–	–	–	–	–
Neoplasic process	2	–	–	–	–	–	–
Physiologic function	2	–	–	–	–	–	–
Anatomical abnormality	1	–	–	–	–	–	–
Functional concept	1	–	–	–	–	–	–
Hazardous or poisonous substance	1	–	–	–	–	–	–
Health care activity	1	–	–	–	–	–	–
Individual behaviour	1	–	–	–	–	–	–
Injury or poisoning	2	–	–	–	–	–	–
Medical device	1	–	–	–	–	–	–
Organ or tissue function	1	–	–	–	–	–	–

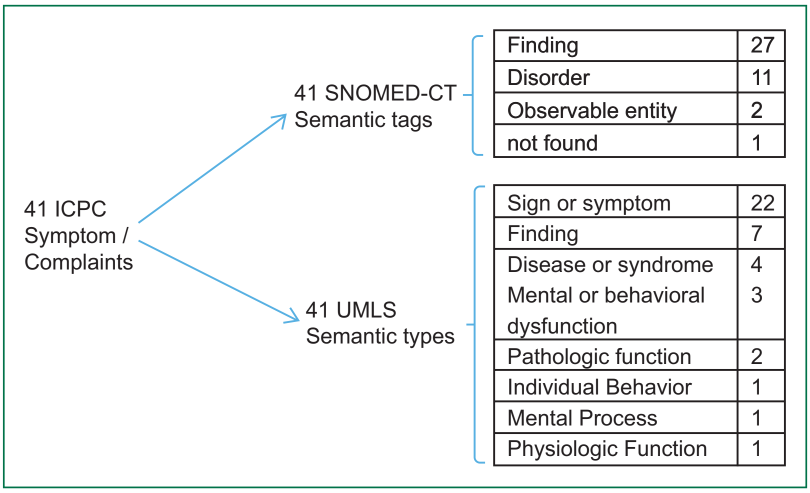


Figure 1 Semantic values mapping distribution in 41 items classed as Symptoms in ICPC-2 on 160 analysed

to international concepts was only performed by a single researcher, albeit physician and expert in terminology. The process included an element of translation, which is always a shift, not only between two languages but between two cultures.[22] The chosen terms and proposed mappings reflect the bias, background, errors, omissions and misinterpretations of the coder. Validation by a consensus process, including medical translators, is needed. Consultations between representatives of multiple languages will be necessary to decide on the ultimate selection of a collection of reference concepts, suitable for a feasible mapping to international classifications and nomenclatures.

We observed important differences between the four international terminology systems. Classifications and nomenclatures are built for different purposes, and not based on the same world of reference. UMLS Semantic types[23,24] or SNOMED-CT Semantic tags[25] are different constructs. SNOMED-CT represents the views of the pathologists in a world where symptoms are non-existing and where there is only a place for disorder, finding, observable entities, morphology or structure. SNOMED-CT shares also with UMLS authors a pre-Canguilheim[26] view about the normal and the pathological by referring to the concept of abnormality. ICPC-2 reflects the point of view of family doctors who are close to their patients.[27] ICD-10 is disease-centred and an extension of a historical mortality classification, although augmented by several symptoms and procedures. Classifications and terminologies are not neutral; they reflect the world view of the human classifier. Bowker[28] summed the subjective nature as 'to classify is human'. Nevertheless, despite the difficulty of reconciling these apparently so different worlds, we must focus on our common interest, the safe and effective care of the human person.[29]

Guidelines in narrative form are difficult to implement in EHRs. In the last 20 years, major steps have been made in methods to formalise guidelines into computer usable information, often in the field of heart diseases.[30–31] This formalisation requires the use of standardised clinical vocabularies and terminologies.[32] Efforts are needed from physicians to understand the necessity to use a common standardised terminology when producing guidelines and performing medical registration. However, physicians in clinical practice also need to be helped by sophisticated linguistic and expert-based mapping support to several classifications and nomenclatures. This will require close cooperation between clinicians, knowledge engineers, computational linguists and health informaticians.

Acknowledgements

This work was supported by a grant from the Belgian EBMPracticeNet network (www.ebmpracticenet.be) and by the institutions participating to the MERITERM consortium (www.meriterm.org): Heymans Institute of Pharmacology, University of Ghent, Ghent, Belgium (ugent.be); Fondazione Bruno Kessler, Trento, Italy (fbk.eu); Centre of Expertise in Technology for Information and Communication, Charleroi, Belgium (cetic.be).

REFERENCES

1. Eco U. Dire quasi la stessa cosa, Esperienze di traduzione. [Experiences in translation] Bompiani, 2003. PMid:12514644.
2. Cardillo E, Hernandez G and Bodenreider O. Integrating consumer-oriented vocabularies with selected professional ones from the UMLS using semantic web technologies. Szomszor M and Kostkova P (Eds). *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*. Berlin Heidelberg: Springer, 2010. Available from: http://mor.nlm.nih.gov:8000/pubs/pdf/2010-eHealth-ec.pdf.
3. Jain SH. Practicing medicine in the age of Facebook. *New England Journal of Medicine* 2009;361(7):649–51. Available from: http://dx.doi.org/10.1056/NEJMp0901277. http://dx.doi.org/10.1056/NEJMp0901277.
4. Singh PM, Wight CA, Sercinoglu O, Wilson DC, Boytsov A and Raizada MN. Language preferences on websites and in Google searches for human health and food information. *Journal of Medical Internet Research* 2007;9(2):e18. doi:10.2196/jmir.9.2.e18. Available from: http://dx.doi.org/10.2196/jmir.9.2.e18.
5. Sonnenberg FA, Hagerty CG, Acharya J, Pickens DS and Kulikowski CA. Vocabulary requirements for implementing clinical guidelines in an electronic medical record: a case study. *Proceedings of AMIA Annual Symposium* 2005, pp. 709–13. PMid:16779132; PMCid:PMC1560696.
6. Van Royen P, Chevalier P, Dekeulenaer G, Goossens M, Koeck P, Vanhalewyn M et al. Recommandation de Bonne Pratique Insuffisance cardiaque [Guidelines on Heart Failure]. Belgium: Domus Medica. SSMG, 2011.
7. Roumier J, Jamoulle M, Romary L, Cardillo E and Vander Stichele RH. Towards a terminologies support system in primary care [Letter to the editor]. *Informatics in Primary Care* 2011;19:257–8. PMid:22828581.
8. Cardillo E, Warnier M, Roumier J, Jamoulle M and Vander Stichele RH. Using ISO and semantic web standards for creating a multilingual medical interface terminology: a use case for heart failure. *Proceedings of the 10th International Conference on Terminology and Artificial Intelligence (TIA2013) Paris, IRIT*, 2013, pp. 27–34.
9. de Lusignan S. Codes, classifications, terminologies and nomenclatures: definition, development and application in practice. *Informatics in Primary Care* 2005;13(1):65–70. PMid:15949178.
10. European Commission DE & I. ICT standards in the health sector: current situation and prospects. Special Study No. 1, 2008, p. 82. Available from: http://goo.gl/15R8CQ.
11. Liyanage H, Liaw ST, Kuziemsky C, Terry AL, Jones S, Soler JK et al. The evidence-base for using ontologies and semantic integration methodologies to support integrated chronic disease management in primary and ambulatory care: realist review. Contribution of the IMIA Primary Health Care Informatics WG. *Yearbook of Medical Informatics* 2013;8(1):147–54. PMid:23974562.
12. Smith B and Ceusters W. An ontology-based methodology for the migration of biomedical terminologies to electronic

health records. *Proceedings of AMIA Annual Symposium 2005*, pp. 704–8. PMCid:PMC1560617.

13. Cardillo E, Warnier M, Roumier J, Jamoulle M and Vander Stichele R. Using ISO and Semantic Web standards for creating a Multilingual Medical Interface Terminology: a use case for heart failure. *Proceedings of the 10th International Conference on Terminology and Artificial Intelligence* (TIA2013), October 28, 2013. Available from: http://hdl.handle.net/2268/171534.
14. Okkes IM, Jamoulle M, Lamberts H and Bentzen N. ICPC-2-E: the electronic version of ICPC-2. Differences from the printed version and the consequences. *Family Practice* 2000;17:101–107. Available from: http://dx.doi.org/10.1093/fampra/17.2.101. PMid:10758069
15. World Health Organization. *The ICD-10 Classification of Mental and Behavioral Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva: WHO, 1992.
16. International Health Terminology Standards Development Organization, Copenhagen. *Systematized Nomenclature of Medicine – Clinical Term (SNOMED CT)*. Available from: http://www.ihtsdo.org/snomed-ct/
17. National Library of Medicine, National Institutes of Health, U.S. Department of Health and Human Services. *Unified Medical Language System (UMLS)*. Available from: http://www.nlm.nih.gov/research/umls/.
18. International Health Terminology Standards Development Organization, Copenhagen. *SNOMED-CT Technical Implementation Guide 2014*. Available from: http://www.ihtsdo.org/snomed-ct/.
19. Ceusters W, Rogers J, Consorti F and Rossi-Mori A. Syntactic–semantic tagging as a mediator between linguistic representations and formal models: an exercise in linking SNOMED to GALEN. *Artificial Intelligence in Medicine* 1999;15(1):5–23. Available from: http://dx.doi.org/10.1016/S0933-3657(98)00043-8.
20. Jamoulle M. An information science eye at a general practice/family medicine dedicated guideline. Report to the EBMPracticeNet network. 2011. Part 1 (84p.), Brussels. Available from: http://trix.docpatient.net/files/others/K77_guideline_mj_studydata_part1.pdf
21. Jamoulle M and Roumier J. An information science eye at a general practice/family medicine dedicated guideline. Extension of data categories. Report to the EBMPracticeNet network. 2012 Part 2 (12p.), Brussels. Available from: http://trix.docpatient.net/files/others/K77_guideline_mj_studydata_part2.pdf
22. Eco U. *Mouse or Rat? Translation As Negotiation*. Philippines : Phoenix, 2004.
23. Long W. Extracting diagnoses from discharge summaries. In *Proceedings of AMIA Annual Symposium* 2005, pp. 470–74. PMCID:PMC1560678.
24. Fan JW and Friedman C. Word sense disambiguation via semantic type classification. *Proceedings of AMIA Annual Symposium* 2008, pp. 177–81. PMCid:PMC2655949.
25. Gu HH, Hripcsak G, Chen Y, Morrey CP, Elhanan G, Cimino J et al. Evaluation of a UMLS auditing process of semantic type assignments. *Proceedings of AMIA Annual Symposium* 2007, pp. 294–98. PMid:18693845; PMCid:PMC2655790.
26. Canguilhem G (Ed). The Normal and the Pathological. New York, NY: Zone Books, 1991. PMid:1795225.
27. Lamberts H. In het huis van de huisarts. [In the home of the home doctor]. *Verslag van Het Transitie Project*. Meditekst: Lelystad, 1991.
28. Bowker GC and Star SL (Eds). *Sorting Things Out: Classification and Its Consequences*. Massachusetts, Cambridge, MA: MIT Press, 2000.
29. McCracken EC, Stewart MA, Brown JB and McWhinney IR. Patient-centered care: the family practice model. *Canadian Family physician* 1983;29:2313–6. PMid:20469404; PMCid:PMC2153697.
30. Tierney WM, Overhage JM, Takesue BY, Harris LE, Murray MD, Vargo DL et al. Computerizing guidelines to improve care and patient outcomes: the example of heart failure. *JAMIA* 1995;2(5):316–22. PMid:7496881; PMCid:PMC116272
31. Thorne C, Cardillo E, Eccher C, Montali M and Calvanese D. The VeriCliG project: extraction of computer interpretable guidelines via syntactic and semantic annotation. *Proceedings of the Workshop on Computational Semantics in Clinical Text (CSCT)* (19 March Potsdam, Germany), 2013.
32. Sonnenberg FA, Hagerty CG, Acharya J, Pickens DS and Kulikowski CA. Vocabulary requirements for implementing clinical guidelines in an electronic medical record: a case study. *Proceedings of AMIA Annual Symposium* 2005, pp.709–13. PMC1560696/

Chapter 6

Semantic Web and the Future of Health Care Data in Family Practice

Jamoulle M, Vander Stichele RH, Cardillo E, Roumier J, Grosjean J, Darmoni S. Semantic Web and the Future of Health Care Data in Family Practice. Merit Research Journal of Medicine and Medical Sciences. 2015;3(12):586–94.
http://orbi.ulg.ac.be/handle/2268/189292

Merit Research Journal of Medicine and Medical Sciences (ISSN: 2354-323X) Vol. 3(12) pp. 586-594, December, 2015
Available online http ://www.meritresearchjournals.org/mms/index.htm

Review

Semantic Web and the Future of Health Care Data in Family Practice

Marc Jamoulle[1*], MD, Robert Vander Stichele[2], MD, PhD, Elena Cardillo[3], PhD, Joseph Roumier[4], Ir, Julien Grosjean[5], Ir, PhD, Stéfan J. Darmoni[6], MD, PhD

[1]Family Physician, Department of General practice, University of Liege, Belgium

[2]Family Physician, Prof of Pharmacology, University of Ghent, Belgium

[3]Information scientist, Institute of Informatics and Telematics, Italian National Research Council, Italy

[4]Engineer and Information scientist

[5]Engineer and Information scientist, CISMeF, TIBS, LITIS EA 4108, Rouen University Hospital, Normandy, Rouen, France

[6]Prof of Medical Informatics, CISMeF, TIBS, LITIS EA 4108, Rouen University Hospital, Normandy, Rouen, France and LIMICS, U1142, INSERM, Paris, France

***Corresponding Author's Email: marc.jamoulle@doct.ulg.ac.be**

Abstract

Abstract

New tools in data management have emerged in the last decade, which are going to revolutionize electronic health records and health information systems. Family practitioners are the first, main and continuous contact with the patient. They are the source and endpoint of the circle of information, generated by judicious medical documentation and smart secondary use of medical data. This article is an introduction to health terminologies, ontologies, semantic data, and linked open data, all expressions used by computer scientists, preparing themselves for the next step: semantic web for health care data

Keywords ; biological ontologies; classification; controlled vocabulary

INTRODUCTION

The complex world of patient and family doctors

The acronym GP/FM shows by itself the extent and the complexity of the domain. "General Practice" (GP), used in UK and in continental Europe, refers to what first line doctors are doing in the medical field, while "Family Medicine" (FM), most used in the United States and Canada, is addressing for which people the doctor is working. This apparent opposition between what and for whom, process (practice) and humans (family) reflects the tension between the hard and soft sciences, between biomedical and anthropological views.

GP/FM terminologies address multiple scientific facets of medical knowledge in anthropological/psychological and epidemiological/biological sciences (Van Dormael, 2001). The General Practitioner/Family Physician (GP/FP) becomes an interface between the health system and the patient's world. Being at a crucial network point in the complexity of health care (Plsek and Greenhalgh, 2001), s/he has to use understandable terms (lay terms) to communicate with the patient and his social network (Messai et al., 2008). Links between lay

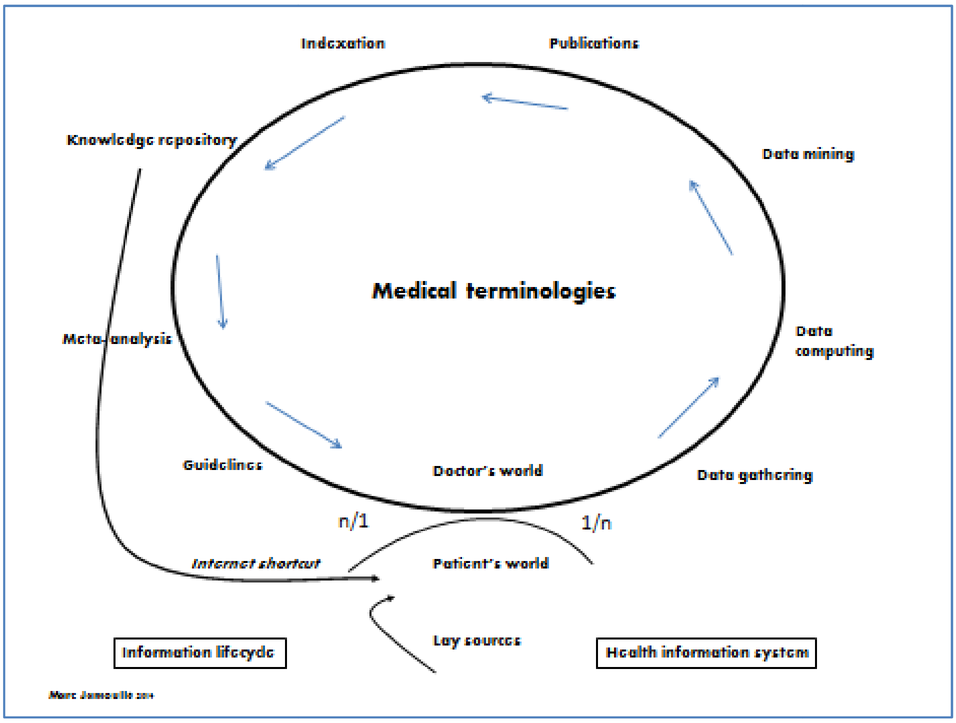


Figure 1. Information lifecycle in medical practice. The consultation is a meeting between patient's and doctor's world of reference

and professional terminologies are becoming a necessity and researchers are working fast in this domain (Cardillo et al., 2014).

Medical data, coming from the health ecosystem, are transformed by research, and finally edited as publications, could create new knowledge and insights (Liyanage et al., 2012). Information used to treat the patient is most often of statistical nature (n/1) while the patient is unique and becomes sources of information (1/n) (Glasziou et al., 1998). In the lifecycle of medical information, guidelines (Woolf et al., 1999) or EBM (Evidence Based Medicine) resources (De Jonghe, 2015) offer a particular way to go back to practice with sets of coherent, comprehensive and consensual recommend-dations which minimize the potential harms to patients (Gray, 2004).

Primary Care terminologies have thus to build bridges between different and distinct worlds: the biomedical world and the patient's world. All this requires a common language or at least common meanings. Hence, there is a crucial role for sophisticated medical terminologies and classifications (Figure 1).

Moreover, patients express their fears and feeling in a particular language, which is sometimes poorly understood by doctors and transformed in preconceived symptoms or disease (Gomes et al., 2015), as shown in Figure 2. This requires a deep professional understanding but also dedicated consumer-oriented terminologies, which bridge the gap between lay language and technical jargon (Cardillo et al., 2010; Marshall, 2000; Smith, 2011).

New tools in information management

Since the seminal proposal of Tim Berners Lee to turn the Internet of documents into an Internet of data (Berners-Lee et al., 2001) (then Internet of information and Internet of knowledge), giant steps have been made by numerous researchers. Gradually the Healthcare Knowledge Management (HKM) field becomes a multidisciplinary area involving family physicians, medical specialists, terminologists, taxonomists, computer scientists, knowledge engineers, and computational linguists, while medical informaticians may often offer the glue between all this heterogeneous competencies.

Figure 2. Bias in patient doctor communication

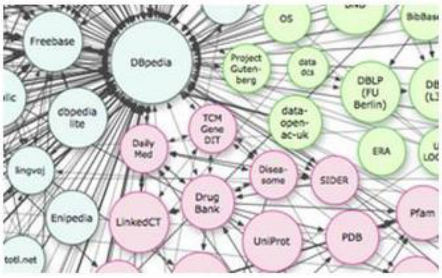


Figure 3. Linked Open Data

The medical semantic web (See Semantic Web Health Care and Life Sciences Interest Group, 2011) and the huge possibility of distributed data have exploded in many different fields: clinical guidelines (Kumar et al., 2004; Camilo et al., 2013; Claudio et al., 2008) consumer information resources (Smith and Fellbaum, 2004; Medline plus http://www.nlm.nih.gov/medlineplus/) mappings between medical classification systems (Cardillo et al., 2008; Ceusters et al., 2005), biomedical data integration (Smith et al., 2007; Ofoghi et al., 2014) semantic interoperability (Qamar, 2008); Linked data (Bizer et al., 2009) e.g. the Bio2RDF framework, to create and provide linked data for the life science (The Bio2RDF frameworkhttp://bio2rdf.org/), medical terminologies and classification systems development or revision (Tudorache et al., 2010; Cramerotti et al., 2015), representation of adverse events (Ceusters et al., 2011), medical education (Blaum et al., 2013), Supervised Machine Learning (SVM) (Pham et al., 2014; Volker et al., 2005-2007), Natural Language Processing (NLP), ontologies (Liu et al., 2011) and many others.

Semantic web and linked data

Semantic web technologies and, above all, Linked data (Figure 3) have emerged as a future solution to exchange scientific data distributed between so many providers as family physicians and colleagues, spread around the globe. The considerable development in the field of medical ontologies shows the vitality of this field of discovery.

Advances in information technology, ontologies, new languages such as the Resource Description Framework (RDF) (Allemang and Hendler, 2008), SKOS (SKOS http://www.w3. org/2004/02/skos/), or SPARQL (Salvadores et al., 2012) for semantic queries, transform the Internet into a huge distributed data base. When associated with Natural Language Processing techniques (Ittoo and Bouma, 2013), what will allow strides in information management in general and family medicine.

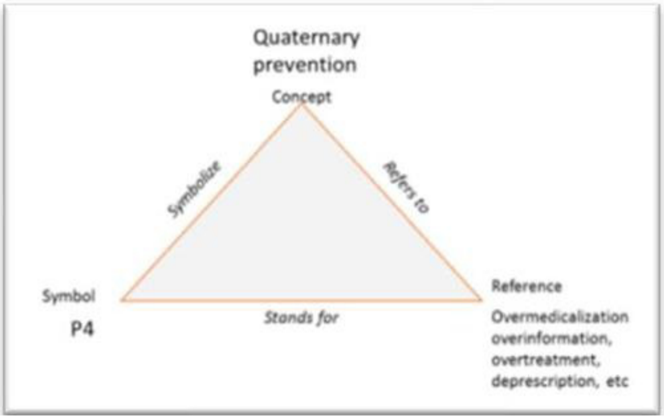


Figure 4. Ogden and Richards triangle (1923) Conceptualization of an item of knowledge.

It brings us closer to fulfill the prediction that "information is to general practice what technology is to specialized medicine" (Van Dormael, 2001).

Ontology, what's in a name?

In its traditional philosophical meaning, Ontology is the study of being or existence (Sadeg-Zadeh, 2015).

In the world of reference of computer scientists, computational linguists, librarians, the word *ontology* has a different meaning. Here it refers to small or huge collections of interlinked concepts, sometimes spread over the Internet, with the aim to describe the knowledge content of a domain in a dedicated machine readable format.

All the fundamental papers in the field are referring to Gruber for a standardized definition: "Ontology is a formal specification of a conceptualization" (Gruber, 1993).

To be informative for general practitioners, this definition needs further explanation of its three terms "formal", "conceptualisation", " and specification".

The first term to understand is the term "formal". Computer scientists are using the term "formal"as a set of symbols understandable by machines. The collections of terms have to be read and understood both by human beings and by machines. Ontologies are not primarily built for human use but for machine use in three types of communications: i) from human to machine, ii) from machine to machine, and finally, iii) from machine to human. So, one can say that ontologies are an arrangement of human knowledge ready to be used by machines.

The second term of Gruber's definition is "conceptualisation". This term has to be understood through the triangle of Ogden and Richards (Figure 4). The concepts, defined as mental representations of reality are symbolized by a machine (and/or human) readable sign, and are finally used to describe specific physical objects or actual practices. All concepts pertaining to a particular domain can be collected and have their relationships defined. Although ontologies often use English labels to describe the concepts to humans, their logical organisation is language-independent, and hence, suitable for multilingual applications. See for further insight the website www.babelnet.org (Navigli and Ponzetto, 2012).

The third and last term of the definition proposed by Gruber is "s*pecification".* It refers to a number of technology tools patiently elaborated by hundreds of computer scientists and computational linguists.

Every single data in a semantic information system receives identification with a Unique Resource Identifier (URI), as proposed by Tim Berners Lee (Berners-Lee, 2002).

Also concepts receive an URI and their origin will be specified as a part of a specific ontological class. For instance, the concept of "feeling tired" will be identified as specific subclass of an OBO Foundry Emotion Ontology (Hastings et al., 2011) on subjective symptoms, perceived by the patient (Figure 5).

Finally, data are no longer recorded in two-dimensional databases of records and fields but are stored in triples. A triple contains a subject, a predicate and an object (Vander Stichele and Dipak, 2015) (Figure 6).

Data structured in triples can be linked by programs. Resource Definition Framework (RDF) is the name of one

```
<!-- http://purl.obolibrary.org/obo/MFOEM_000080->
<owl:Class rdf:about="&obo;MFOEM_000080">
<rdfs:label>feeling tired</rdfs:label>
<rdfs:subClassOf rdf:resource="&obo;MFOEM_000006"/>
<obo:MFOEM_000165>tired</obo:MFOEM_000165>
<obo:IAO_0000115>The subjective emotional feeling
of tiredness, needing sleep.</obo:IAO_0000115>
</owl:Class>
```

Figure 5. The concept « feeling tired » in the Ontology of emotion in RDF (Obo Foundry)

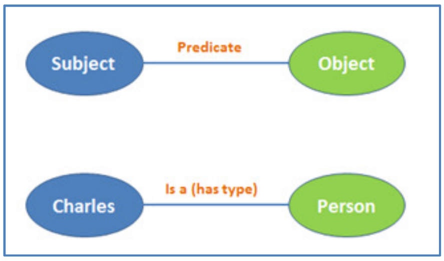


Figure 6. The triplet

of the languages which allow the link of data elements with other data elements. In addition, a whole stack of tools has been build which allow to query, organise and reformat information stored in disparate places on the internet.

Linked data and data differentiation

The linked data cloud on http://linkeddata.org/ shows billions of interlinked data elements published in RDF and submitted to queries by dedicated tools. Whatever is the activity (e.g. business, cars selling, books selling, civil affairs management or health care) there are a lot of sites already managing linked-data.

The difference with the Internet “of documents” is striking. If one queries the BBC linked data web site asking information about an artist, he/she will not receive a link to an already prepared page, but rather it will be created instantly based on semantic data (including semantic relations). This is achieved by retrieving data from the relevant semantic repositories that can be updated independently of the BBC.

The semantic web scientists do not use the concept “retrieved” for this, they use the strange term “dereferencing data” instead. It means that the knowledge needed on a subject is collected based on the URI that identifies it. As URIs is the addressing system of the internet, internet is now the database, or better the “knowledge base”

Practical application in health care terminology: the HeTOP portal

Hetop.eu interactive web site offers cross lingual multi-terminological mappings on a semantic basis (Grosjean et al., 2012). Developed initially on MeSH and French mapping of MeSH for French speaking users, it has evolved towards a dynamic semantic interface supporting a two-dimensional navigation across terminologies/ ontologies and languages. It also provides access to PubMed via the InfoRoute infobutton which is a powerful tool to leverage Medline querying based on the semantic mappings (Thirion et al., 2007) (Figure 7)

The future of health care information systems

The next EHR

Now imagine that the data elements on the patients are not sent by e-mail to GPs electronic medical record (EMR) anymore but that, with the due authorizations and encryption, GP is allowed to dereference the needed data in all the hospitals or primary care settings which store information about his/her patient in the correct format. GP will reconstitute in seconds and only for the time needed the current record of his/her patient. Health-Data becomes Linked-Data (Dowling, 2015). As stated by Pierce et al., “*Semantic Web technologies offer the potential to revolutionize management of health care data*

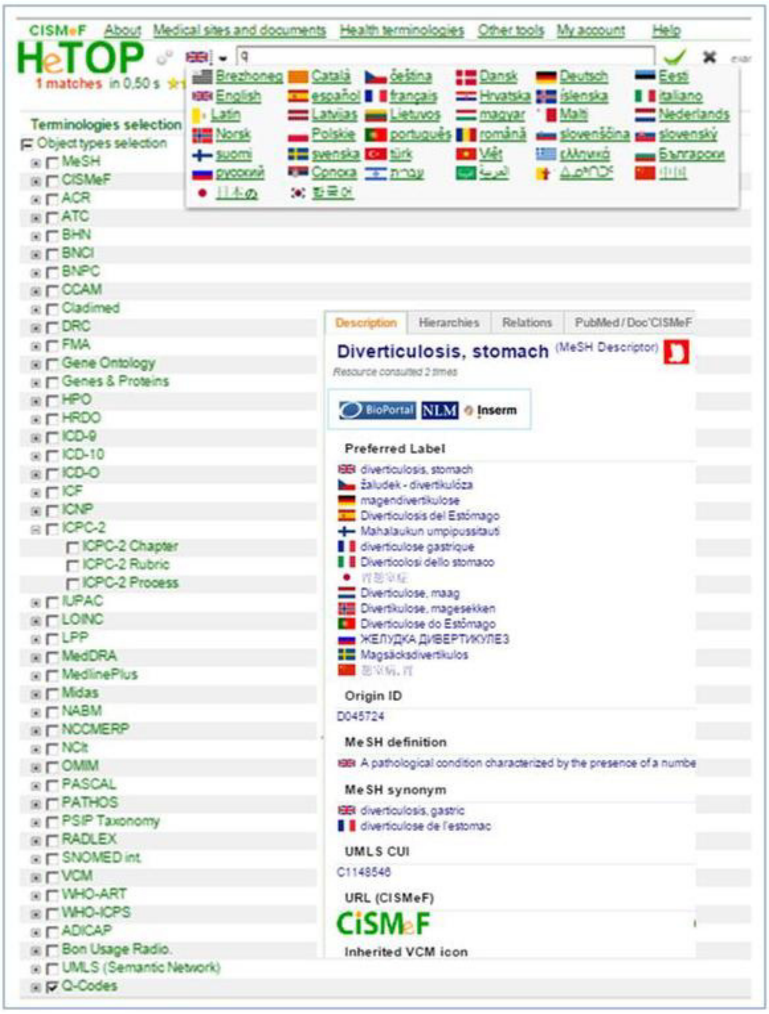


Figure 7. The cross-lingual multi-terminology HeTOP portal and an opening of a MeSH with tabs for access to bibliographic databases

by increasing interoperability and reusability while reducing the need for redundant data collection and storage" (Pierce et al., 2014). This is the future of health care information system and of electronic health records (EHR). This is behind our door (Fernández-Breis et al., 2013). That is why scientists are working so hard to develop medical ontologies like the Open Biomedical Ontologies (OBO) consortium (Smith et al., 2007), National Center for Biomedical Ontology (Musen et al., 2011) or Linking Open Drug Data (LODD) for pharmaceutical research (Samwald et al., 2011).

Natural Language Processing techniques like Information Extraction or Supervised Machine Learning and Data mining can be applied to discover terms (e.g. smoking, cancer) and their semantic relationships (e.g. "causes" as in "smoking causes cancer") from huge text corpora such as patient records. These discovered terms and relationships can be used to populate/augment existing Linked Data ontologies or to create new ones (Fahmi, 2009). EHR warehouses are currently growing all around the world. Multiple research projects are coming with them to ease Information Retrieval and data visualization in EHRs. Some goals are already reached today with complex systems such as I2B2 (Integrating Biology and the Bedside) (Informatics for integrating biology and the bedside. URL: https://www.i2b2.org). These frameworks are not so friendly for physicians because they are dealing with complex models, huge data and heterogeneous information. However, the future is right there with exciting perspectives: intelligent semantic search among EHRs, patient cohorts automatic creation, decision making support, etc (Garde et al., 2007).

Interlinked publications

In the same way, the gray literature and non-published works like more than 50% of abstracts presented by doctors in congresses (Post et al., 2013) could be tagged by semantic specification and stay in the local database of congress organizers or local organizations. Then, a semantic web tool could crawl these websites, dereferencing when needed the asked information through semantically interoperable indexation systems.

One hopes to apply such techniques to an indexing system for communications of family doctors, and link them through the use of dedicated ontologies in a semantic web GP/FM universe.

CONCLUSION

Health knowledge management is a difficult but promising new research domain, in which family physicians should be embedded as the main interface between the healthcare system and the patient, with the Electronic Health Record as the main tool. Family physicians should develop their understanding of and their capacities to be involved in the practical applications of this domain.

AKNOWLEDGEMENT

This paper has been edited thanks to a grant from the private health centre 'Espace Temps, Maison de Santé', Gilly, Belgium in which the first author is practicing as family doctor since 41 years

REFERENCES

Allemang D, Hendler JA (2008). Semantic Web for the Working Ontologist: Effective Modeling in Rdfs and Owl. Morgan Kaufmann Publishers.

Berners-Lee T (2002). « What do HTTP URIs Identify? W3C. http://www.w3.org/DesignIssues/HTTP-URI

Berners-Lee T, Hendler J, Lassila O (2001). The Semantic Web. Scientific American,.284(5):34-43

Bizer C, Heath T, Berners-Lee T (2009). Linked data-the story so far, International Journal on Semantic Web and Information Systems, 5(3), 1-22

Blaum WE, Jarczweski A, Balzer F, Stötzner P, Ahlers O (2013). Towards Web 3.0: taxonomies and ontologies for medical education -- a systematic review. GMS Z Med Ausbild.;30(1):Doc13. doi:10.3205/zma000856.

Camilo Thorne, Elena Cardillo, Claudio Eccher, Marco Montali, Diego Calvanese (2013). The VeriCliG Project: Extraction of Computer Interpretable Guidelines via Syntactic and Semantic Annotation, Computational Semantics in Clinical Text (CSCT), Germany, March 19).

Cardillo E, Eccher C, Serafini L, Tamilin A (2008). Logical Analysis of Mappings between Medical Classification Systems. In: Dochev D, Pistore M, Traverso P, eds. Artificial Intelligence: Methodology, Systems, and Applications SE - 26.Vol 5253. Lecture Notes in Computer Science. Springer Berlin Heidelberg; 311–321. doi:10.1007/978-3-540-85776-1_26.

Cardillo E, Hernandez G, Bodenreider O (2010). Integrating consumer-oriented vocabularies with selected professional ones from the UMLS using SemanticWeb Technologies. In: Szomszor M, Kostkova P, eds. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Springer Berlin Heidelberg;. Available at: http://dx.doi.org/10.1007/978-3-642-23635-8_7.

Cardillo E, Tamilin A, Serafini L (2010). A Methodology for Knowledge Acquisition in Consumer-oriented Healthcare. In Knowledge Discovery, Knowledge Engineering and Knowledge Management: IC3K Revised Selected Papers, pp. 64 -71, Springer Berlin

Cardillo E, Folino A (2014). Towards the reuse of standardized thesauri into ontologies. Workshop on Ontology and Semantic Web Patterns (WOP2014). 2014. 8p ; Available from: http://ceur-ws.org/Vol-1302/paper3.pdf

Ceusters W, Capolupo M, de Moor G, Devlies J, Smith B (2011). An evolutionary approach to realism-based adverse event representations. Methods Inf Med.;50(1):62–73. doi:10.3414/ME10-02-0016.

Ceusters W, Smith, B, De MoorG (2005). Ontology-Based Integration of Medical Coding Systems and Electronic Patient Records. In MIE2005.

Claudio Eccher, Antonella Ferro, Andreas Seyfang, Marco Rospocher, Silvia Miksch (2008). Modeling Clinical Protocols Using Semantic MediaWiki: The case of the Oncocure Project, in: David Riano (eds.), Knowledge Management for Health Care Procedures. ECAI 2008 Workshop, K4Help, Revised Selected Papers, Berlin Heidelberg, Springer, 2009, pp. 42 -54.

De Jonghe M. (2015) Quaternary prevention as a guideline for the editorial team of an evidence-based medicine journal. Rev Bras Med Família e Comunidade.10(35):1. Available from: http://www.rbmfc.org.br/rbmfc/article/view/1067

Dowling C. (2015) Using RDF to produce one coherent, definitive dataset about drugs. Part 3 of Yosemite Series. Webinar, Recorded August 6, 2015 available at http://yosemiteproject.org/2015/ webinars/drugdocs/

Cramerotti S, Buccio M,Vaschetto G,Rospocher M, Serafini L, Cardillo E, Donadello Y.(2014) ePlanning: an Ontology-based System for Building Individualized Education Plans for Students with Special Educational Needs. Proceedings of the Industry Track at the International Semantic Web Conference 2014 ({ISWC} 2014 CEUR-WS.org. Workshop Proceedings, http://ceur-ws.org/Vol-1383/paper7.pdf

Fahmi I (2009). Automatic term and relation extraction for medical question answering system. Groningen University: 244p. http://dissertations.ub.rug.nl/faculties/arts/2009/i.fahmi/.

Fernández-Breis JT, Maldonado JA, Marcos M, et al. Leveraging electronic healthcare record standards and semantic web technologies for the identification of patient cohorts. *J Am Med Inform Assoc*. 2013;20(e2):e288–96. doi:10.1136/amiajnl-2013-001923.

Garde S, Knaup P, Hovenga E, Heard S. Towards semantic interoperability for electronic health records. Methods Inf Med 2007;46 (3), 332-43. doi:101160/ME5001

Glasziou P, Guyatt GH, Dans AL, Dans LF, Straus S, Sackett DL (1998). Applying the results of trials and systematic reviews to individual patients. ACP J Club. Nov-Dec;129(3):A15-6.

Gomes L, Gusso G, Jamoulle M (2015). Learning and teaching quaternary prevention. Revista Brazileira de Medicina Familiar y Communautaria.

Gray JAM (2004). New concepts in screening. Br J Gen Pract.;54(501):292–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1314856&tool=pmcentrez&rendertype=abstract.

Grosjean, J; Kerdelhué, G; Merabti, T & Darmoni, SJ (2012). The HeTOP: indexing Health resources in a multi-terminology/ontology and cross-lingual world. EAHIL.http://ceur-ws.org/Vol-952/paper_17.pdf

Gruber TR (1993). A Translation Approach to Portable Ontology Specifications. Knowl Acquis.;5(April):199–220. http://dx.doi.org/10.1006/knac.1993.1008

Hastings J, Ceusters W, Smith B, Mulligan K (2011). Dispositions and Processes in the Emotion Ontology. Proceedings of the 2nd International Conference on Biomedical Ontology. CEUR Workshop Proceedings, 833; Available from: http://philpapers.org/rec/HASDAP

Informatics for integrating biology and the bedside. URL: https://www.i2b2.org

Ittoo A, Bouma G (2013). Minimally-supervised learning of domain-specific causal relations using an open-domain corpus as knowledge base. Data Knowl Eng.;88:142–163.

Kumar A, Smith B, Pisanelli DM, Gangemi A, Stefanelli M (2004). An ontological framework for the implementation of clinical guidelines in health care organizations. Stud Health Technol Inform.;102:95–107. http://www.ncbi.nlm.nih.gov/pubmed/15853266.

Liu K, Hogan WR, Crowley RS (2011). Natural Language Processing methods and systems for biomedical ontology learning. J Biomed Inform.; 44(1):163–79. doi:10.1016/j.jbi.2010.07.006.

Liyanage H, Liaw ST, De Lusignan S (2012). Accelerating the development of an information ecosystem in health care, by stimulating the growth of safe intermediate processing of health information (IPHI). Inform Prim Care. 20(2):81–6. DOI: http://dx.doi.org/10.14236/jhi.v20i2.28

Marshall PD (2000). Bridging the Terminology Gap Between Health Care Professionals and Patients with the Consumer Health Terminology (CHT), in Proceedings of AMIA Annual Symposium, AMIA2000, p. 1082

Medline plus http://www.nlm.nih.gov/medlineplus/

Messai R, Simonet M, Mousseau M (2008). Bridging the Gap between Lay Language and Medical Terminology. In: IPCV'08 - The 2008 International Conference on Image Processing, Computer Vision, and Pattern Recognition.;. Available at: http://nguyendangbinh.org/Proceedings/IPCV08/Papers/IKE4547.pdf

[1]Musen MA, Noy NF, Shah NH, Whetzel PL, Chute CG, Story MA, Smith B; NCBO team. The National Center for Biomedical Ontology. J Am Med Inform Assoc. 2012 Mar-Apr;19(2):190-5. Epub 2011 Nov 10.

Navigli R.and Ponzetto S (2012). BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network. Artificial Intelligence, 193, Elsevier, pp. 217-250.

Ofoghi B, Lopez-Campos GH, Martin Sanchez FJ, Verspoor K (2014). Mapping Biomedical Vocabularies: A Semi-Automated Term Matching Approach. In Studies in Health Technology and Informatics, Volume 202: Integrating Information Technology and Management for Quality of Care, pp. 16 – 19. doi=10.3233/978-1-61499-423-7-16.

Pham A-D, Névéol A, Lavergne T, Yasunaga D, Clément O, Meyer G, Morello R, Burgun A (2014) Natural language processing of radiology reports for the detection of thromboembolic diseases and clinically relevant incidental findings. BMC Bioinformatics, 15(1), 266. http://doi.org/10.1186/1471-2105-15-266

Pierce CD, Booth D, Ogbuji C, Deaton C, Blackstone E, Lenat D. SemanticDB: A Semantic Web Infrastructure for Clinical Research and Quality Reporting. *Curr Bioinform*. 2014;7(3). Available at: http://benthamscience.com/journal/abstracts.php?journalID=cbio&articleID=101604.

Plsek PE, Greenhalgh T (2001). Complexity science: The challenge of complexity in health care. BMJ.;323(7313):625–8. Available at: http://www.pubmedcentral.nih.gov/1121189 .

Post RE, Mainous AG, O'Hare KE, King DE, Maffei MS. Publication of research presented at STFM and NAPCRG conferences. Ann Fam Med. 2013;11(3):258-61. doi:10.1370/afm.1503.

Qamar R (2008). Semantic mapping of clinical model data to biomedical terminologies to facilitate interoperability . 2008:257. Available at: http://www.openehr.org/publications/archetypes/RQamar-PhD-Thesis-Mar2008.pdf.

Sadeg-Zadeh K.(2015) Handbook of Analytical Philosophy of Medicine. Chap. 23 medical Ontology. 2015 p. 773–804.

Salvadores M, Horridge M, Alexander PR, Fergerson RW, Musen MA, and Noy NF (2012). Using SPARQL to Query BioPortal Ontologies and Metadata. International Semantic Web Conference. Boston US. LNCS 7650, pp. 180195

Samwald M, Jentzsch A, Bouton C, Kallesøe CS, Willighagen E, Hajagos J, Marshall MS, Prud'hommeaux E, Hassenzadeh O, Pichler E, Stephens S. (2011) Linked open drug data for pharmaceutical research and development. J Cheminform . 3(1):19. Available from: http://www.jcheminf.com/content/3/1/19

See Semantic Web Health Care and Life Sciences Interest Group (2011). - http://www.w3.org/2011/09/HCLSIGCharter.html SKOS http://www.w3.org/2004/02/skos/

Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. Nature Publishing Group. 25(11):1251–5. Available from: http://dx.doi.org/10.1038/nbt1346

Smith B, Ashburner M, Rosse C, et al., (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol.;25(11):1251-5. doi:10.1038/nbt1346.

Smith B, Fellbaum C (2004). Medical WordNet : A New Methodology for the Construction and Validation of Information Resources for Consumer Health. In: Proceedings of Coling: The 20th International Conference on Computational Linguistics, Geneva; 31–38.

Smith CA (2011). Consumer language, patient language, and thesauri: a review of the literature. J Med Libr Assoc. Apr;99(2):135-44. DOI= 10.3163/1536-5050.99.2.005.

The Bio2RDF frameworkhttp://bio2rdf.org/

Thirion, B; Pereira, S; Névéol, A; Dahamna, B and Darmoni, SJ (2007). French MeSH Browser: a cross-language tool to access MEDLINE/PubMed.. AMIA symp., Pages 1132, IOS Press

Tudorache T, Falconer S, Nyulas C, Storey M-A, Ustün TB, Musen MA (2010). Supporting the Collaborative Authoring of ICD-11 with WebProtégé. AMIA Annu Symp Proc.; 2010:802–806. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21347089.

Vander Stichele R, Dipak K. (2015) Report to clinical stakeholders on the creation and implementation of domain-specific semantic interoperability resources. Deliverable 3.3 of the Semantic Interoperability for Health Network of Excellence. http://www.semantichealthnet.eu/index.cfm/deliverables/

Van Dormael M (2001). Identités professionnelles en médecine générale et en soins de santé primaires.[Professional identity in GP/FM and in Primary Care] Chap VI in Hours B (ed): Systèmes et politiques de santé. De la santé publique à l'anthropologie. Karthala: Médecines du Monde;

Volker T, Markus B, Achim R, Yi H (2005-2007). Towards Machine Learning on the Semantic Web. In Uncertainty Reasoning for the Semantic Web I. ISWC International Workshops, URSW 2005-2007, Revised Selected and Invited Papers, Lecture Notes in Computer Science, Springer Berlin Heidelberg 2008, vol. 5327, pp 282-314. DOI= 10.1007/978-3-540-89765-1_17

Woolf SH, Grol R, Hutchinson A, Eccles M, Grimshaw J (1999). Potential benefits, limitations, and harms of clinical guidelines. BMJ : British Medical Journal;318(7182):527-530.

Chapter 7

Analysis of definitions of General Practice/Family Medicine and Primary Health Care.

Jamoulle, M., Resnick, M., Ittoo, A., Cardillo, E., Vander Stichele, R., & Vanmeerbeek, M. (2017). Analysis of definitions of General Practice/Family Medicine and Primary Health Care. British Journal of General Practice - Open, accepted.
http://orbi.ulg.ac.be/handle/2268/210049

BJGP Open

Analysis of definitions of general practice, family medicine, and primary health care: a terminological analysis

Marc Jamoulle, MD[1*], Melissa Resnick, MLS[2], Robert Vander Stichele, MD, PhD[3], Ashwin Ittoo, PhD[4], Elena Cardillo, PhD[5], Marc Vanmeerbeek, MD, PhD[1,6]

[1]GP, Researcher, and PhD applicant, Department of General Practice, University of Liège, Liège, Belgium; [2]Medical Librarian and PhD applicant, Health Science Center, University of Texas at Houston, Houston, TX, US; [3]GP and Professor of Pharmacology, Heymans Institute of Pharmacology, University of Ghent, Ghent, Belgium; [4]Associate Professor in Health Information Systems, HEC Management School, University of Liège, Liège, Belgium; [5]Senior Researcher in Computational Linguistics, Institute of Informatics and Telematics, National Research Council, Rende, Italy; [6]GP and Professor of General Practice, Department of General Practice, University of Liège, Liège, Belgium

Abstract

Background: There are numerous definitions of general practice/family medicine (GP/FM) and primary health care (PHC), but the distinction between the two concepts is unclear.

Aim: To conduct a terminological analysis of a set of definitions of GP/FM and of PHC, to clarify the commonalities and differences between these two concepts.

Design: Sets of 20 definitions were collected in two 'bags of words' (one for GP/FM and one for PHC terms). A terminological analysis of these two collections was performed to prioritise the terms and analyse their universe of discourse.

Method: The two collections were extracted with VocabGrabber®, configured in two 'term clouds' using Wordle®, and further explored for similarities using Tropes®. The main terms were analysed using the Aristotelian approach to the categorisation of things.

Results: Although continuity of care (characterised by a person-centred approach and shared decision making) is common to both sets, the two sets of definitions differ greatly in content. The main terms specific to GP/FM (*community, medicine, responsibility, individual, problem,* and *needs*) are different from those specific to PHC (*home, team, promotion, collaborator, engagement, neighbourhood,* and *medical* centre).

Conclusion: Terminological analysis of the definitions for GP/FM and PHC shows two overlapping but distinct entities, necessitating a different taxonomic approach and different bibliographic search strategies.

***For correspondence:** marc@jamoulle.com

Competing interests: The authors declare that no competing interests exist.

Received: 25 January 2017
Accepted: 27 March 2017
Published: 06 September 2017

Author Keywords: primary health care, general practice, terminology as topic, qualitative research

DOI:10.3399/bjgpopen17X101049

How this fits in

There are numerous definitions of GP/FM and PHC. The governance of these concepts is related to their use in two distinct organisations: the World Organization of family doctors (WONCA) and the World Health Organization (WHO). In GP/FM textbooks and bibliographic retrieval systems, there is often confusion between these concepts. A clear understanding of the similarities and differences between the two concepts is needed for the organisation of medical training, for the development of the profession and of health policy, and for optimal information and retrieval in this scientific discipline.

Introduction

General practice designates a branch of medicine characterised by its broad scope. The term general, also extended to generalism,[1] encompasses the comprehensive range of transactions performed, and thus the scope and nature of the work of the practitioner.

Family medicine emphasises the relationship with the patient and seeing the person as a whole, in the context of their family (next of kin or relevant others) and their wider community. The WONCA dictionary states: 'Many medical practitioners in the primary health care prefer the terms family physician and family medicine in order to emphasise the recognition of their branch of medical practice as a specialty in its own right.'[2] In other countries, other terms are used such as general practitioner (UK), 'hausart' (Germany), 'huisarts' (Netherlands), 'me´decin ge´ne´raliste' and 'me´decin de famille' (France), and family physician (US). WONCA has always used the pair of acronyms GP/FM in order to present and discuss the situation, taking into account the members of this professional organisation. Hence, GP/FM is a people-oriented profession aiming at the management of an extended and general set of human health problems.[3] Core values of GP/FM have been extensively discussed. Patient-centredness, as well as the biopsychosocial model, are now definitely considered as undisputable attributes of a profession directed towards building personal relationships during the patient's lifetime.[4]

The concept of primary health care (PHC), endorsed by the WHO in 1978 at Alma-Ata, is an organisational concept.[5] It addresses the place, management, and workload of the first (primary) level of health care, as well as its inclusion in the network of care facilities. 'Strong primary health care is the foundation of healthy communities' remains a WHO motto.

The aim of this study was to conduct a terminological analysis of a set of definitions of GP/FM and of PHC, in order to clarify the commonalities and differences between these two concepts.

Method

To construct a set of relevant definitions for each of the two concepts (GP/FM and PHC), a search of PubMed, Google Scholar, Global Index Medicus (the WHO bibliographic database),[6] and books related to the discipline was made. For GP/FM, the following keywords were used: *family practice; general practice; general practitioners; physicians, family; physicians, primary care*. For PHC we used: *primary health care; community health centres; community health services; rural health services; home care services*.

Definitions that were repetitive or yielded no further information were disregarded. We aimed for geographical and cultural spread, stopping after 10 definitions for each concept, because new definitions did not provide any additional significant information.

Furthermore, a terminological analysis of these two sets of 10 definitions for GP/FM and PHC was performed to prioritise the terms used in each of the two sets. To this end, we first targeted the key vocabulary in the definitions by using VocabGrabber®, a text analysing tool, which ranks the relevance of all of the words appearing in a source text by comparing the frequency of their use in the presented text to their overall frequency of use in written English (https://www.visualthesaurus.com/vocabgrabber). In this system, the more frequent words can be displayed in a tabular list with the numerical frequency and relevance of each word shown, or in a semantic map with a view of the relationships between words and meanings. The relative relevance of terms can be displayed in a 'tag cloud' through the use of a specific 'word cloud' generator such as Wordle® (http://www.wordle.net). Here words that appear more frequently in the source text are given greater prominence in the cloud (they appear in a larger font).

In addition, we used Tropes®, a natural language processing software program designed for semantic classification, keyword extraction, and linguistic and qualitative analysis. (http://tropes.fr/).

Finally, the prioritised terms within each set of definitions and their semantic relationships were then used to perform a comparative analysis of the two concepts (GM/FM and PHC). To clarify what links and what separates the two concepts, we used the classical category theory approach of Greek philosopher Aristotle (4th century BCE), in which the meaning of a term is explored by asking 10 fundamental questions about the universal categories of things:

1. essence;
2. quantity;
3. quality;
4. relation;
5. place;
6. time;
7. posture;
8. state;
9. action;
10. passion.

This approach is still used, for example, in the development of taxonomies and ontologies to identify relevant concepts of a domain of application and to categorise these concepts.[7,8]

Results

Twenty definitions (10 relating to GP/FM and 10 relating to PHC) were selected from the results of a larger, exhaustive search. These definitions were in English, Spanish, Portuguese, and French, spanning Europe, the US, Canada, South America, Australia, and India. The dates of the 10 GP/FM definitions ranged from 1974 to 2016, while those of the 10 PHC definitions ranged from 1996 to 2016 (**Box 1**).

Using the above-mentioned VocabGrabber® tool, 319 words were taken from the GP/FM set of definitions, and 262 words from the PHC set, and displayed in two tag clouds generated using Wordle® (**Figure 1** and **Figure 2**).

Words or compound words mentioned at least three times in both sets of definitions, or appearing in only one set, analysed using Tropes® software, are outlined below (**Box 2**).

Finally, the main terms were analysed, using the 10 seminal Aristotelian categories of things,[26] and integrated to a statement in response to the philosophical questions, relevant for each category, for each of the two sets (**Box 3**).

Both sets share the terms *continuity of care*, *patient centredness*, *community health*, and *shared decision making*. Although *care* is the central issue of the two sets, they differ greatly in content.

Box 1. Sources of the two sets of 10 definitions of GP/FM and PHC.

General Practice/Family Medicine definitions	Primary Health Care definitions
Leeuwenhorst definition (1974)[9]	Institute of Medicine (1996)[10]
AAFP primary care physician definition (US) (1977)[11]	PAHO primary health care statement (Pan America) (2007)[12]
AAFP family medicine definition (US) (1984)[13]	EU expert panel definition of primary care (2014)[14]
Olesen's proposal for a new definition of general practice (2000)[15]	Brazil: organisation of primary health care (2013)[16]
WONCA dictionary (2003)[2]	WHO glossary (2016)[17]
WONCA Europe / EURACT definition (2011)[18]	PHCRIS (Australia) (2015)[19]
CIMF Carta de Quito definition (Latin America) (2014)[20]	FMMCSF (Belgium) (2016)[21]
The Role Definition Group definition (US) (2014)[22]	AHRQ Primary Care Medical Home model (US) (2016)[23]
NBE definition (India) (2015)[24]	FFMPS (France) (2016)[25]
AAFP Primary Care Physician (2016)[11]	AAFP Primary Care (US) (2016)[11]

AAFP = American Academy of Family Physicians. AHRQ = Agency for Healthcare Research and Quality. CIMF = Confederación Iberoamericana de Medicina Familiar / Iberoamerican Confederation of Family Medicine. EU = European Union. FMMCSF = Fédération des Maisons Médicales et des Collectifs de Santé Francophone / Federation of Medical Homes and French-speaking Health Centres. FFMPS = Fédération Française des Maisons et Pôles de Santé / French Federation of Housing and Health Centers. NBE = National Board of Examination. PAHO = Pan American Health Organisation. PHCRIS = Primary Health Care Research and Information Service. WHO = World Health Organization. WONCA = World Organization of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians.

Figure 1. Tag cloud for General Practice/Family Medicine.

As indicated in *Box 2*, the main terms specific to each set differ greatly. GP/FM is determined by such terms as *medicine*, *responsibility*, *individual*, *problem*, *disease*, and *peculiarity*. PHC is quite service oriented with *home*, *team*, *promotion*, *collaborator*, *engagement*, *neighbourhood*, and *medical centre*.

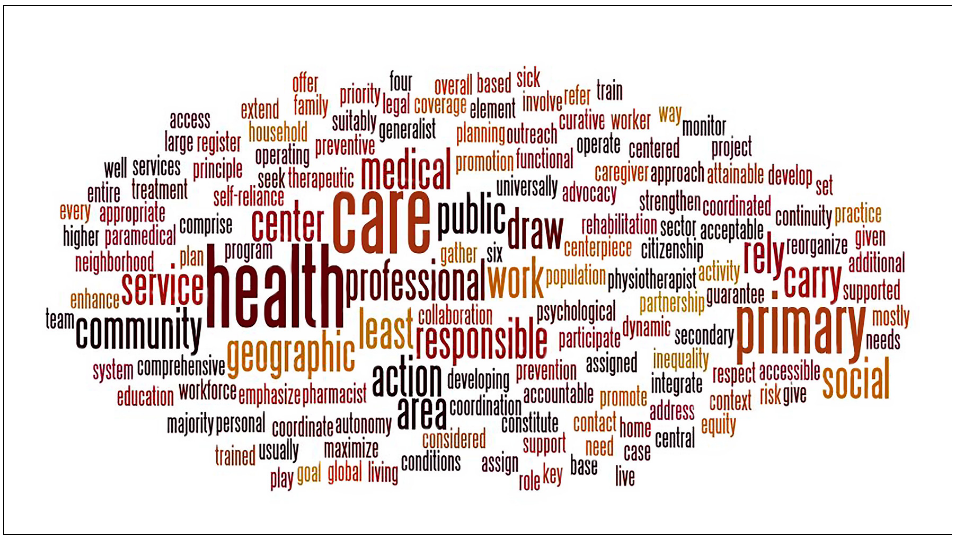


Figure 2 Tag cloud for Primary Health Care.

Box 2. Terms that unite and separate the two concepts: GP/FM and PHC.

What unites	• Listed at least three times in both sets	Care, health, patient, service, family, community, health care, system, prevention, doctor, population, needs, provision, junction.
What separates	• Listed at least three times only in GP/FM set	Medicine, responsibility, individual, general practitioner, sex, illness, disease, problem, peculiarity, specialist, factor, management, science, basis, age, resource, point.
	• Listed at least threetimes only in the PHC set	Home, team, promotion, person, part, activity, health professional, righteousness, nurse, majority, action, professional, partnership, access, level, improvement, time, insurance, collaborator, engagement, neighbourhood, medical centre.

Among the top 10 terms for both GP/FM and PHC concepts, the terms *global health*, *environmental hazard*, *ethics*, economic aspects, and the recent concept of quaternary prevention (danger

Box 3. Analysis of the distribution of the main terms of the 10 definitions of GP/FM and PHC according to Aristotle's categories of things.

Aristotelian categories Greek and Latin Translations	Questions	General Practice/Family Medicine words	Primary Health Care words
Essence οὐαία *(ousia)* *Quod est?* *Essentia*	What is it? Essence or substance?	Licensed medical graduate who provides care, specialty charterised by breadth, primary care services, take care, promotion of health, prevention of disease, early diagnosis, initial decision, provision of clinical care, rehabilitation, palliative care, education, research	Clinician provides healthcare services, care, health, prevention, promotion, first-contact primary care, intersectoral action, health promotion, illness prevention, treatment and care of the sick, community development, rehabilitation
Quantity Ποσόν *(poson)* *Quantum,* *Quatitas*	How much, how many, how tall?	General, every, both sexes, all age, irrespective of age, each organ system, every disease entity, repeated contacts, entire family	Large majority, any, set, variety, every family, entire population
Quality Ποιόν *(poion)* *Quale* *Qualitas*	How is it? What kind or quality?	Personal, access, available, comprehensive, effective, necessary, personal, respecting autonomy, safety, satisfaction, sustainability	Universal coverage, comprehensive, integrated, person-centred, accessible, socially appropriate, critical, effective, scientifically sound, include, partner, professional, specifically, undiagnosed, undifferentiated, whole-person care
Relation πρός τι *(pros ti)* *Relativum*	What is it related to? Towards something?	In the context of their family, their community, and their culture, family doctor, general practitioner, individual, population, undifferentiated patient, cultural diversity	Multiprofessional health teams, partnership, patient, caregivers, population, family, communities, local network
Place ποῦ *(pou)* *Ubi*	Where?	Where necessary, at the point of first contact, entry point, in the front line, consulting room, homes, acute and chronic care settings	Set of functional and structural elements, home, office, setting, coverage area, geographic, territory
Time Πότε *(pote)* *Quando*	When?	First contact, prolonged contact, continuing, repeated, maintaining, always, preventing, chronic, recurrent, terminal	First contact, first level, over time, prevention, primary, continuing, acute, chronic, limited, long term
Posture κεῖσθαι *(keisthai)* *Situ*	From what action does it result?	Autonomy, balance, basis, clinical, contact, cultural, disease, existential, health, illness, needs, self, physical, biomedical, psychological, social and behavioural sciences	Autonomy, behavioural, biological, communication, concern, consultation, contact, disease, health, illness, organ, problem, sign, social, symptom, living conditions, health risks, health status, health inequalities
State ἔχειν *(echein)* *Habitus*	What is it required to have or be?	Socially responsible, reliable, leader, professional, advocate, trust, knowledge, personal balance and value	Equity enhancing, responsible, concern, accountable, cost, role, professional, self-reliance, participation and control, advocacy, social justice, equity, solidarity
Action ποιεῖν *(poiein)* *Agere*	What is it doing? (change), to make or do	Provide, train, integrate, intervene, promote, maintain, prevent, serve, manage, practice, define, optimise, negotiate, coordinate, monitor, devote, gather information, organise, assist	Perform, participate, measure, utilise, monitor, understand, reorganise, maximise, collaborate, assess, inform, integrate, gather, encourage, enable
Passion Πάσχειν *(paschein)* *Pati*	How is it being acted upon (be changed)?	Must be trained, developing and maintaining their skills, personal balance and values, discipline, professional role	Accomplish, appropriate, perform, skilled, trained

of overmedicalisation)[27] are almost absent. None of the definitions specifically addressed medical anthropology. Only in the GP/FM definition from Latin America ('*Carta de Quito*' [letter from Quito]),[20] are the terms *sustainability* and *social responsibility* mentioned.

Discussion

Summary

To the best of our knowledge, this is the first terminological analysis of the terms used to depict workforce and structure of primary-level care as found in published definitions of GP/FM and PHC.

Although continuity of care (characterised by a person-centred approach and shared decision making) is core to the two sets, the two sets of definitions differ greatly in content.

The main terms obtained from an analysis of 10 definitions of GP/FM pertain to a professional discipline, conducted by *practitioners* who are *responsible physicians* shaped by *science* and who care for *family problems* in the context of a *social* role.

The main terms from the 10 definitions of PHC still speak of *care* and *health* as central elements but, here, it is a *service* to the population made by *unspecified professionals* in a *geographic area*.

Strengths and limitations

This study provides an innovative method to examine the nature of GP/FM and PHC through a terminological analysis.

The prioritisation of terms based on software tools may be subject to variation over time, as tools evolve. The qualitative interpretation of the terminological findings is a potentially subjective process that needs further validation.

Comparison with existing literature

As stated by Olesen (2000) and Pereira Gray (2017), many definitions confuse the setting with the role and the person.[15,28] However, the American Academy of Family Physicians (AAFP) clearly distinguishes between the two concepts, arguing that 'The terms "primary care" and "family medicine" are not interchangeable'.[11]As stated on the website of the WHO Primary Health Care Performance Initiative, PHC is deeply embedded in the following main values: people's first contact, people-centred, comprehensive, continuous, coordinated, accessible (also echoed by a Canadian analysis of 25 attributes of PHC).[29,30] Worldwide, general practitioners and family physicians, referring to comprehensiveness, personal and patient-centred care and universal accessibility, provide and sometimes organise primary care in PHC settings.[31] In this terminological analysis, we also found that the two concepts (GP/FM and PHC) are related but distinct.

Implications for information science and health policy

This terminological analysis of the definitions of GP/FM and PHC may have implications on the construction of field-specific filters for bibliographic searches (for example, a GP/FM filter, a PHC filter). In the filters usually published in the literature, the two concepts tend to be mixed.[32, 33]

The present study is part of the development of a taxonomy for the organisational aspects of the activities in GP/FM, as an extension of the International Classification of Primary Care (ICPC-2)[34] for contextual profesionnal aspects.[35]

This study may facilitate a dialogue between the two organisations, which have pioneered these two concepts and are still governing them, that is WONCA for GP/FM, and WHO for PHC. These organisations could come to a better understanding of the commonalities and complementarities of their endeavors, to foster mutual collaboration.[36–38] In addition, it was observed that in both sets of definitions important aspects are missing. Environmental issues are very poorly addressed as well as ethical challenges. Those are numerous and are a core task for general practitioners (for example, ethics of information and ethics of prevention).[27] There is also a need to adapt the definitions to take into account 21st century insights and developments in Information and Communication Technology. Both organisations should collaborate to produce updated, profound and distinct definitions for both GP/FM and PHC.

Funding

This study was not funded.

Acknowledgements

We are grateful to Dr. Marie-Dominique Beaulieu MSc, MD, CFPF, FCFP, at the Department of Family and Emergency Medicine, University of Montre´al, for her wise suggestions, and to Mr. Pierre Chevalier BA, at the University of Louvain, Belgium, for his kind comments on Aristotle category translation from Latin and Greek.

Provenence

Freely submitted; externally peer reviewed.

References

1. Royal College of General Practitioners. *A blueprint for building the new deal for general practice in England*. London: RCGP, 2015.
2. Bentzen N. *WONCA Dictionary of General/Family Practice*. Copenhagen: Maanedsskr, 2003.
3. Allen J, Heyrman J, Gay B, *et al*. *Being a general practitioner (Royal College of General Practitioners curriculum statement 1)*. London: Royal College of General Practitioners, 2007.
4. Howie JG, Heaney D, Maxwell M. Quality, core values and the general practice consultation: issues of definition, measurement and delivery. *Fam Pract* 2004; **21(4):** 458–468. doi: 10.1093/fampra/cmh419
5. Declaration of Alma-Ata. International Conference on Primary Health Care, Alma-Ata, USSR. 6–12 September 1978.
6. World Health Organization. Global Index Medicus. http://www.globalhealthlibrary.net/php/index.php (accessed 19 Jul 2017)
7. Aristotle. The Categories, 4, 1b25–27. In: Barnes J (ed). *The complete works of Aristotle: the revised Oxford translation*. Bollingen Series LXXI-2. 2. Princeton, NJ: Princeton University Press, 1984.
8. Munn K, Smith BJ. *Applied ontology. An introduction*. Frankfurt: Ontos Verlag, 2008.
9. Allen J, Gay B, Crebolder H, *et al* The European definitions of the key features of the discipline of general practice: the role of the GP and core competencies. *Br J Gen Pract* 2002; **52(479):** 526–527.
10. Institute of Medicine. *Primary care: America's health in a new era*. Washington, DC: The National Academies Press, 1996.
11. American Academy of Family Physicians. Primary care (definitions). http://www.aafp.org/about/policies/all/primary-care.html (accessed 20 Jul 2017).
12. Macinko J, Montenegro H, Nebot C. *Renewing primary health care in the Americas: a position paper of the Pan American Health Organization/World Health Organization (PAHO/WHO)*. Washington, DC: Pan American Health Organization, 2007.
13. American Academy of Family Physicians. Definition of family medicine. http://www.aafp.org/about/policies/all/family-medicine-definition.html (accessed 20 Jul 2017)
14. Expert Panel on Effective Ways of Investing in Health (EXPH). *Definition of a frame of reference in relation to primary care with a special emphasis on financing systems and referral systems*. Brussels: European Union, 2014.
15. Olesen F, Dickinson J, Hjortdahl P. General practice — time for a new definition. *BMJ* 2000; **320(7231):** 354–357. doi: 10.1136/bmj.320.7231.354
16. Dmytraczenko T, Couttolenc B. *Brazil's primary care strategy*. Washington DC: The World Bank, 2013 .1–40
17. World Health Organization. Health systems strengthening glossary. http://www.who.int/healthsystems/hss_glossary/en/index8.html (accessed 20 Jul 2017).
18. Commission of the Council of Wonca Europe. *The European Definition of General Practice/Family Medicine* Bangkok. Wonca Europe, 2011.
19 Primary Health Care Research and Information Service (PHCRIS). Introduction to primary health care. http://www.phcris.org.au/guides/about_phc.php (accessed 20 Jul 2017).
20. Copolillo F, Urquiza R, Jure H. Avances de la medicina familiar y comunitaria en Iberoamerica: V cumbre de WONCA Iberoamerica CIMF y la Carta de Quito. *Archivos De Medicina Familiar Y General* 2014; **11(1):** 33–38.
21. Fédération des Maisons Médicales et des Collectifs de Santé Francophone. *Integrated Health Centres*. Brussels: FFMCSF, 2016.
22. Phillips RL, Brundgardt S, Lesko SE, *et al*. The future role of the family physician in the United States: a rigorous exercise in definition. *Ann Fam Med* 2014; **12(3):** 250–255. doi: 10.1370/afm.1651
23 Agency for Healthcare Research and Quality (US Dept of Health and Human Services). Defining the primary care medical home (PCMH). https://pcmh.ahrq.gov/page/defining-pcmh (accessed 20 Jul 2017).
24. Academy of Family Physician of India. The concept of family medicine. *2nd National Conference of Family Medicine and Primary Care,* 2015, New Delhi.
25 Fédération Française des Maisons et Pôles de Santé. Qu'est ce qu'une Maison et un Pôle de Santé?. http://www.ffmps.fr/index.php/definitions/quest-ce-quune-maison-de-sante-et-un-pole-de-sante (accessed 20 Jul 2017).
26. Sadegh-Zadeh K. Medical ontology. In: Sadegh-Zadeh K, *Handbook of analytic philosophy of medicine*. 2nd ed. Dordrecht, Netherlands: Springer, 2015.
27. Jamoulle M. Quaternary prevention, an answer of family doctors to overmedicalization. *Int J Health Policy Manag* 2015; **4(2):** 61–64. doi: 10.15171/ijhpm.2015.24

28. Pereira Gray D. Towards research-based learning outcomes for general practice in medical schools: Inaugural Barbara Starfield Memorial Lecture. *BJGP Open* 2017. doi: 10.3399/bjgpopen17X100569
29. Haggerty J, Burge F, Lévesque JF, *et al*. Operational definitions of attributes of primary health care: consensus among Canadian experts. *Ann Fam Med* 2007; **5(4):** 336–344. doi: 10.1370/afm.682
30. Primary Health Care Performance Initiative. Why PHC?. http://phcperformanceinitiative.org/about-us/why-phc (accessed 20 Jul 2017).
31. Lakhani M, Baker M, Field S. *Future direction of general practice: A roadmap*. London: Royal College of General Practitioners, 2007.
32. Gill PJ, Roberts NW, Wang KY, *et al*. Development of a search filter for identifying studies completed in primary care. *Fam Pract* 2014; **31(6):** 739–745. doi: 10.1093/fampra/cmu066
33. Pols DH, Bramer WM, Bindels PJ, *et al*. Development and validation of search filters to identify articles on family medicine in online medical databases. *Ann Fam Med* 2015; **13(4):** 364–366. doi: 10.1370/afm.1780
34. Okkes I, Jamoulle M, Lamberts H, *et al*. ICPC-2-E: the electronic version of ICPC-2. Differences from the printed version and the consequences. *Fam Pract* 2000; **17(2):** 101–107. doi: 10.1093/fampra/17.2.101
35. Jamoulle M, Grosjean J, Resnick M*et al*. A terminology in general practice/family medicine to represent non-clinical aspects for various usages: the Q-Codes. In: Randell R, Cornet R, McCowan C *Informatics for health: connected citizen-led wellness and population health*. Amsterdam: IOS Press BV, 2017.471–475
36. De Maeseneer J, van Weel C, Egilman D, *et al*. Strengthening primary care: addressing the disparity between vertical and horizontal investment. *Br J Gen Pract* 2008; **58(546):** 3–4. doi: 10.3399/bjgp08X263721
37. Starfield B, Shi L, Macinko J. Contribution of primary care to health systems and health. *Milbank Q* 2005; **83 (3):** 457–502. doi: 10.1111/j.1468-0009.2005.00409.x
38. Meads G, Ashcroft J, Barr H *et al*. The case for interprofessional collaboration. In: *Health and Social Care*. Oxford: Blackwell Publishing, 2005.

Chapter 8

A terminology in General Practice / Family Medicine to represent non-clinical aspects for various usages: the Q-Codes

Citation; Jamoulle M, Grosjean J, Resnick M, Ittoo A, Treuherz A, Vander Stichele RH, et al. A terminology in General Practice / Family Medicine to represent non-clinical aspects for various usages : the Q-Codes. Medical Informatics Europe (MIE2017) Informatics for Health 2017 /. 2017; (April):1–5.(reproduced in full hereunder)
http://orbi.ulg.ac.be/handle/2268/206527

A terminology in General Practice / Family Medicine to represent non-clinical aspects for various usages: the Q-Codes

Marc JAMOULLE(a), Julien GROSJEAN(b), Melissa RESNICK(c), Ashwin ITTOO(d), Arthur TREUHERZ(e), Robert VANDER STICHELE(f), Elena CARDILLO(g), Stéfan J. DARMONI(b), Frank S. SHAMENEK(h) and Marc VANMEERBEEK(a)
a Department of General Practice, University of Liège, Belgium
b Département d'Information et d'Informatique Médicale, University of Rouen, France
c University of Texas, Health Science Center at Houston, TX USA
d HEC Management School, University of Liège, Belgium
e Department of Health Sciences Terminology, BIREME/PAHO/WHO, Sao Paulo, Brazil
f Heymans Institute of Pharmacology, University of Ghent, Belgium
g Institute of Informatics and Telematics, Rende, Italy
h Consultant, New York, NY USA

Abstract

The hereby proposed terminology called "Q-Codes" can be defined as an extension of the International Classification of Primary Care (ICPC-2). It deals with non-clinical concepts that are relevant in General Practice/Family Medicine (GP/FM). This terminology is a good way to put an emphasis on underestimated topics such as Teaching, Patient issues or Ethics. It aims at indexing GP/FM documents such as congress abstracts and theses to get a more comprehensive view about the GP/FM domain. The 182 identified Q-Codes have been very precisely defined by a college of experts (physicians and terminologists) from twelve countries. The result is available on the Health Terminology/Ontology Portal (http://www.hetop.org/Q) and formatted in OWL-2 for further semantic considerations and will be used to index the 2016 WONCA World congress communications.

Keywords

Q-Codes, general practice, family medicine, ICPC, categorization, controlled vocabulary, qualitative analysis, abstracts

Introduction

The field of medicine is blessed with a rich array of terminologies that support structured documentation of clinical information, and storage and retrieval of research publications. Overarching resources are the International Classification of Diseases (ICD), the Medical Subject Headings (MeSH) and the SNOMED CT. For most of the specialized medical domains have been built proper nomenclatures and classifications (e.g., the Systematized Nomenclature of Pathology, SNOP for pathologists) [1].

General Practice / Family Medicine (GP/FM) is a peculiar domain characterized by a very broad scope and a large array of research methods, and encompassing both clinical and non-clinical issues. Clinical issues are those aspects pertaining to signs and symptoms, reasons for encounter, and processes and diagnoses covered by the International Classification of Primary Care (ICPC) [2].

The non-clinical content of GP/FM is not fully represented either in indexes of textbooks in GP/FM, or in specific classifications or terminologies. The existing representation suffers from a top-down approach which is not always reflecting the complexity of the discipline.

To tackle this lack of non-clinical content in vocabularies, we hereby propose a new terminology for pragmatic use in real-life situations which can be used to index specific content of GP/FM (communications, literature or event managerial content of the contact with patients). Thus this resource is complementary to the ICPC-2.

In this paper, we aimed to a) describe the methodology applied for the creation of a taxonomy, b) on this basis, to present a domain-specialized terminological resource called the "Q-Codes" to facilitate the indexing of GP/FM non-clinical content and c) show how this resource could support the information retrieval of specific bibliographic information.

By doing this, we aim to contribute defining the limits of GP/FM. This is also a preliminary work to address the future necessity of the application of semantic web technology on the GP/FM domain [3]. This approach refers to an ontology building process [4].

Methods

The first step of the methodology had the aim to design the taxonomy by identifying relevant concepts in a compiled corpus that includes GP/FM texts. We have studied the concepts identified in hundreds of communications of GPs during congresses from a bottom-up approach. The relevant concepts belong to the fields that are focusing on GP/FM activities (e.g. teaching, ethics, or environmental hazard issues).

The second step was the development of a terminological resource for each category of the resulting taxonomy. This has been formalized by defining concepts, hierarchy and mappings relationships. Several methods and tools were used to perform this step: i) Cimino's standard set of desiderata was applied to build the terminology content and structure [5], ii) we highlighted each concept by relevant bibliographic citations as well as by linking them to BabelNet , DBpedia and to other reference terminologies, iii) we relied on the HeTOP multi-lingual and multi-terminology portal to fulfill each conceptual content and to manage the translations of terms and their definitions.

The final step was to evaluate and discuss the conceptual content of each created code of the taxonomy by involving several experts from all over the world (twenty four GPs and terminology experts from ten countries).

Selection of abstracts and corpus compilation

A total of 1,702 abstracts were selected from six sources in English or French: (i) Wonca Europe 2007 (n=998), (ii) Portuguese 18th national conference of family medicine, 2013 (n=128), (iii) Congrès Confédération des Généralistes Enseignants (CNGE) Clermont-Ferrand 2013 (n=205), (iv) CNGE Lille 2014 (n=289), (v) SwissFamilyDocs Zurich 2014 (n=45), and (vi) Belgian GP/FM research congress Brussels 2014 (n=37). These six sources were selected due to the ease of accessibility of their abstracts.

Qualitative methods for development of the taxonomy

Data collection stemmed from analyzing abstract proceedings from the mentioned GP/FM conferences. These data were analyzed in a grounded theory approach. It involves construction of a hypothesis or discovery of concepts through data analysis [6].

Tools

HeTOP is the Health Terminology/Ontology Portal which provides access to 60 main health terminologies in several different languages. Due to these multilingual terminologies, HeTOP is used for many purposes. It is very useful not only for translators, terminologists and ontologists, but also for physicians coding patient records and using services on demand, e.g. info buttons. Finally, and most important for this research, HeTOP assists in indexing resources on the Internet.

We relied on HeTOP for this study to: a) find similar concepts in other vocabularies; b) create the Q-Codes terminology; c) manage each concept (labels, synonyms, definitions); d) perform manual mappings and e) give access to the final Q-Codes terminology through a web site.

The Q-Codes terminology

The ambition was to create a terminology to represent the non-clinical activities of GP/FM, by extending the 17 chapters of the clinical classification ICPC with an 18th chapter, called Q-Codes ("Q" being a letter in the alphabet not yet used in ICPC).

Each term identified in the first step has been converted into a concept (a "Q-Code"). Each concept has received a definition explaining its conceptual value. The extension of the concept, i.e. its use in several other online databases, has been documented through a careful search in a set of online dictionaries and terminologies.

For each Q-Code, a minimum of properties was fulfilled to define the concept: a) the Preferred Label, b) one or more definitions, c) synonyms and linguistic variants, d) a sample of pertinent articles to the understanding of the main subject and e) BabelNet and DBpedia unique Ids and f) relevant MeSH terms related to the Q-Codes were mapped. Those mappings can lead to ease querying bibliographic databases (e.g. PubMed) for each specific Q-Code but they are also an important way to evaluate and assess the quality of Q-Codes definitions and conceptual content.

With regards to implementation, the Q-Codes (concepts, properties and relations) were formalized and implemented in Web Ontology Language (OWL). Based on the Resource Description Framework (RDF) standard, OWL is a knowledge-representation language which is considered the de facto language for ontology implementation. This task was a rough OWL-2 export from HeTOP to Web Protégé without any description logic.

Results

The complete analysis of 1,702 French/English abstracts lead to the construction of a taxonomy composed by 182 terms. This taxonomy was enhanced to a terminological level according to Cimino's desiderata [5] and thanks to a complete support of semantic web technologies (HeTOP, Web Protégé, etc.) The resulting terminology called the "Q-Codes" consists of 182 concepts divided among 8 domains. Each Q-Code (concept) was contentiously defined and tagged with English terms and definitions. Bibliographic citations and external concept URIs have been added to ensure semantic extension and validity. Table 1 gives an overview of the main Q-Codes domains with included covered topics.

Table 1. Q-Codes domains overview Q-Code domain Label Examples of covered topics QC Patient's category age, gender issues, abuse QD Family doctor's issue communication, clinical prevention, medico legal issues QE Medical ethics bioethics, professional ethics, info ethics QH Planetary

health environmental health, biological hazards, nuclear hazards QP Patient issue patient safety, patient centeredness, quality of health care QR Research research methods, research tools, epidemiology of primary care QS Structure of practice primary care setting, primary care provider, practice relationship QT Knowledge management teaching, training, knowledge dissemination

Results of the final step are related to the translation of the labels and definitions of the Q-Codes by different general practitioners in their native languages which are: French, Spanish, Brazilian-Portuguese, Dutch, Turkish, Korean and Vietnamese. Two terminologists reviewed the translations for three of the languages. One terminologist validated the English translation, and the second terminologist validated the Spanish and Portuguese translations. The Q-Codes multi-lingual terminology is available on HeTOP at http://www.hetop.org/Q (authentication required as wicc/wiccdemo).

Discussion

To the best of our knowledge, this is the first attempt to expand the ICPC coding system with an extension for managerial issues, thus covering nonclinical content, with the intent to improve performance in information storage and retrieval for research purposes in this broad, eclectic, and underserved domain of medicine. 4.1. Implications for Practice We expect that the creation of this terminological resource for indexing abstracts and for facilitating Medline searches for general practitioners, researchers and students in medicine will reduce loss of knowledge in the domain of GP/FM. In addition, through better indexing of the grey literature (congress abstracts, master's and doctoral theses), we hope to enhance the accessibility of research results of general practitioners. 4.2. Implications for research End-users are often not very well-versed in knowledge-representation formalisms, and it remains to be proven that our proposed terminology will help them in dealing with more complex systems, such as MeSH, to support their information storage and retrieval activities. Nevertheless, the Q-Codes base is aimed at several uses: a) online repository of knowledge specific to GP/FM in several languages; b) online PubMed linked bibliographic system easy to use for training in GP/FM; c) resource for online e-learning; d) resource for the analysis of content of congresses in GP/FM in joint usage with ICPC-2 and indexation of gray literature in GP/FM; e) automatic or semi-automatic congresses indexing system; f) linking GP/FM to the web of data and the linked data initiative. Further work could be conducted to enhance the Q-Codes formalism with ontology building as the current OWL-2 version is a rough export.

Conclusion

"Q-Codes" are a terminology of non-clinical subjects in GP/FM. This work is the result of a two-year cooperative project between participants from twelve countries and eight languages. The 182 concepts have been elaborated in a bottom-up approach, by retrieving the topics most frequently addressed by GPs when they met in congresses. This work is freely available online at http://www.hetop.org/Q. The data is also available in the OWL-2 language for future use in the semantic web. The product, user guide and e-learning are available at http://3cgp.docpatient.net.

References

1. R. Cornet & N. de Keizer, Forty years of SNOMED: a literature review, BMC Medical Informatics and Decision Making 8 (2008), Suppl 1, S2.
2. I. Okkes, M. Jamoulle, H. Lamberts & N. Bentzen, ICPC-2-E: the electronic version of ICPC2. Differences from the printed version and the consequences, Family Practice 17(2) (2000), 101–107.
3. M. Jamoulle, R.H. Vander Stichele, E. Cardillo, J. Roumier, J. Grosjean & S.J. Darmoni, Semantic Web and the Future of Health Care Data in Family Practice, Merit Research Journal of Medicine and Medical Sciences 3(12) (2015), 586–594.
4. M. Fernández-López & A. Gómez-Pérez, Overview and analysis of methodologies for building ontologies, The Knowledge Engineering Review 17(02) (2002), 129–156.
5. J.J. Cimino, Desiderata for controlled medical vocabularies in the twenty-first century, Methods of Information in Medicine 37(4-5) (1998), 394–403.
6. C. Faggiolani, Perceived Identity: applying Grounded Theory in Libraries, JLIS.it (2011)

Chapter 9

Development, dissemination, and applications of a new terminological resource the Q-Code taxonomy for professional aspects of General Practice / Family Medicine.

Jamoulle, M., Resnick, M., Grosjean, J., Ittoo, A., Cardillo, E., Stichele, V., Darmoni S., Vanmeerbeek, M. (2017, October). Development , dissemination , and applications of a new terminological resource the Q-Code taxonomy for professional aspects of General Practice / Family Medicine. European Journal of General Practice, (accepted), 1–13. (abstract only)

Keywords: General practice – Primary health care - Terminology as Topic – Semantic web - Qualitative Research.

Abstract

Background:

While documentation of clinical aspects of General Practice/Family Medicine (GP/FM) is assured by the International Classification of Primary Care (ICPC), there is no taxonomy for the professional aspects (context and management) of GP/FM.

Aim:

To present the development, dissemination, applications, and resulting face validity of the Q-Codes taxonomy specifically designed to describe contextual features of GP/FM, proposed as an extension to the ICPC

Development:

The Q-Codes taxonomy was developed from Lamberts' seminal idea for indexing contextual content (1987) by a multi-disciplinary team of knowledge engineers, linguists and general practitioners, through a qualitative and iterative analysis of 1702 abstracts from six GP/FM conferences using Atlas.ti software. A total of 182 concepts, called Q-codes, representing professional aspects of GP/FM, were identified and organised in a taxonomy.

Dissemination:

The taxonomy is published as an on-line terminological resource, using semantic web techniques and Web ontology language (OWL) (www.hetop.eu/Q). Each Q-code is identified with an Unique Resource Identifier (URI), and provided with preferred terms, formal definitions in eight languages (pt, es, en, fr, nl, ko, vi, tr) and search filters for Medline and web searches.

Applications:

This taxonomy has already been used to support queries in bibliographic databases (e.g. Medline), to facilitate indexing of grey literature in GP/FM as congress abstracts, master theses, websites and as an educational tool in vocational teaching,

Conclusions:

The rapidly growing list of practical applications provides face-validity for the usefulness of this freely available new terminological resource

Chapter 10

Indexing grey multilingual literature in General Practice in the era of Semantic Web

Jamoulle, M., Resnick, M., Ittoo, A., Vander Stichele, R., Cardillo, E., Grosjean, J., Darmoni, S., & Vanmeerbeek, M. (2017). Indexing grey multilingual literature in General Practice in the era of Semantic Web. The Grey Journal (accepted)
http://hdl.handle.net/2268/210490 (abstract only)

Abstract :

Problem/Goal:

Sharing the results of research with General Practitioners (GPs) is crucial for the survival of the discipline of General Practice / Family Medicine (GP/FM). The production of abstracts in GP/FM exceeds 15,000 per year worldwide. Each abstract often represents two years of work for its authors and is expressed in local languages. Only 45% of them are published in indexed medical journals.

Usual indexation systems like MeSH are not multilingual nor adapted to the particular field of GP/FM. Consequently, these abstracts are lacking bibliographic control and more than half of the research presented by GPs at congresses is lost. Considering the absence of appropriate domain-specific terminologies or classification systems, we propose a new multilingual indexing system.

The existing International Classification of Primary Care (ICPC) is currently used for clinical purposes and has now been expanded with a taxonomy related to contextual aspects (called Q-Codes) such as education, research, practice organization, ethics or policy in GP/FM, currently not captured. The set is proposed under the name Core Content Classification in General Practice (3CGP).

The aim is to facilitate indexing of GP/FM specific scientific work and to improve performance in information storage and retrieval for research purposes in this field.

Research Method/Procedure:

Using qualitative analysis, a corpus of 1,702 abstracts from six GP/FM-related European congresses was analyzed to identify main themes discussed by GPs (as continuity, accessibility or medical ethics), handled in a domain-specific taxonomy called Q-Codes and translated in 8 languages. In addition, a methodology for building a lightweight ontology (in OWL-2) was applied to Q-Codes, adding object and datatype properties to the hierarchical relations, including mapping to the MeSH thesaurus, Babelnet (www.babelnet.org) and Dbpedia. Finally, the ICPC-2 in 19 languages and Q-Codes in 8 languages have been integrated in a healthcare terminology service (www.hetop.eu/q) with a companion website (http://3cgp.docpatient.net)

Anticipated Results of the Research:

The creation and the on-line publication of this multilingual terminological resource for indexing abstracts and for facilitating Medline searches could reduce loss of knowledge in the domain. In addition, through better indexing of the grey literature (congress abstracts, master's and doctoral thesis), we hope to enhance the accessibility of research results of GP/FM domain and promote the emergence of networks of researchers. First result of experimental implementations of the new indexing system will be presented.

Indication of costs related to the project:

This project has not been funded. 3CGP is placed under Attribution-Non-Commercial-Share-Alike 4.0 International (CC BY-NC-SA 4.0). ICPC is copyrighted by WONCA.

Appendix A

Q-Codes Copy desk 2007

Jamoulle M. 3CGP/FM Core Content Classification of GP/FM. In: WONCA Europe Conference 2007, Paris, France [Internet]. 2007. Available from: http://docpatient.net/mj/wonca2007/3CGPFMdeskcopy.pdf

3CGP/FM **Core Content Classification of GP/FM** © marc@jamoulle.com ver 2.0 Oct. 2007

Domain name / Category name / Sub-category name	Code
C Patient's categories	QC
Age groups	QC1
Infants	QC11
Children	QC12
Adolescents	QC13
Ageing	QC14
Gender issues	QC2
Men's health	QC21
Women's health	QC22
Social high risk	QC3
Ethnic subgroups	QC31
Migrants & refugees	QC32
Homeless	QC33
In jail	QC34
Addiction	QC4
legal products	QC41
street drugs	QC42
gaming	QC43
Assault	QC5
battered women	QC51
victims of abuses	QC52
torture	QC53
ritual mutilations	QC54
D Provider (Doctor) issues	QD
Communicator	QD1
Encounter management	QD11
Doctor patient relationship	QD12
Counselling	QD13
Systemic	QD14
Caregiver	QD2
Problem solving	QD21
Comprehensiveness	QD22
Health education	QD23
Clinical skills	QD24
Continuity of care	QD25
Palliative care	QD26
A & E	QD27
Care manager	QD3
Health risk management	QD31
Health issue management	QD32
Health status assessment	QD33
Outcome assessment	QD34
Genetic issues	QD35
Agent of prevention	QD4
Primairy prevention	QD41
Secondairy prevention	QD42
Tertiairy prevention	QD43
Quaternairy prevention	QD44
Complementary medicine	QD5
Medico legal issues	QD6
Professional image & identity	QD7
Health provider personal life	QD8
E Ethics	QE
Personal views	QE1
Professional ethics	QE2
Bioethics	QE3
Euthanasia	QE31
Infoethics	QE4
Confidentiality	QE41
Informed consent	QE42
H Hazards	QH
Environmental	QH1
Indoor pollution	QH11
Outdoor pollution	QH12
Biological	QH2
Nuclear	QH3
P Patient issues	QP
Diagnostic process	QP1
Availability diag. process	QP11
Safety diagnostic process	QP12
Therapeutic process	QP2
Availability of ther. proces	QP21
Over The Counter	QP22
Comfort ther. process	QP23
Safety of ther. process	QP24
Practice & health care organisation	QP3
Availability of health care	QP31
Accessibility of health care	QP32
Acceptability health care	QP33
Safety of health care org.	QP34
Participation	QP36
Patient's views	QP4
Patient demand	QP40
Patient appraisal	QP41
Patient satisfaction	QP42
Patient knowledge	QP43
Patient autonomy/depend	QP44
Patient cultural backgr.	QP45
Patient expenses	QP46
Patient health habits	QP5
Nutrition	QP51
Sexuality	QP52
Self care & hygiene	QP53
Travel	QP54
R R & D tools	QR
Science philosophy	QR1
Epidemiology	QR2
Pharmacoepidemiology	QR21
Community health	QR22
Functional status	QR3
Research methods	QR4
Qualitative study	QR41
Research network	QR42
Classification	QR5
Scales & Questionnaires	QR6
Health economy	QR7
PHC planification & organisation	QR8
S Structure of practice	QS
Infrastructure	QS1
Setting (incl. rural)	QS11
Economy of practice	QS12
Practice management	QS13
Manpower	QS14
Health Inform. Manag.	QS15
Practice equipment	QS16
Security	QS17
Relationship	QS2
Collaboration	QS21
Referral/ countereferral	QS22
Coordination of care	QS23
Transdisciplinarity	QS24
Professional bodies	QS3

T Training & Knowledge management	QT
Teaching	QT1
Teaching methods	QT11
Teaching curriculum	QT12
Teaching program	QT13
Training & knowl. Eval.	QT14
Training	QT4
Undergrad. or basic educ.	QT41
Post graduate education	QT42
Continuing medical educ.	QT43
Supervision methods	QT44
Trainers & Supervisors	QT45
Academics	QT46
Quality assurance	QT5
Theory & principles	QT50
Evidence based medicine	QT51
Guidelines	QT52
Critical reading & review	QT53
Peer review	QT54
Accreditation process	QT55
Practice assesment	QT56
Health device assesment	QT57
Editing	QT6
Printed	QT61
On line Information	QT62
Digital libraries	QT63
Reporting	QT7
Sentinel network	QT71
Drug reporting	QT72
Events reporting	QT73
O Others	QO

This authority list, adapted from an orginal work of Prof. Dr. Henk Lamberts about Q codes in 1987, has to be completed by further analysis of publications in General Practice and Family Medicine.

Careful analysis of the definitions, inclusion and exclusion criteria is now necessary in order to avoid as far as possible the heterogeneity and the overlap of the classes.

This tool is complementary to ICPC and is designed to describe the metaclinical concepts refering to GP/FM. The letter Q is not used in ICPC and has been chosen to make the link with this clinical tool.

This tool is not validated nor is it endorsed by WICC.

As such it's a proposal for a future work proposed to WICC members during the Dunedin (NZ) meeting in 2007.

Please do refer to marc@jamoulle.com for any questions.

M.J.
WICC member
Researcher, Departm.of General Practice, UCL, Brussels.

Citation

Jamoulle M, Core content Classification of GP/FM, ver 0.2, desk copy, 2007. Available on http://docpatient.net/mj/wonca2007

Appendix B

Core Content Classification in GP/FM (3CGP) copy desks

ICPC-2, ICPC-Process and Q-Codes copy desks are reproduced here in English. More languages on http://3cgp.docpatient.net

- ICPC-2 copy desk (English)
  Wonca International Classification Committee (WICC). International Classification of Primary Care – 2nd Edition (ICPC-2). Copy desk.
  http://www.kith.no/templates/kith_WebPage____1633.aspx

- ICPC-2 Process classified term list (English)
  Wonca International Classification Committee (WICC). International Classification of Primary Care Process codes–2nd Edition (beta version 2016), classified term list in A4 format recto-verso.
  http://3cgp.docpatient.net/wp-content/uploads/2017/03/process_icpc_en.pdf

- Q-Codes ver. 2.5 copy desk (English)
  Jamoulle, M., Ouvrard, P., & Schuers, M. (2016). Terminologie multilingue de médecine générale et de famille - Version française. Charleroi: Care Editions. Jumet. Belgique. 2016.
  http://3cgp.docpatient.net/wp-content/uploads/2016/11/en_Q_2.5_tabular_list.pdf

ICPC-2 – English

International Classification of Primary Care – 2nd Edition

Wonca International Classification Committee (WICC)

Process codes

-30 Medical Exam/Eval-Complete
-31 Medical Examination/Health Evaluation-Partial/Pre-op check
-32 Sensitivity Test
-33 Microbiological/Immunological Test
-34 Blood Test
-35 Urine Test
-36 Faeces Test
-37 Histological/Exfoliative Cytology
-38 Other Laboratory Test NEC
-39 Physical Function Test
-40 Diagnostic Endoscopy
-41 Diagnostic Radiology/Imaging
-42 Electrical Tracings
-43 Other Diagnostic Procedures
-44 Preventive Imunisations/Medications
-45 Observe/Educate/Advice/Diet
-46 Consult with Primary Care Provider
-47 Consultation with Specialist
-48 Clarification/Discuss Patient's RFE
-49 Other Preventive Procedures
-50 Medicat-Script/Reqst/Renew/Inject
-51 Incise/Drain/Flush/Aspirate
-52 Excise/Remove/Biopsy/Destruction/Debride
-53 Instrument/Catheter/Intubate/Dilate
-54 Repair/Fixate-Suture/Cast/Prosthetic
-55 Local Injection/Infiltration
-56 Dress/Press/Compress/Tamponade
-57 Physical Medicine/Rehabilitation
-58 Therapeutic Counselling/Listening
-59 Other Therapeutic Procedure NEC
-60 Results Tests/Procedures
-61 Results Exam/Test/Record
-62 Administrative Procedure
-63 Follow-up Encounter Unspecified
-64 Encounter Initiated by Provider
-65 Encounter Initiated third person
-66 Refer to Other Provider (EXCL. M.D.)
-67 Referral to Physician/Specialist/Clinic/Hospital
-68 Other Referrals NEC
-69 Other Reason for Encounter NEC

General and Unspecified A

A01 Pain general/multiple sites
A02 Chills
A03 Fever
A04 Weakness/tiredness general
A05 Feeling ill
A06 Fainting/syncope
A07 Coma
A08 Swelling
A09 Sweating problem
A10 Bleeding/haemorrhage NOS
A11 Chest pain NOS
A13 Concern/fear medical treatment
A16 Irritable infant
A18 Concern about appearance
A20 Euthanasia request/discussion
A21 Risk factor for malignancy
A23 Risk factor NOS
A25 Fear of death/dying
A26 Fear of cancer NOS
A27 Fear of other disease NOS
A28 Limited function/disability NOS
A29 General symptom/complaint other
A70 Tuberculosis
A71 Measles
A72 Chickenpox
A73 Malaria
A74 Rubella
A75 Infectious mononucleosis
A76 Viral exanthem other
A77 Viral disease other/NOS
A78 Infectious disease other/NOS
A79 Malignancy NOS
A80 Trauma/injury NOS
A81 Multiple trauma/injuries
A82 Secondary effect of trauma
A84 Poisoning by medical agent
A85 Adverse effect medical agent
A86 Toxic effect non-medicinal substance
A87 Complication of medical treatment
A88 Adverse effect physical factor
A89 Effect prosthetic device
A90 Congenital anomaly OS/multiple
A91 Abnormal result investigation NOS
A92 Allergy/allergic reaction NOS
A93 Premature newborn
A94 Perinatal morbidity other
A95 Perinatal mortality
A96 Death
A97 No disease
A98 Health maintenance/prevention
A99 General disease NOS

Blood, Blood Forming Organs and Immune Mechanism B

B02 Lymph gland(s) enlarged/painful
B04 Blood symptom/complaint
B25 Fear of aids/HIV
B26 Fear cancer blood/lymph
B27 Fear blood/lymph disease other
B28 Limited function/disability
B29 Sympt/complt lymph/immune other
B70 Lymphadenitis acute
B71 Lymphadenitis non-specific
B72 Hodgkin's disease/lymphoma
B73 Leukaemia
B74 Malignant neoplasm blood other
B75 Benign/unspecified neoplasm blood
B76 Ruptured spleen traumatic
B77 Injury blood/lymph/spleen other
B78 Hereditary haemolytic anaemia
B79 Congen.anom. blood/lymph other
B80 Iron deficiency anaemia
B81 Anaemia, Vitamin B12/folate def.
B82 Anaemia other/unspecified
B83 Purpura/coagulation defect
B84 Unexplained abnormal white cells
B87 Splenomegaly
B90 HIV-infection/aids
B99 Blood/lymph/spleen disease other

PROCESS CODES
SYMPTOMS/COMPLAINTS
INFECTIONS
NEOPLASMS
INJURIES
CONGENITAL ANOMALIES
OTHER DIAGNOSES

Digestive D

D01 Abdominal pain/cramps general
D02 Abdominal pain epigastric
D03 Heartburn
D04 Rectal/anal pain
D05 Perianal itching
D06 Abdominal pain localized other
D07 Dyspepsia/indigestion
D08 Flatulence/gas/belching
D09 Nausea
D10 Vomiting
D11 Diarrhoea
D12 Constipation
D13 Jaundice
D14 Haematemesis/vomiting blood
D15 Melaena
D16 Rectal bleeding
D17 Incontinence of bowel
D18 Change faeces/bowel movements
D19 Teeth/gum symptom/complaint
D20 Mouth/tongue/lip symptom/complt.
D21 Swallowing problem
D23 Hepatomegaly
D24 Abdominal mass NOS
D25 Abdominal distension
D26 Fear of cancer of digestive system
D27 Fear of digestive disease other
D28 Limited function/disability (d)
D29 Digestive symptom/complaint other
D70 Gastrointestinal infection
D71 Mumps
D72 Viral hepatitis
D73 Gastroenteritis presumed infection
D74 Malignant neoplasm stomach
D75 Malignant neoplasm colon/rectum
D76 Malignant neoplasm pancreas
D77 Malig. neoplasm digest other/NOS
D78 Neoplasm digest benign/uncertain
D79 Foreign body digestive system
D80 Injury digestive system other
D81 Congen. anomaly digestive system
D82 Teeth/gum disease
D83 Mouth/tongue/lip disease
D84 Oesophagus disease
D85 Duodenal ulcer
D86 Peptic ulcer other
D87 Stomach function disorder
D88 Appendicitis
D89 Inguinal hernia
D90 Hiatus hernia
D91 Abdominal hernia other
D92 Diverticular disease
D93 Irritable bowel syndrome
D94 Chronic enteritis/ulcerative colitis
D95 Anal fissure/perianal abscess
D96 Worms/other parasites
D97 Liver disease NOS
D98 Cholecystitis/cholelithiasis
D99 Disease digestive system, other

Eye F

F01 Eye pain
F02 Red eye
F03 Eye discharge
F04 Visual floaters/spots
F05 Visual disturbance other
F13 Eye sensation abnormal
F14 Eye movements abnormal
F15 Eye appearance abnormal
F16 Eyelid symptom/complaint
F17 Glasses symptom/complaint
F18 Contact lens symptom/complaint
F27 Fear of eye disease
F28 Limited function/disability (f)
F29 Eye symptom/complaint other
F70 Conjunctivitis infectious
F71 Conjunctivitis allergic
F72 Blepharitis/stye/chalazion
F73 Eye infection/inflammation other
F74 Neoplasm of eye/adnexa
F75 Contusion/haemorrhage eye
F76 Foreign body in eye
F79 Injury eye other
F80 Blocked lacrimal duct of infant
F81 Congenital anomaly eye other
F82 Detached retina
F83 Retinopathy
F84 Macular degeneration
F85 Corneal ulcer
F86 Trachoma
F91 Refractive error
F92 Cataract
F93 Glaucoma
F94 Blindness
F95 Strabismus
F99 Eye/adnexa disease, other

Ear H

H01 Ear pain/earache
H02 Hearing complaint
H03 Tinnitus, ringing/buzzing ear
H04 Ear discharge
H05 Bleeding ear
H13 Plugged feeling ear
H15 Concern with appearance of ears
H27 Fear of ear disease
H28 Limited function/disability ear
H29 Ear symptom/complaint other
H70 Otitis externa
H71 Acute otitis media/myringitis
H72 Serous otitis media
H73 Eustachian salpingitis
H74 Chronic otitis media
H75 Neoplasm of ear
H76 Foreign body in ear
H77 Perforation ear drum
H78 Superficial injury of ear
H79 Ear injury other
H80 Congenital anomaly of ear
H81 Excessive ear wax
H82 Vertiginous syndrome
H83 Otosclerosis
H84 Presbyacusis
H85 Acoustic trauma
H86 Deafness
H99 Ear/mastoid disease, other

Cardiovascular K

K01 Heart pain
K02 Pressure/tightness of heart
K03 Cardiovascular pain NOS
K04 Palpitations/awareness of heart
K05 Irregular heartbeat other
K06 Prominent veins
K07 Swollen ankles/oedema
K22 Risk factor cardiovascular disease
K24 Fear of heart disease
K25 Fear of hypertension
K27 Fear cardiovascular disease other
K28 Limited function/disability (k)
K29 Cardiovascular sympt./complt. other
K70 Infection of circulatory system
K71 Rheumatic fever/heart disease
K72 Neoplasm cardiovascular
K73 Congenital anomaly cardiovascular
K74 Ischaemic heart disease w. angina
K75 Acute myocardial infarction
K76 Ischaemic heart disease w/o angina
K77 Heart failure
K78 Atrial fibrillation/flutter
K79 Paroxysmal tachycardia
K80 Cardiac arrhythmia NOS
K81 Heart/arterial murmur NOS
K82 Pulmonary heart disease
K83 Heart valve disease NOS
K84 Heart disease other
K85 Elevated blood pressure
K86 Hypertension uncomplicated
K87 Hypertension complicated
K88 Postural hypotension
K89 Transient cerebral ischaemia
K90 Stroke/cerebrovascular accident
K91 Cerebrovascular disease
K92 Atherosclerosis/PVD
K93 Pulmonary embolism
K94 Phlebitis/thrombophlebitis
K95 Varicose veins of leg
K96 Haemorrhoids
K99 Cardiovascular disease other

Musculoskeletal L

L01 Neck symptom/complain
L02 Back symptom/complaint
L03 Low back symptom/complaint
L04 Chest symptom/complaint
L05 Flank/axilla symptom/complaint
L07 Jaw symptom/complaint
L08 Shoulder symptom/complaint
L09 Arm symptom/complaint
L10 Elbow symptom/complaint
L11 Wrist symptom/complaint
L12 Hand/finger symptom/complaint
L13 Hip symptom/complaint
L14 Leg/thigh symptom/complaint
L15 Knee symptom/complaint
L16 Ankle symptom/complaint
L17 Foot/toe symptom/complaint
L18 Muscle pain
L19 Muscle symptom/complaint NOS
L20 Joint symptom/complaint NOS
L26 Fear of cancer musculoskeletal
L27 Fear musculoskeletal disease other
L28 Limited function/disability (l)
L29 Sympt/complt. Musculoskeletal other
L70 Infections musculoskeletal system
L71 Malignant neoplasm musculoskeletal
L72 Fracture: radius/ulna
L73 Fracture: tibia/fibula
L74 Fracture: hand/foot bone
L75 Fracture: femur
L76 Fracture: other
L77 Sprain/strain of ankle
L78 Sprain/strain of knee
L79 Sprain/strain of joint NOS
L80 Dislocation/subluxation
L81 Injury musculoskeletal NOS
L82 Congenital anomaly musculoskeletal
L83 Neck syndrome
L84 Back syndrome w/o radiating pain
L85 Acquired deformity of spine
L86 Back syndrome with radiating pain
L87 Bursitis/tendinitis/synovitis NOS
L88 Rheumatoid/seropositive arthritis
L89 Osteoarthrosis of hip
L90 Osteoarthrosis of knee
L91 Osteoarthrosis other
L92 Shoulder syndrome
L93 Tennis elbow
L94 Osteochondrosis
L95 Osteoporosis
L96 Acute internal damage knee
L97 Neoplasm benign/unspec musculo.
L98 Acquired deformity of limb
L99 Musculoskeletal disease, other

Neurological N

N01 Headache
N03 Pain face
N04 Restless legs
N05 Tingling fingers/feet/toes
N06 Sensation disturbance other
N07 Convulsion/seizure
N08 Abnormal involuntary movements
N16 Disturbance of smell/taste
N17 Vertigo/dizziness
N18 Paralysis/weakness
N19 Speech disorder
N26 Fear cancer neurological system
N27 Fear of neurological disease other
N28 Limited function/disability (n)
N29 Neurological symptom/complt. other
N70 Poliomyelitis
N71 Meningitis/encephalitis
N72 Tetanus
N73 Neurological infection other
N74 Malignant neoplasm nervous system
N75 Benign neoplasm nervous system
N76 Neoplasm nervous system unspec.
N79 Concussion
N80 Head injury other
N81 Injury nervous system other
N85 Congenital anomaly neurological
N86 Multiple sclerosis
N87 Parkinsonism
N88 Epilepsy
N89 Migraine
N90 Cluster headache
N91 Facial paralysis/bell's palsy
N92 Trigeminal neuralgia
N93 Carpal tunnel syndrome
N94 Peripheral neuritis/neuropathy
N95 Tension headache
N99 Neurological disease, other

Psychological P

P01 Feeling anxious/nervous/tense
P02 Acute stress reaction
P03 Feeling depressed
P04 Feeling/behaving irritable/angry
P05 Senility, feeling/behaving old
P06 Sleep disturbance
P07 Sexual desire reduced
P08 Sexual fulfilment reduced
P09 Sexual preference concern
P10 Stammering/stuttering/tic
P11 Eating problem in child
P12 Bedwetting/enuresis
P13 Encopresis/bowel training problem
P15 Chronic alcohol abuse
P16 Acute alcohol abuse
P17 Tobacco abuse
P18 Medication abuse
P19 Drug abuse
P20 Memory disturbance
P22 Child behaviour symptom/complaint
P23 Adolescent behav. Symptom/complt.
P24 Specific learning problem
P25 Phase of life problem adult
P27 Fear of mental disorder
P28 Limited function/disability (p)
P29 Psychological symptom/complt other
P70 Dementia
P71 Organic psychosis other
P72 Schizophrenia
P73 Affective psychosis
P74 Anxiety disorder/anxiety state
P75 Somatization disorder
P76 Depressive disorder
P77 Suicide/suicide attempt
P78 Neuraesthenia/surmenage
P79 Phobia/compulsive disorder
P80 Personality disorder
P81 Hyperkinetic disorder
P82 Post-traumatic stress disorder
P85 Mental retardation
P86 Anorexia nervosa/bulimia
P98 Psychosis NOS/other
P99 Psychological disorders, other

Respiratory R

R01 Pain respiratory system
R02 Shortness of breath/dyspnoea
R03 Wheezing
R04 Breathing problem, other
R05 Cough
R06 Nose bleed/epistaxis
R07 Sneezing/nasal congestion
R08 Nose symptom/complaint other
R09 Sinus symptom/complaint
R21 Throat symptom/complaint
R23 Voice symptom/complaint
R24 Haemoptysis
R25 Sputum/phlegm abnormal
R26 Fear of cancer respiratory system
R27 Fear of respiratory disease, other
R28 Limited function/disability (r)
R29 Respiratory symptom/complaint oth.
R71 Whooping cough
R72 Strep throat
R73 Boil/abscess nose
R74 Upper respiratory infection acute
R75 Sinusitis acute/chronic
R76 Tonsillitis acute
R77 Laryngitis/tracheitis acute
R78 Acute bronchitis/bronchiolitis
R79 Chronic bronchitis
R80 Influenza
R81 Pneumonia
R82 Pleurisy/pleural effusion
R83 Respiratory infection other
R84 Malignant neoplasm bronchus/lung
R85 Malinant neoplasm respiratory, other
R86 Benign neoplasm respiratory
R87 Foreign body nose/larynx/bronch
R88 Injury respiratory other
R89 Congenital anomaly respiratory
R90 Hypertrophy tonsils/adenoids
R92 Neoplasm respiratory unspecified
R95 Chronic obstructive pulmonary dis
R96 Asthma
R97 Allergic rhinitis
R98 Hyperventilation syndrome
R99 Respiratory disease other

PROCESS CODES

SYMPTOMS/COMPLAINTS

INFECTIONS

NEOPLASMS

INJURIES

CONGENITAL ANOMALIES

OTHER DIAGNOSES

Skin S

S01 Pain/tenderness of skin
S02 Pruritus
S03 Warts
S04 Lump/swelling localized
S05 Lumps/swellings generalized
S06 Rash localized
S07 Rash generalized
S08 Skin colour change
S09 Infected finger/toe
S10 Boil/carbuncle
S11 Skin infection post-traumatic
S12 Insect bite/sting
S13 Animal/human bite
S14 Burn/scald
S15 Foreign body in skin
S16 Bruise/contusion
S17 Abrasion/scratch/blister
S18 Laceration/cut
S19 Skin injury other
S20 Corn/callosity
S21 Skin texture symptom/complaint
S22 Nail symptom/complaint
S23 Hair loss/baldness
S24 Hair/scalp symptom/complaint
S26 Fear of cancer of skin
S27 Fear of skin disease other
S28 Limited function/disability (s)
S29 Skin symptom/complaint other
S70 Herpes zoster
S71 Herpes simplex
S72 Scabies/other acariasis
S73 Pediculosis/skin infestation other
S74 Dermatophytosis
S75 Moniliasis/candidiasis skin
S76 Skin infection other
S77 Malignant neoplasm of skin
S78 Lipoma
S79 Neoplasm skin benign/unspecified
S80 Solar keratosis/sunburn
S81 Haemangioma/lymphangioma
S82 Naevus/mole
S83 Congenital skin anomaly other
S84 Impetigo
S85 Pilonidal cyst/fistula
S86 Dermatitis seborrhoeic
S87 Dermatitis/atopic eczema
S88 Dermatitis contact/allergic
S89 Diaper rash
S90 Pityriasis rosea
S91 Psoriasis
S92 Sweat gland disease
S93 Sebaceous cyst
S94 Ingrowing nail
S95 Molluscum contagiosum
S96 Acne
S97 Chronic ulcer skin
S98 Urticaria
S99 Skin disease, other

Endocrine/Metabolic and Nutritional T

T01 Excessive thirst
T02 Excessive appetite
T03 Loss of appetite
T04 Feeding problem of infant/child
T05 Feeding problem of adult
T07 Weight gain
T08 Weight loss
T10 Growth delay
T11 Dehydration
T26 Fear of cancer of endocrine system
T27 Fear endocrine/metabolic dis other
T28 Limited function/disability (t)
T29 Endocrine/met./sympt/complt other
T70 Endocrine infection
T71 Malignant neoplasm thyroid
T72 Benign neoplasm thyroid
T73 Neoplasm endocrine oth/unspecified
T78 Thyroglossal duct/cyst
T80 Congenital anom endocrine/metab.
T81 Goitre
T82 Obesity
T83 Overweight
T85 Hyperthyroidism/thyrotoxicosis
T86 Hypothyroidism/myxoedema
T87 Hypoglycaemia
T89 Diabetes insulin dependent
T90 Diabetes non-insulin dependent
T91 Vitamin/nutritional deficiency
T92 Gout
T93 Lipid disorder
T99 Endocrine/metab/nutrit. dis. other

Urological U

U01 Dysuria/painful urination
U02 Urinary frequency/urgency
U04 Incontinence urine
U05 Urination problems other
U06 Haematuria
U07 Urine symptom/complaint other
U08 Urinary retention
U13 Bladder symptom/complaint other
U14 Kidney symptom/complaint
U26 Fear of cancer of urinary system
U27 Fear of urinary disease other
U28 Limited function/disability urinary
U29 Urinary symptom/complaint other
U70 Pyelonephritis/pyelitis
U71 Cystitis/urinary infection other
U72 Urethritis
U75 Malignant neoplasm of kidney
U76 Malignant neoplasm of bladder
U77 Malignant neoplasm urinary other
U78 Benign neoplasm urinary tract
U79 Neoplasm urinary tract NOS
U80 Injury urinary tract
U85 Congenital anomaly urinary tract
U88 Glomerulonephritis/nephrosis
U90 Orthostatic albumin./proteinuria
U95 Urinary calculus
U98 Abnormal urine test NOS
U99 Urinary disease, other

Pregnancy, Childbearing, Family Planning W

W01 Question of pregnancy
W02 Fear of pregnancy
W03 Antepartum bleeding
W05 Pregnancy vomiting/nausea
W10 Contraception postcoital
W11 Contraception oral
W12 Contraception intrauterine
W13 Sterilization
W14 Contraception other
W15 Infertility/subfertility
W17 Post-partum bleeding
W18 Post-partum symptom/complaint oth.
W19 Breast/lactation symptom/complaint
W21 Concern body image in pregnancy
W27 Fear complications of pregnancy
W28 Limited function/disability (w)
W29 Pregnancy symptom/complaint other
W70 Puerperal infection/sepsis
W71 Infection complicating pregnancy
W72 Malignant neoplasm relate to preg.
W73 Benign/unspec. neoplasm/pregnancy
W75 Injury complicating pregnancy
W76 Congenital anomaly complicate preg.
W78 Pregnancy
W79 Unwanted pregnancy
W80 Ectopic pregnancy
W81 Toxaemia of pregnancy
W82 Abortion spontaneous
W83 Abortion induced
W84 Pregnancy high risk
W85 Gestational diabetes
W90 Uncomplicate labour/delivery live
W91 Uncomplicate labour/delivery still
W92 Complicate labour/ delivery livebirth
W93 Complicate labour/delivery stillbirth
W94 Puerperal mastitis
W95 Breast disorder in pregnancy other
W96 Complications of puerperium other
W99 Disorder pregnancy/delivery, other

Female Genital X

X01 Genital pain female
X02 Menstrual pain
X03 Intermenstrual pain
X04 Painful intercourse female
X05 Menstruation absent/scanty
X06 Menstruation excessive
X07 Menstruation irregular/frequent
X08 Intermenstrual bleeding
X09 Premenstrual symptom/complaint
X10 Postponement of menstruation
X11 Menopausal symptom/complaint
X12 Postmenopausal bleeding
X13 Postcoital bleeding
X14 Vaginal discharge
X15 Vaginal symptom/complaint other
X16 Vulval symptom/complaint
X17 Pelvis symptom/complaint female
X18 Breast pain female
X19 Breast lump/mass female
X20 Nipple symptom/complaint female
X21 Breast symptom/complt. female other
X22 Concern breast appearance female
X23 Fear sexually transmitted disease (f)
X24 Fear of sexual dysfunction female
X25 Fear of genital cancer female
X26 Fear of breast cancer female
X27 Fear genital/breast disease other (f)
X28 Limited function/disability (x)
X29 Genital symptom/complt female oth.
X70 Syphilis female
X71 Gonorrhoea female
X72 Genital candidiasis female
X73 Genital trichomoniasis female
X74 Pelvic inflammatory disease
X75 Malignant neoplasm cervix
X76 Malignant neoplasm breast female
X77 Malignant neoplasm genital other (f)
X78 Fibromyoma uterus
X79 Benign neoplasm breast female
X80 Benign neoplasm female genital
X81 Genital neoplasm oth/unspecied (f)
X82 Injury genital female
X83 Congenital anomaly genital female
X84 Vaginitis/vulvitis NOS
X85 Cervical disease NOS
X86 Abnormal cervix smear
X87 Uterovaginal prolapse
X88 Fibrocystic disease breast
X89 Premenstrual tension syndrome
X90 Genital herpes female
X91 Condylomata acuminata female
X92 Chlamydia infection genital (f)
X99 Genital disease female, other

Male Genital Y

Y01 Pain in penis
Y02 Pain in testis/scrotum
Y03 Urethral discharge
Y04 Penis symptom/complaint other
Y05 Scrotum/testis sympt/complt. other
Y06 Prostate symptom/complaint
Y07 Impotence NOS
Y08 Sexual function sympt./complt.(m)
Y10 Infertility/subfertility male
Y13 Sterilization male
Y14 Family planning male other
Y16 Breast symptom/complaint male
Y24 Fear of sexual dysfunction male
Y25 Fear sexually transmitted dis. male
Y26 Fear of genital cancer male
Y27 Fear of genital disease male other
Y28 Limited function/disability (y)
Y29 Genital sympt./complt.male other
Y70 Syphilis male
Y71 Gonorrhoea male
Y72 Genital herpes male
Y73 Prostatitis/seminal vesiculitis
Y74 Orchitis/epididymitis
Y75 Balanitis
Y76 Condylomata acuminata male
Y77 Malignant neoplasm prostate
Y78 Malign neoplasm male genital other
Y79 Benign/unspec. neoplasm gen. (m)
Y80 Injury male genital
Y81 Phimosis/redundant prepuce
Y82 Hypospadias
Y83 Undescended testicle
Y84 Congenital genl anomaly (m) other
Y85 Benign prostatic hypertrophy
Y86 Hydrocoele
Y99 Genital disease male, other

Social Problems Z

Z01 Poverty/financial problem
Z02 Food/water problem
Z03 Housing/neighbourhood problem
Z04 Social cultural problem
Z05 Work problem
Z06 Unemployment problem
Z07 Education problem
Z08 Social welfare problem
Z09 Legal problem
Z10 Health care system problem
Z11 Compliance/being ill problem
Z12 Relationship problem with partner
Z13 Partner's behaviour problem
Z14 Partner illness problem
Z15 Loss/death of partner problem
Z16 Relationship problem with child
Z18 Illness problem with child
Z19 Loss/death of child problem
Z20 Relationship prob. parent/family
Z21 Behaviour problem parent/family
Z22 Illness problem parent/family
Z23 Loss/death parent/family member
Z24 Relationship problem friend
Z25 Assault/harmful event problem
Z27 Fear of a social problem
Z28 Limited function/disability (z)
Z29 Social problem NOS

Abbreviations

Anom	anomaly
behav.	behaviour
bronch.	bronchus
complicat.	complication
congen.	congenital
dis.	disease
eval.	evaluation
exam.	examination
gen.	genital
malig.	malignant
metab.	metabolic
musculo.	musculoskeletal
NEC	not elsewhere classified
NOS	not otherwise specified
nutrit.	nutrition
oth	other
preg.	pregnancy
prob.	problem
RFE	reason for encounter
sympt.	symptom
unspec.	unspecified
w	with
w/o	without

ICPC-2 Process 2016
Version beta. 1 March 1, 2017
Classification

30	medical examination / health evaluation complete
31	medical examination / health evaluation partial
32	sensitivity test
33	microbiological/ immunological test
34	blood test
35	urine test
36	faeces test
37	histological/exfoliative cytology
38	other laboratory test nec
39	physical function test
40	diagnostic endoscopy
41	diagnostic radiology/imaging
42	electrical tracing
43	other diagnostic procedure
44	preventive immunization/medication
45	observation / health education / advice / diet
46	consultation with primary care provider
47	consultation with specialist
48	clarification / discussion of patient's RFE / demand
49	other preventive procedure
50	medication / prescription / renewal / injection
51	Incision / drainage / flushing / aspiration / removal body fluid
52	excision / tissue removal / biopsy / destruction / desbridement / cauterization
53	instrumentation / catheterization / intubation / dilatation
54	repair/fixation-suture/cast/prosthetic device (apply/remove)
55	local injection / infiltration
56	dressing / pressure / compression / tamponade
57	physical medicine / rehabilitation
58	therapeutic counselling / listening
59	other therapeutic procedure / minor surgery nec
60	result test / procedure
61	result examination / test / record / letter from other provider
62	administrative procedure
63	follow-up encounter unspecified
64	encounter / problem initiated by provider
65	encounter / problem initiated by other than patient / provider
66	referral to other provider / nurse / therapist / social worker
67	referral to physician / specialist / clinic / hospital
68	other referral nec
69	other process nec

nec: not elsewhere classified

ICPC-2 Process 2016
Version beta 1 March 1, 2017
Classified terms

Code	Preferred label
30.000	complete check-up
30.001	health evaluation complete
30.002	well-baby exam
30.003	school health care exam
31.000	focused examination
31.001	health evaluation partial
31.002	specific body system examination
31.003	clinical screening of single health problem
31.004	heart auscultation
31.005	blood pressure measurement
31.006	body temperature measurement
32.000	sensitivity test
32.001	skin prick-test
32.002	patch test
32.003	radioallergosorbent (RAST) test
32.004	food challenge
32.005	methacholine challenge test
32.006	Mantoux test
33.000	microbiological test
33.001	immunological test
33.002	blood culture
33.003	microbiological test of feces
33.004	microbiological test of sputum
33.005	microbiological test of throat swab
33.006	microbiological test of vaginal swab
33.007	antigen
33.008	antibody testing
33.009	testing for human immunodeficiency virus (HIV)
33.010	testing for human papilloma virus (HPV)
33.011	antibody titer
33.012	testing for Hepatitis B
34.000	blood test
34.001	blood group and type
34.002	erythrocyte sedimentation rate (ESR)
34.003	white blood cell count (WBC)
34.004	blood hormone level
34.005	blood enzyme level
34.006	blood drug level
34.007	blood alcohol level
34.008	C-reactive protein (CRP)
34.009	prothrombin time (PTT)
34.010	D-dimer
34.011	arterial blood gases
34.012	calculation of creatinine clearance
34.013	blood pregnancy test
34.014	basic metabolic panel
35.000	urine test
35.001	urine strip test
35.002	urine drug test
35.003	urine pregnancy test
35.004	urine albumin/creatinine ratio
36.000	faeces test
36.001	stool parasite test
36.002	stool ova test
36.003	stool cyst test
36.004	test for occult blood
37.000	histological-exfoliative cytology
37.001	tissue pathology-cytology
37.002	fluid pathology-cytology
37.003	biopsy pathology-cytology
37.004	puncture pathology-cytology
37.005	excision pathology-cytology
37.006	swabbing pathology-cytology
37.007	urine cytology
37.008	sputum pathology-cytology
37.009	semen analysis
38.000	other laboratory test nec
38.001	antenatal chromosome screening
38.002	chromosome test
38.003	genetic test
38.004	DNA test
38.005	paternity test
38.006	helicobacter pylori breath test
38.007	sweat test
38.008	tear drop test
39.000	physical function test
39.001	audiometry
39.002	visual field test
39.003	urinary flowmetry
39.004	spirometry
39.005	calorimetric test
39.006	vestibular function test
39.007	color vision test
39.008	visual test
39.009	tonometry
39.010	tympanometry
40.000	diagnostic endoscopy
40.001	anoscopy
40.002	arthroscopy
40.003	bronchoscopy
40.004	colonoscopy
40.005	colposcopy
40.006	gastroscopy
40.007	hysteroscopy
40.008	laparoscopy
40.009	laryngoscopy
40.010	mediastinoscopy
40.011	pharyngoscopy
40.012	rectoscopy
40.013	rhinoscopy
40.014	sigmoidoscopy
40.015	tracheoscopy
41.000	diagnostic radiology
41.001	imaging
41.002	angiography
41.003	nuclear medicine
41.004	x-rays
41.005	ultrasound
41.006	magnetic resonance imaging (MRI)
41.007	cardiac scintigraphy
41.008	isotope scanning
41.009	bone scintigraphy
41.010	thyroid scintigraphy
41.011	thermography
41.012	computed tomography (CT)
41.013	positron emission tomography (PET)
42.000	electrical tracing
42.001	electrocardiogram (ECG)
42.002	exercise electrocardiogram
42.003	electroencephalogram (EEG)
42.004	electromyogram (EMG)
42.005	electronystagmogram (ENG)
42.006	Holter monitoring
43.000	other diagnostic procedure
43.001	dermatoscopy
43.002	assessment of activities of daily living
43.003	genealogical chart
43.004	diagnostic laparotomy
43.005	ophthalmoscopy
43.006	skin photo
43.007	anxiety test
43.008	depression test
43.009	dementia test
43.010	intelligence test
44.000	preventive immunization
44.001	preventive medication
44.002	desensitization
44.003	active or passive immunization
44.004	malaria prophylaxis
44.005	prophylactic treatment with drugs
45.000	health observation
45.001	health education
45.002	health advice
45.003	diet
45.004	advice on healthy behaviour
45.005	advice regarding social problems
45.006	advice on pregnancy
45.007	advice on family planning
45.008	advice on occupational health
45.009	advice regarding the use of health services
45.010	watchful waiting
45.011	advice on prevention of health problems
46.000	consultation with primary care provider
46.001	telemedicine consultation with primary care provider
47.000	consultation with specialist
47.001	telemedicine consultation with specialist
48.000	clarification of patient's reason for encounter
48.001	discussion of patient's demand
48.002	discussion about the need of screening tests
49.000	other preventive procedure
49.001	assessment of health conditions and risks
49.002	assessment of environment conditions and risks
49.003	medication review
50.000	medication
50.001	prescription
50.002	renewal
50.003	injection
50.004	administration of medication
50.005	injection of medication with systemic effect
50.006	prescribing of medication
50.007	renewal of medication
51.000	Incision
51.001	drainage
51.002	flushing
51.003	aspiration of body fluid
51.004	aspiration of a hematoma
51.005	incision of an abscess
51.006	irrigation of an eye
51.007	irrigation of an ear
51.008	paracentesis
51.009	ascitic fluid aspiration
51.010	lumbar puncture
51.011	puncture

51.012	aspiration of sinus
51.013	aspiration of a ganglion
51.014	aspiration of a cyst
51.015	aspiration of a bursa
51.016	aspiration of a joint
51.017	aspiration of the stomach
51.018	aspiration of the lungs
51.019	aspiration of the bladder
51.020	venesection
51.021	thoracentesis
51.022	tympanostomy
51.023	amniocentesis
52.000	excision
52.001	biopsy
52.002	tissue destruction
52.003	tissue removal
52.004	cauterisation
52.005	amputation
52.006	maggot therapy
52.007	biopsy of skin
52.008	biopsy of an organ
52.009	chemical cauterisation
52.010	electric cauterisation
52.011	cold cauterisation
52.012	laser cauterisation
52.013	curettage of the skin
52.014	uterine curettage
52.015	autolytic debridement
52.016	chemical debridement
52.017	surgical debridement
52.018	mechanical debridement
52.019	nerve destruction
52.020	vein destruction
52.021	embolectomy
52.022	vein stripping
52.023	excision, removal of tissue
52.024	excision, removal of nail
52.025	tooth extraction
52.026	sterilization
53.000	catheterization
53.001	intubation
53.002	dilatation
53.003	instrumentation
53.004	intravenous cannulation
53.005	catheterization
53.006	urethral dilatation
53.007	uterine cervical dilatation
53.008	lacrimal dilatation
53.009	tympanostomy tube insertion or removal
53.010	endotracheal intubation
53.011	enema
53.012	artificial rupture of the amniotic membranes
53.013	tracheotomy
54.000	repair (apply/remove)
54.001	fixation (apply/remove)
54.002	suture (apply/remove)
54.003	cast (apply/remove)
54.004	prosthetic device (apply/remove)
54.005	hearing aids and device
54.006	technical aids and device
54.007	visual aids and device
54.008	brace
54.009	sling
54.010	strip-plaster
54.011	taping
54.012	surgical glue suture
54.013	hormonal implant
54.014	dental implantation
54.015	insertion of vaginal pessary
54.016	orthosis
54.017	stent
54.018	Intra-uterine device (IUD)
54.019	orthopedic prosthesis
54.020	applying hernia truss
54.021	taping for sprains
54.022	treatment of fracture
54.023	reduction of luxation
55.000	local injection
55.001	local infiltration
55.002	nerve block
55.003	joint injection
55.004	bursa injection
55.005	tendon sheath injection
55.006	skin injection
55.007	subcutaneous tissue injection
55.008	epidural injection
55.009	sclerosing injection for varices
56.000	dressing
56.001	pressure
56.002	compression
56.003	tamponade
56.004	support bandage
56.005	wound bandage
56.006	application of an eye patch
57.000	rehabilitation
57.001	chiropractic
57.002	fitness training
57.003	muscle strengthening
57.004	ergotherapy
57.005	muscle exercise
57.006	physical medicine
57.007	podiatry
57.008	physiotherapy
57.009	specific disability rehabilitation
57.010	nerve stimulator
57.011	electronic muscle stimulator
57.012	speech therapy
57.013	use of heat for therapeutic purposes
57.014	use of ultrasound for therapeutic purposes
57.015	use of electrical current for therapeutic purposes
57.016	use of cold for therapeutic purposes
58.000	therapeutic counselling
58.001	therapeutic listening
58.002	psychotherapy
58.003	therapeutic counseling for a specific disease
58.004	motivational interview
58.005	supportive psychotherapy
58.006	psychosocial alcohol rehabilitation
58.007	psychosocial drug rehabilitation
59.000	other therapeutic procedure
59.001	minor surgery nec
59.002	delivery
59.003	acupuncture
59.004	caesarean section
59.005	plastic surgery
59.006	dialysis
59.007	removal of a foreign body
59.008	oxygen therapy
59.009	cardiopulmonary resuscitation
59.010	blood transfusion
60.000	test result
60.001	result of procedure
61.000	result examination from another provider
61.001	test from another provider
61.002	record from another provider
61.003	letter from another provider
62.000	administrative procedure
62.001	sick leave document
62.002	death certificate
62.003	certificate for driver's license
62.004	billing issues
62.005	health questionnaire
62.006	health record issues
62.007	filling out documents
62.008	filling out forms
63.000	follow-up encounter
64.000	problem initiated by provider
64.001	encounter initiated by provider
65.000	problem initiated by third party
65.001	encounter initiated by third party
65.002	referral by family
65.003	referral by nurse
65.004	referral by social worker
65.005	referral by psychologist
66.000	referral to a provider (not doctor)
66.001	referral to a nurse
66.002	referral to a physiotherapist
66.003	referral to an occupational therapist
66.004	referral to a chiropractor
66.005	referral to a social worker
66.006	referral to an optician
66.007	referral to a midwive
66.008	referral to a dietician
66.009	referral to a dentist
66.010	referral to a home health worker
66.011	referral to an orthodontist
66.012	referral to a psychologist
67.000	referral to physician
67.001	referral to a specialist
67.003	referral to a clinic
67.004	referral to hospital
68.000	referral to patient associations
68.001	referral to unemployment service
68.002	referral to hospice
68.003	referral to a nursing home
68.004	referral to a non-medical service
68.005	referral to an institution for rehabilitation
69.000	other process nec

Process group of the Wonca International Classification Committee http://ph3c.org

Chair :
Marten Kwist, Finlande

Adaptation :
Melissa P. Resnick (USA) &
Marc Jamoulle, Belgium

Q-Codes tabular list Version 2.5

November 2016

We reproduced the list of titles of each section of the Q-codes in a tabular list comparable to that produced with the ICPC-2 in A4 format and known as "desk copy". Both tabular lists are used to have on hand for an easy coding process of literature in General Practice Family Medicine.
Each Q-Code has a definition. Please refer to Jamoulle M, Resnick MP. General Practice / Family Medicine Multilingual Terminology.Charleroi: Care Editions; 2016. 62 p
Consider http://3cgp.docpatient.net for further informations, user guide and other languages.

QC PATIENT'S CATEGORY

- **QC1 age group**
  - QC11 infant
  - QC12 child
  - QC13 adolescent
  - QC14 adult
  - QC15 elderly
- **QC2 gender issue**
  - QC21 men's health
  - QC22 women's health
  - QC23 sex difference
  - QC24 transgender
- **QC3 social high risk**
  - QC31 ethnic subgroup
  - QC32 refugee
  - QC33 homeless
  - QC34 prisoner
- **QC4 addict**
  - QC41 prescribed drug addict
  - QC42 street drug addict
  - QC43 game addict
- **QC5 victim of violence**
  - QC51 gender-based violence
  - QC52 child abuse
  - QC53 elder abuse
  - QC54 victim of torture
  - QC55 ritual mutilation
- **QC6 survivor**

QD DOCTOR'S ISSUE

- **QD1 communicator**
  - QD11 encounter
  - QD12 doctor-patient relationship
  - QD13 counselling
  - QD14 systems thinking
  - QD15 motivational interviewing
- **QD2 doctor as carer**
  - QD21 problem solving
  - QD22 comprehensiveness
  - QD23 health education
  - QD24 clinical competence
  - QD25 continuity of care
  - QD26 palliative care
  - QD27 family planning
- **QD3 care manager**
  - QD31 health risk management
  - QD32 health issue management
    - QD321 medically unexplained symptom
    - QD322 multimorbidity
    - QD323 shared decision making
    - QD324 incidentaloma
    - QD325 prescribing behaviour
    - QD326 deadoption
  - QD33 health status assessment
  - QD34 genetic issue
- **QD4 clinical prevention**
  - QD41 primary prevention
  - QD42 secondary prevention
  - QD43 tertiary prevention
  - QD44 quaternary prevention
    - QD441 overmedicalisation
    - QD442 disease mongering
    - QD443 overinformation
    - QD444 overscreening
    - QD445 overdiagnosis
    - QD446 overtreatment
- **QD5 complementary medicine**
  - QD51 homeopathy
- **QD6 medico legal issue**
- **QD7 professional image**
- **QD8 work–life balance**

QE MEDICAL ETHICS

- **QE1 personal view**
- **QE2 professional ethics**
- **QE3 bioethics**
  - QE31 euthanasia
- **QE4 infoethics**
  - QE41 confidentiality
  - QE42 informed consent

QH PLANETARY HEALTH

- **QH1 environmental health**
  - QH11 indoor pollution
  - QH12 outdoor pollution
- **QH2 biological hazard**
- **QH3 nuclear hazard**

QO OTHER

- **QO1 Unable to code, unclear**
- **QO2 Acronym**
- **QO3 Out of scope of family medicine**
- **QO4 Consider new code**

QP PATIENT ISSUE

- **QP1 patient safety**
- **QP2 patient-centredness**
  - QP21 accessibility
  - QP22 temporal accessibility
  - QP23 cultural competency
  - QP24 affordability
  - QP25 acceptability
- **QP3 quality of health care**
- **QP4 patient perspective**
  - QP41 patient satisfaction
  - QP42 patient knowledge
  - QP43 patient autonomy
  - QP44 patient culture
  - QP45 patient expense

QP5 health behaviour
- QP51 patient alimentation
- QP52 patient's sexuality
- QP53 self-care
- QP54 over the counter

QP6 patient participation
- QP61 social network

QP7 patient advocacy

QR RESEARCH

QR1 science philosophy

QR2 epidemiology of primary care
- QR21 pharmacoepidemiology
- QR22 community-based study

QR3 research method
- QR31 qualitative study
- QR32 quantitative study
  - QR321 case-control study
  - QR322 cohort study
  - QR323 cross-sectional study
  - QR324 longitudinal study
  - QR325 intervention study
- QR33 mixed study
- QR34 validation study
- QR35 action research
- QR36 case report

QR4 research network

QR5 research tools
- QR51 taxonomy
- QR52 scale
- QR53 functional status
- QR54 observation material

QR6 expert advice

QR7 economics, primary health care

QS STRUCTURE OF PRACTICE

QS1 primary care setting
- QS11 management of practice
- QS12 economy of practice
- QS13 health information management
- QS14 practice equipment

QS2 out-of-hours

QS3 practice relationship
- QS31 practice collaboration
- QS32 referral
- QS33 coordination of care

QS4 primary care provider
- QS41 family doctor
- QS42 nurse practitioner
- QS43 midwife
- QS44 allied health professional
  - QS441 physiotherapist
  - QS442 social worker
  - QS443 psychologist
- QS45 family caregiver
- QS46 pharmacist

QT KNOWLEDGE MANAGEMENT

QT1 teaching
- QT11 pedagogic method
- QT12 teaching organization
- QT13 teaching evaluation
  - QT131 simulated patient
- QT14 academic organization

QT2 training
- QT21 undergraduate
- QT22 vocational training
- QT23 continuous medical education
- QT24 Balint group
- QT25 trainer

QT3 quality assurance
- QT31 evidence based medicine
- QT32 guideline
- QT33 critical reading
- QT34 peer review
- QT35 accreditation
- QT36 quality indicator

QT4 knowledge translation
- QT41 publishing
- QT42 online knowledge-sharing
- QT43 digital library
- QT44 electronic communication

QT5 health data reporting
- QT51 sentinel network
- QT52 adverse event
- QT53 health database

This taxonomy is complementary to ICPC. It is intended to retrieve non-clinical concepts in GP/FM
Citation : Jamoulle M, Q-Code, version 2.5, tabular list, copy desk. 2016. http://3CGP.docpatient.net

Appendix C

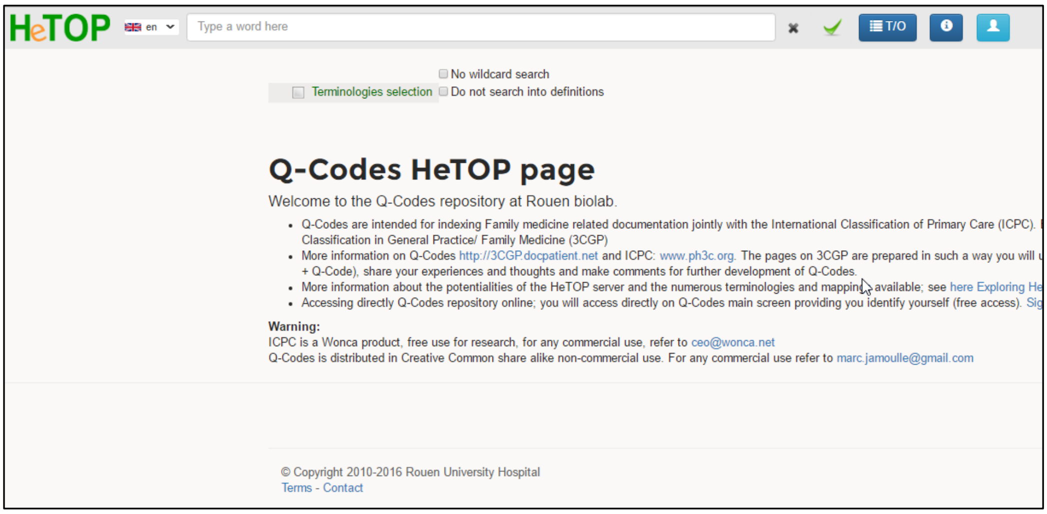
Pages about Q-Codes and 3CGP on Internet

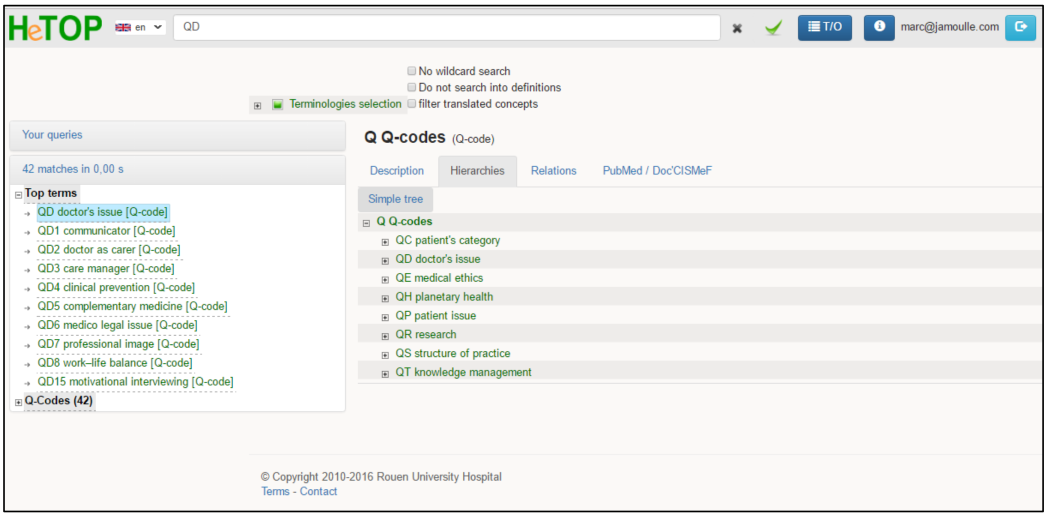
Q-Codes knowledge base on Health Terminology Portal at Rouen University (HeTOP) http://www.hetop.eu/Q

Core Content Classification web pages on http://3cgp.woncaeurope.org

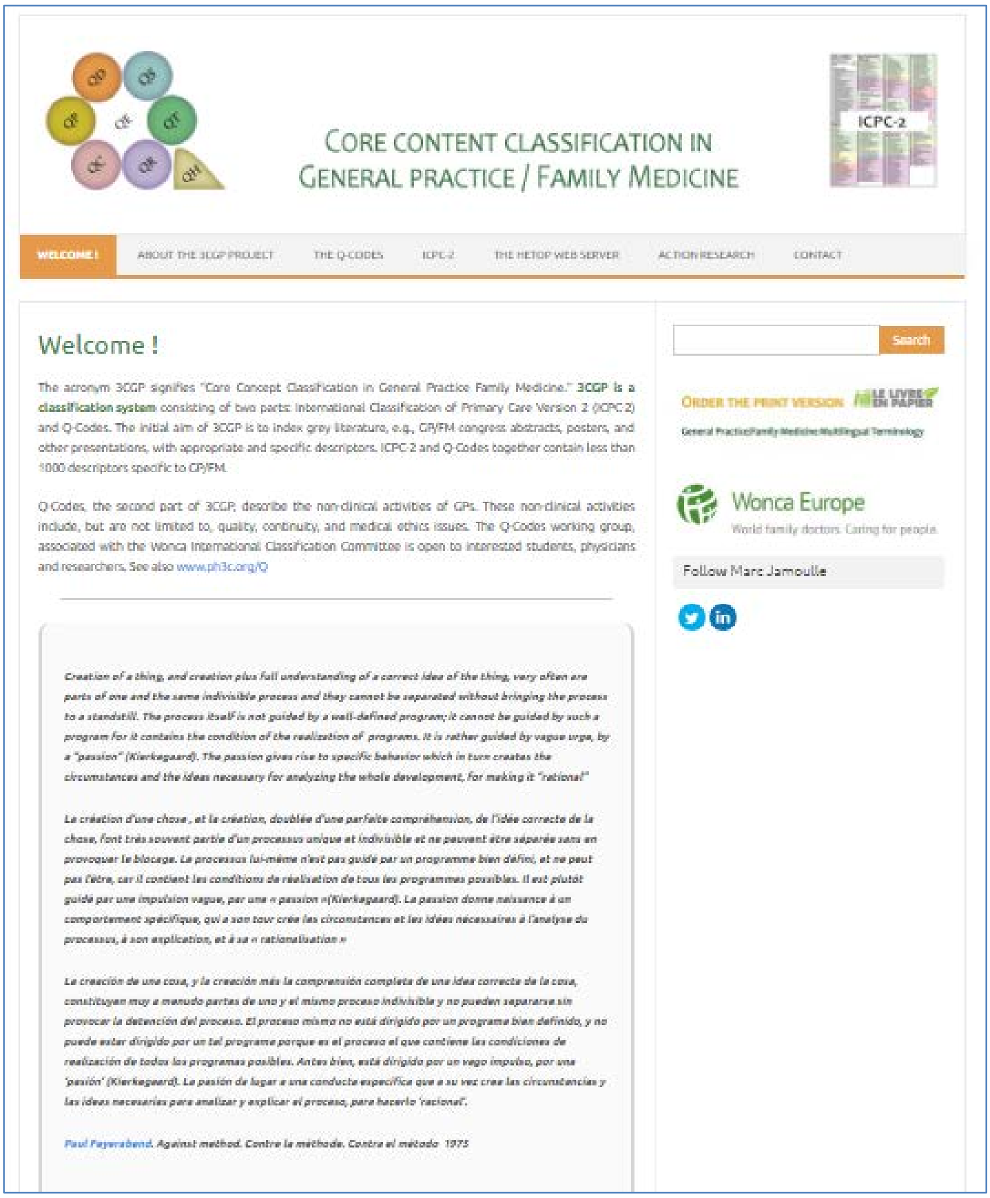
Q-Codes pages on the web site of the Wonca International Classification Committee (WICC). http://www.ph3c.org/Q

The Q-Codes HeTOP page on http://www.hetop.eu/Q





The Core Content Classification web page on http://3cgp.docpatient.net



The Q-Codes rubrics on the Wonca International Classification Committee web site www.ph3c.org

Appendix D

Study of the publication of EGPRN 2010

In 2010, Carsten Kruschinski, Maaike Lange, Christos Lionis, Chris van Weel and Eva Hummers-Pradier on behalf of the European General Practice Research Network (EGPRN) have studied the content of EGPRN conferences abstracts under the title *Themes and methods of research presented at European General Practice Research Network conferences*.

The author of the EGPRN 2010 study, referring to 2007 work, stated that ; "*The EGPRN study was performed as a part of the project on developing a Wonca/EGPRN research agenda. It analyzed the research themes covered within recent EGPRN conferences with a focus on study design and methods that were used. Abstract of past EGPRN conferences were classified on the basis of content and methodology by content analysis. This study revealed a broad range of research themes that had been addressed*." (Kruschinski, Lange et al. 2010).

Quoting the Wonca 2007 e-archive proposal (Jamoulle and Dekeuster 2007), they stated also that their result was similar. Through the careful analysis of this work, the research domain (QR) has been revised. The proposals of the EGPRN study authors to classify the abstracts presented during EGPRN conference have been numbered. It has been possible to identify in this publication 57 entries of which 22 classed in categories and subcategories in the research domain (see also fig 1.38, page 65).

The results of the study are published below.

Study of the publication of EGPRN 2010 M. Jamoulle 2016

Themes and methods of research presented at European General Practice Research Network conferences
Carsten Kruschinski, Maaike Lange, Christos Lionis, Chris van Weel, Eva Hummers-Pradier and EGPRN
Family Practice 2010; 0; 1-9; doi:10.1093/fampra/cmq023

The publication has been read carefully and proposals of the authors to classify the abstracts presented during EGPRN conference have been carefully numbered (left column of the table)

Corresponding Q-Codes have been searched for in the Q-Codes version 2.2.

EGPRN 2010	Q-Codes ver 2.3	Rem
1. Guidelines:	QT32	guidelines
1.1. Adherence	QT32	guidelines
1.2. Implementation	QT32	guidelines
1.3. development	QT32	guidelines
2. Clinical:	QD	Doctor issue
2.1. Therapy	*50	ICPC Process codes
2.2. Diagnosis	*30 to 49	ICPC Process codes
2.3. Prevention	QD4	
2.4. Risk factors	QD31	
2.5. Disease	ICPC	ICPC rubrics
2.6. prognosis	QD33	
3. Epidemiology:	QR2	
3.1. epidemiological	QR2	
4. GP/Health service (HS):	QS	
4.1. Organization	QS11	
4.2. Quality of care	QP3	
4.3. Undergraduate	QT21	
4.4. Attitudes, opinion	QD28	
4.5. Prescribing behavior	QD32	included in health issue management
4.6. Performance	QT13	included in competence
4.7. Vocational ; CME	QT22 ; QT23	Vocational training Continuous med ed.
4.8. Doctor patient relationships	QD12	
4.9. Hygiene	QP53	
4.10.Medical errors	QD31	included in health risk assessment , consider QD44 quaternary prevention
4.11.Health care utilization	QS11	
4.12.Health care access	QP21	as the patient is at the core of the system, accessibility and quality are seen from patient side
5. Patient:	QP	Patient issue
5.1. Attitudes	QP4	
5.2. Sociodemographic characteristics	QC	Category of patients
5.3. Compliance	Z11	compliance is included in ICPC
5.4. Quality of life	QP4	
5.5. Participation	QP6	
5.6. Education	QP42	
5.7. Lifestyle/health behavior	QP53	
6. Research:	QR	Research issue
6.1. Electronic patient records	QS13	
6.2. Primary care	QS1	
6.3. Evidence based medicine	QT31	
6.4. Research methods	QR3	
6.4.1. Report for example the develope-ment of a guideline	QT32	Guideline is in the Q-Codes
6.4.2. Original study	_	this approach not considered in Q-Codes and could hardly be considered as it is a temporal judgment
6.4.2.1. Qualitative	QR31	
6.4.2.2. Quantitative	QR32	and subcategories of QR32
6.4.2.2.1. Intervention study	QR325	
6.4.2.2.1.1. wih randomization	QR325	Included in QR325 this approach was not considered in Q-Codes ver2.2, included in entry term on HeTOP
6.4.2.2.1.2. without randomization	QR325	Included in QR325 this approach was not considered in Q-Codes ver2.2, included in entry term on HeTOP
6.4.2.2.2. Observational study.	QR321 to QR323	Observational is Implicit in the four following this approach was not considered in Q-Codes ver2.2 included in entry term on HeTOP
6.4.2.2.2.1. cross-sectional surveys	QR323	
6.4.2.2.2.2. case–control studies	QR321	this approach was not considered in Q-Codes ver2.2
6.4.2.2.2.3. cohort studies (both prospective and retrospective)	QR322	
6.4.2.2.2.4. longitudinal (prospective data collection without control group)	QR324	
6.4.2.2.2.5. other, for example case reports	QR36	Assuming other means other quantitative. Only case report taken in account
6.4.2.3. Qualitative and quantitative	QR33	Qualitative and quantitative = Mixed study
6.4.2.4. Instrumental research. validation and reliability questionnaires	QR52	QR52 includes scale and questionnaire
6.4.3. Systematic review / Meta-analysis	QT33	Systematic review included in QT33
6.4.4. Other. lacking information on the exact study design.	QO	
7. Remaining:	QO	

Appendix E

MeSH and search strategies in GP/FM and PHC

PHC search strategy We have gone through MeSH interface to retrieve eight MeSH which could be considered as inclusions of the concept of PHC. We have carefully excluded all reference to hospital based care, concentrating on domiciliary and locally accessible health facilities and settings. Applying [mh:noexp] allows to remain with major headings.

GP/FM search strategy Seven MeSH which could be considered as inclusions of the concept of GP/FM are studied and presented below. There are either used alone either with entry terms as text word (TW) and the number of citations retrieved through Inforoute facilities on HeTOP.

Appendix E: MeSH and search strategies in GP/FM and PHC

1. Analyzing Primary Health Care (PHC)

1.1. PHC related MeSH

Acronyms; **PHC**: Primary Health Care **GP/FM** : general practice / Family Medicine **MeSH** or **MH** ; Medical Sujcet Headings / **TW** : Text word / **mh:noexp** : tag for not exploding the main descriptor

Considering the various definitions gathered about PHC, we will retain those MeSH which are dealing with the institutional role of this particular organization of care. We have gone through MeSH interface to retrieve 8 MeSH which could be considered as inclusions of the concept of PHC. As for Dec 04, 2016, this search strategy yields 456, 822 results if MeSH are taken full. We have carefully excluded all reference to hospital based care, concentrating on domiciliary and locally accessible health facilities and settings. Applying [mh:noexp] allow to remain with major headings.

Table 1 MeSH , definitions, entry terms, year of inclusion of the eight MeSH to be considered in PHC search strategy and consideration for no exploding

MeSH	Definition	MeSH terms found below in the MeSH hierarchy.	Entry terms of the chosen MeSH	Year introduced:	Results on June 22, 2016,	Consider use in PHC
Community health centers	Facilities which administer the delivery of health care services to people living in a community or neighborhood	Substance Abuse Treatment Centers	Center, Community Health Centers, Community Health Community Health Center Health Center, Community Health Centers, Community Satellite Centers Center, Satellite Centers, Satellite Satellite Center Neighborhood Health Centers Center, Neighborhood Health Centers, Neighborhood Health Health Center, Neighborhood Health Centers, Neighborhood Neighborhood Health Center	1979	10.980	[mh:noexp] Below term includes substance treatment centers[MH]
Community Mental Health Services	Diagnostic, therapeutic and preventive mental health services provided for individuals in the community.	none	Services, Community Mental Health Mental Health Services, Community Health Services, Community Mental Services, Mental Health Community Assertive Community Treatment Community Treatment, Assertive Treatment, Assertive Community	1967	17.142	full

Home Care Agencies	Public or private organizations that provide, either directly or through arrangements with other organizations, home health services in the patient's home. (Hospital Administration Terminology, 2d ed)	none	Agencies, Home Care Agency, Home Care Care Agencies, Home Care Agency, Home Home Care Agency Home Health Care Agencies Home Health Agencies Agencies, Home Health Agency, Home Health Home Health Agency	1995	1.257	full
Community Mental Health Centers	Facilities which administer the delivery of psychologic and psychiatric services to people living in a neighborhood or community	Child Guidance Clinics	CMHC	1979	3881	[mh:noexp] Below term include Child Guidance Clinics
Home Care Services	Community health and NURSING SERVICES providing coordinated multiple services to the patient at the patient's homes. These **home**-care services are provided by a visiting nurse, **home** health agencies, HOSPITALS, or organized community groups using professional staff for care delivery. It differs from **HOME** NURSING which is provided by non-professionals.	Hemodialysis, Home Home Care Services, Hospital-Based Home Health Nursing Home Infusion Therapy Home Nursing Respite Care Homemaker Services Parenteral Nutrition, Home Parenteral Nutrition, Home Total	Home Care Service Service, Home Care Care Services, Home Domiciliary Care Care, Domiciliary Services, Home Care Home Care Care, Home	1967	42.003	[mh:noexp] Below terms include hospital based techniques

Primary Health Care	Care which provides integrated, accessible health care services by clinicians who are accountable for addressing a large majority of personal health care needs, developing a sustained partnership with patients, and practicing in the context of family and community. JAMA 1995;273 3 :192	Continuity of Patient Care Patient Discharge Patient Handoff Patient Transfer Transition to Adult Care Transitional Care Patient-Centered Care Patient Navigation Refusal to Treat	Care, Primary Health Health Care, Primary Primary Care Care, Primary Primary Healthcare Healthcare, Primary	1974	114.350	[mh:noexp] Heteroclites below term
Rural Health Services	Health services, public or private, in rural areas. The services include the promotion of health and the delivery of health care.	Rural Nursing	Health Services, Rural Health Service, Rural Rural Health Service Service, Rural Health Services, Rural Health Rural Health Center Center, Rural Health Centers, Rural Health Health Center, Rural Health Centers, Rural Rural Health Centers	1996	10.144	[mh:noexp] suheading include rural nursing specific to nursing

Community health services	Diagnostic, therapeutic and preventive health services provided for individuals in the community.	Adult Day Care Centers Child Day Care Centers Child Health Services Early Intervention (Education) Maternal-Child Health Services Community Health Nursing Home Health Nursing Parish Nursing Community Mental Health Services Community Networks Community Pharmacy Services Consumer Participation Patient Participation Counseling Distance Counseling Sex Counseling Family Planning Services Foster Home Care Home Care Services Hemodialysis, Home Home Care Services, Hospital-Based Home Health Nursing Home Infusion Therapy Home Nursing + Homemaker Services Parenteral Nutrition, Home + Hospices Maternal Health Services Maternal-Child Health Services Perinatal Care + Preconception Care Prenatal Care Occupational Health Services Senior Centers	Health Services, Community Community Health Service Health Service, Community Service, Community Health Services, Community Health Community Health Care Care, Community Health Health Care, Community Community Healthcare Community Healthcares Healthcare, Community Healthcares, Community	1967	260.670	[mh:noexp] Below term includes too much heteroclites entries

1.2. PHC research strategy

Eight MeSH about Primary Health care, either used alone either with entry terms as text word (TW) and the number of citations retrieved through Inforoute facilities on HeTOP. Copy the queries and input them in the PubMed interface.

Table 2 Query of 8 PHC without TW gives 123,571 citations on July 30, 2017 see https://tinyurl.com/y85kuk66

"Community Health Centers"[Mesh:NoExp] 0R "Community Health Services"[Mesh:NoExp] OR "Community Mental Health Centers"[Mesh:NoExp] OR "Community Mental Health Services"[Mesh] OR "Home Care Agencies"[Mesh] OR "Home Care Services"[Mesh:NoExp] OR "Primary Health Care"[Mesh:NoExp] OR "Rural Health Services"[Mesh:NoExp]

Table 3 Query of 8 PHC with no restriction and TW gives 477,056 citations on July 30, 2017 . The same query with restrictions, adding [MH:NOEXP] gives 237.858 on July 30, 2017

(("community health centers"[MH:NOEXP] OR ("community health centers"[TW] OR "health centers, community"[TW] OR "satellite center"[TW] OR "center, satellite"[TW] OR "centers, neighborhood health"[TW] OR "community health center"[TW] OR "centers, community health"[TW] OR "neighborhood health centers"[TW] OR "health centers, neighborhood"[TW] OR "centers, satellite"[TW] OR "neighborhood health center"[TW] OR "center, community health"[TW] OR "health center, neighborhood"[TW] OR "health center, community"[TW] OR "center, neighborhood health"[TW] OR "satellite centers"[TW] OR "primary care setting"[TW] OR "Community Health Centre"[TW] OR "Community Health Centre"[TW])) OR ("community mental health services"[MH] OR ("community mental health services"[TW] OR "services, community mental health"[TW] OR "community treatment, assertive"[TW] OR "mental health services, community"[TW] OR "assertive community treatment"[TW] OR "treatment, assertive community"[TW] OR "services, mental health community"[TW] OR "health services, community mental"[TW])) OR ("community mental health centers"[MH:NOEXP] OR ("community mental health centers"[TW] OR "cmhc"[TW])) OR ("home care agencies"[MH] OR ("home care agencies"[TW] OR "agency, home health"[TW] OR "care agencies, home"[TW] OR "agencies, home health"[TW] OR "agencies, home care"[TW] OR "home health agency"[TW] OR "home health care agencies"[TW] OR "home care agency"[TW] OR "care agency, home"[TW] OR "agency, home care"[TW] OR "home health agencies"[TW])) OR ("community health services"[MH:NOEXP] OR ("community health services"[TW] OR "service, community health"[TW] OR "health services, community"[TW] OR "care, community health"[TW] OR "health care, community"[TW] OR "health service, community"[TW] OR "community health care"[TW] OR "services, community health"[TW] OR "community healthcare"[TW] OR "healthcare, community"[TW] OR "healthcares, community"[TW] OR "community healthcares"[TW] OR "community health service"[TW])) OR ("rural health services"[MH:NOEXP] OR ("rural health services"[TW] OR "centers, rural health"[TW] OR "services, rural health"[TW] OR "rural health centers"[TW] OR "health centers, rural"[TW] OR "rural health center"[TW] OR "health center, rural"[TW] OR "service, rural health"[TW] OR "rural health service"[TW] OR "health service, rural"[TW] OR "center, rural health"[TW] OR "health services, rural"[TW])) OR ("home care services"[MH:NOEXP] OR ("home care services"[TW] OR "home care service"[TW] OR "service, home care"[TW] OR "care, domiciliary"[TW] OR "care services, home"[TW] OR "services, home care"[TW] OR "domiciliary care"[TW] OR "care, home"[TW] OR "home care"[TW])) OR ("primary health care"[MH:NOEXP] OR ("primary health care"[TW] OR "healthcare, primary"[TW] OR "care, primary"[TW] OR "primary care"[TW] OR "primary healthcare"[TW] OR "care, primary health"[TW] OR "health care, primary"[TW])))

Comments : the growing of citations retrieved is impressive but this is up to the fact that the inforoute facilities don't allow to restrict the MeSH extension as there is no way to apply the [mh:noexp] restriction. Thus the search is done with all the text words of all the below terms in the MeSH hierarchy. This explain the 477.000 citations which are falling halve when using the [mh:noexp]. The addition of TW (terms in free text) is interesting to retrieve not yet indexed citations in PubMed. Nevertheless the noise is important as for instance Text words are searched for in any text including author affiliation.

2. Analyzing General Practice / Family Medicine (GP/FM)

2.1. GP/FM related MeSH

In relation with GP/FM fundamental concepts, we are considering 7 MeSH of which definitions and year of introduction are represented in the below table.

Table 4 MeSH , definitions, entry terms, year of inclusion of the MeSH to be considered in GP/FM search strategy

MeSH	Definition	Entry terms	Year introduced:	Results on June 22, 2016,	Consider use in PHC
Community Medicine	A branch of medicine concerned with the total health of the individual within the home environment and in the community, and with the application of comprehensive care to the prevention and treatment of illness in the entire community.	Medicine, Community	1977	1.925	full
Family Practice	A medical specialty concerned with the provision of continuing, comprehensive primary health care for the entire family.	Family Practices, Practice, Family, Practices, Family	1978-2010	62.608	Full Included in general Practice in 2011
Gatekeeping	The controlling of access to health services, usually by primary care providers; often used in managed care settings to reduce utilization of expensive services and reduce referrals. (From BIOETHICS Thesaurus, 1999)	-	2000	577	full
General Practice	Patient-based medical care provided across age and gender or specialty boundaries.	Practice, General	2011	68.654	Full [Subheadings include Family practice since 2011
General practitioners	Physicians whose practice is not restricted to a specific field of MEDICINE.	General Practitioner, Practitioner, General Practitioners, General, Physicians, General Practice, General Practice Physician, General Practice Physicians, Physician, General Practice, Practice Physicians, General	2011	4.712	full
Physicians, family	Those physicians who have completed the education requirements specified by the American Academy of Family Physicians.	Family Physician, Family Physicians, Physician, Family	1974	15.356	full
Physicians, Primary Care	Providers of initial care for patients. These PHYSICIANS refer patients when appropriate for secondary or specialist care.	Physician, Primary Care, Primary Care Physician Primary Care Physicians	2011	1.860	full

2.2. GP/FM research strategy

Seven MeSH about general practice, family medicine, either used alone, either with entry terms as text word (TW) and the number of citations retrieved as on Dec 4, 2016

Table 5 **Query of 7 GP/FM MeSH gives 91.140 citations on July 30, 2017 see https://tinyurl.com/y8cnzv4v**

"Community Medicine"[Mesh] OR "Family Practice"[Mesh] OR "Gatekeeping"[Mesh] OR "General Practice"[Mesh] OR "General practitioners"[Mesh] OR "Physicians, family"[Mesh] OR "Physicians, Primary Care"

Table 6 **The GP/FM search strategy. The seven MeSH (in bold) and their entry terms for GP/FM as prepared in the HeTOp server: 7 GP/FM with TW = 188,258 on PubMed server using the HeTOP server output. This query generates 195,923 citations on July 30, 2017 see https://tinyurl.com/yd8aywxv**

(("**physicians, primary care"[MH]** OR ("physicians, primary care"[TW] OR "primary care physicians"[TW] OR "primary care physician"[TW] OR "physician, primary care"[TW])) OR **("gatekeeping"[MH]** OR ("gatekeeping"[TW])) OR **("community medicine"[MH]** OR ("community medicine"[TW] OR "medicine, community"[TW])) OR **("general practitioners"[MH]** OR ("general practitioners"[TW] OR "physicians, general practice"[TW] OR "general practice physician"[TW] OR "general practitioner"[TW] OR "practice physicians, general"[TW] OR "practitioners, general"[TW] OR "general practice physicians"[TW] OR "practitioner, general"[TW] OR "physician, general practice"[TW])) OR **("general practice"[MH]** OR ("general practice"[TW] OR "practice, general"[TW])) OR (**"family practice"[MH]** OR ("family practice"[TW] OR "practice, family"[TW] OR "family practices"[TW] OR "general practices"[TW] OR "practices, general"[TW] OR "practices, family"[TW])) OR **("physicians, family"[MH]** OR ("physicians, family"[TW] OR "family physicians"[TW] OR "physician, family"[TW] OR "family physician"[TW])) OR ("family doctor"[TW])) OR "physicians, primary care" OR "gatekeeping" OR "community medicine" OR "general practitioners" OR "general practice" OR "family practice" OR "physicians, family" OR "family doctor "

Comments: those two search strategies could be introduced in the MyNCBI facilities. This allow to cross whatsoever search in PubMed and to retrieve results specific to the two fields

Appendix F

Access to multilingual individual rubrics in URI format for ICPC-2 and the Q-Codes

URI stands for "Uniform Resource Identifier." A URI identifies the name and location of a file or resource in a uniform format. It includes a string of characters for the filename and may also contain the path to the directory of the file. URIs provide a standard way for resources to be accessed by other computers across a network or over the World Wide Web. (https://techterms.com/definition/uri)

Though the URI quoted below, the reader will be able to reach immediately each categories in ICPC-2 and in Q-Codes. URIs have the same structure. Change in language and code give access to the required rubric. (Refer to the figures 1.56, page 88). This access is given on a pedagogical and informative basis.
It is important to take into account that the ICPC classification and its translations are copyrighted by WONCA. On the site url http://www.ph3c.org the reader will find the policy on copyright, licensing and translations.
The Q-Codes belong to Marc Jamoulle and are made available under the Creative Common, share alike, non-commercial license 4.0 (CC BY-NC-SA 4.0). Any use must be reported to the author.

> Citation
> Jamoulle, M., Grosjean, J., & Darmoni, S. (2017). Access to multilingual individual rubrics in URI format for ICPC-2 and the Q-Codes.
> http://orbi.ulg.ac.be/handle/2268/211268

Access to multilingual individual rubrics in URI format for ICPC-2 and the Q-Codes

October 2017
Marc Jamoulle, Julien Grosjean, Stefan Darmoni. D2IM. University of Rouen

ICPC-2 (v6), new ICPC-2 Process (v6) and Q-Codes are freely available rubric by rubric on the University of Rouen's HeTOP server in URI format (Universal Resource Identifier). They are globally available (after free inscription) on http://www.hetop.org A specific Q-Code page is available at http://www.hetop.org/Q

ICPC-2 is available in 19 languages. ICPC-2 rubrics are freely available in 22 languages as permanent resources. For other languages change the two characters *en* in the URI by the corresponding ISO-639 language codes. English language is used by default. Mappings to ICD-10 have been updated in the HeTOP server. Automatic mappings to MeSH are available for all Q-Codes and for 20% of ICPC-2 codes

Access to full ICPC-2 hierarchy in English [1]

- ICPC-2 hierarchy ; http://www.hetop.org/hetop/?la=en&rr=CIP_C_ARBO&tab=1
  - For particular codes, use =CIP_D_ and change the ICPC code at the end.
    Example: ICPC-2 (v6) Pregnancy W78 in English : http://www.hetop.org/hetop/?la=en&rr=CIP_D_W78
  - For other languages, change the ISO-639 for language. Ex; =en for English, =pt for Portuguese, =ja for Japanese etc.
    Example: ICPC-2 (v6) heartburn D03 in Japanese : http://www.hetop.org/hetop/?la=ja&rr=CIP_D_D03
- ICPC-2 Process hierarchy ; http://www.hetop.org/hetop/?la=en&rr=CIP_C_ARBOPROC&tab=1
  - For ICPC-2 Process codes, use =CIP_P_ and the last two digits of the ICPC Process rubric.
    Example: ICPC-2 Process (v6) medical examination -30 ; http://www.hetop.org/hetop/?la=en&rr=CIP_P_30
    Example : ICPC-2 Follow-up 63 in Portuguese ; http://www.hetop.org/hetop/?la=pt&rr=CIP_P_63

[1] URIs are not working on Explorer. Do use Google Chrome or Mozilla

Access to full Q-Codes hierarchy in English [2]

- Q-Codes hierarchy in English;
  `http://www.hetop.eu/hetop/Q?la=en&rr=CGP_CO_Q&tab=1`
- Q-Codes hierarchy in other languages, change the ISO-639 language code. Available (Oct.2017) : English *en*, Spanish *es*, Portuguese *pt*, French *fr*, Dutch *nl*, Korean *ko*, Georgian *ka*, Vietnamese *vi*, Turkish *tr*.
  Example: Full hierachy in Vietnamese
  `http://www.hetop.eu/hetop/Q?la=vi&rr=CGP_CO_Q&tab=1`

Access to each Q-Codes rubric

- For particular Q-Codes, change the Q-Code at the end.
  Example: QC3 social high risk
  `http://www.hetop.eu/hetop/Q?la=en&rr=CGP_QC_QC3`
- For other languages, change the ISO-639 language code.
  Example: QE42 informed consent in Korean *ko*
  `http://www.hetop.eu/hetop/Q?la=ko&rr=CGP_QC_QE42`

[2] URIs are not working on Explorer. Do use Google Chrome or Mozilla

Appendix G

Sources of the definitions of the Q-Codes

Six books were edited of witch the title of the English version is ; *General Practice/Family Medicine Multilingual Terminology - English version*. The same book is available in French, Dutch, Spanish, Portuguese and Vietnamese. Refer to the list of publication at page xxxi to find the citations. The book are the same, co-authored by the translators in each version. The definitions of the Q-Codes were carefully chosen in the international literature and terminological datasets available on-line of which the full list is edited here.

Marc Jamoulle MD. Department of general practice. University of Liège, Belgium
Edited Aug 28, 2016

Sources of the definitions of the Q-Codes

The Q-Codes are a classification of non-clinical items in General Practice / Family Medicine (GP/FM). This list is in an ontology-ready format, and made available at www.hetop.eu/Q by the Laboratory of Bioinformatics, University of Rouen, France.
The Q-Codes, whose classification is complementary to ICPC-2, are part of the 3CGP project and are intended to retrieve the main organizational concepts in GP/FM.
3CGP is an acronym for "Core Content Classification in Primary Care." The 3CGP project joins ICPC-2 and Q-Codes to create a new ontology to improve the indexing of grey literature in GP/FM.

The definitions of the Q-Code have been carefully selected from online dictionaries, terminologies and papers which the list is published here. The purpose of this bibliography is to highlight the content of the Q-Codes. Refer to http://3CGP.docpatient.net for more information about Q-Codes

.

Aamland, Aase, Kirsti Malterud, and Erik L. Werner. 2014. "Patients with Persistent Medically Unexplained Physical Symptoms: A Descriptive Study from Norwegian General Practice." *BMC family practice* 15(1):107. Retrieved December 10, 2014 (http://www.biomedcentral.com/1471-2296/15/107).

Adler, Herbert M. 2002. "The Sociophysiology of Caring in the Doctor-Patient Relationship." *Journal of General Internal Medicine* 17(11):883–90. Retrieved July 25, 2016 (http://link.springer.com/10.1046/j.1525-1497.2002.10640.x).

AHRQ Effective Health Care Program .Glossary. Retrieved October 1, 2015 (http://effectivehealthcare.ahrq.gov/index.cfm/options/glossary/?keywords=glossary).

Allen, Justin et al. 2005. "The European Definition of General Practice/family Medicine." *European Academy of Teachers in General Practice* 1–11. Retrieved (http://www.aile.net/sunum/dersnot/fame/module1/pres1/pres1europeandef.pdf).

American Association of Family Physicians. 2011. "AAFP Policies - Primary Care Definitions." *AAFP web page*. Retrieved February 7, 2016 (http://www.aafp.org/about/policies/all/primary-care.html).

Antoniou, G., and Frank van Harmelen. 2004. *Handbook on Ontologies in Information Systems, Chapter Web Ontology Language: OWL*. Springer Verlag.

APA. 2011. "The Guidelines for Psychological Practice with Lesbian, Gay, and Bisexual Clients.Definition of Terms: Sex, Gender, Gender Identity, Sexual Orientation." Retrieved July 24, 2016 (https://www.apa.org/pi/lgbt/resources/sexuality-definitions.pdf).

Barnes, Danielle M., and Michael H. Logan. 2010. "Dental Modification: An Anthropological Perspective." University of Tennessee - Knoxville. Retrieved (http://trace.tennessee.edu/utk_chanhonoproj\nhttp://trace.tennessee.edu/utk_chanhonoproj/1345).

Bradley, Elizabeth H., Leslie A. Curry, and Kelly J. Devers. 2007. "Qualitative Data Analysis for Health Services Research: Developing Taxonomy, Themes, and Theory." *Health services research* 42(4):1758–72. Retrieved July 9, 2014 (http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1955280&tool=pmcentrez&rendertype=abstract).

Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde. 2007. "DeCS - Descritores Em Ciências Da Saúde Guia de Atualização Do DeCS. / BIREME / OPAS / OMS." 23. Retrieved (http://bvsmodelo.bvsalud.org/download/decs/DeCS-GuiaAtualizacao.pdf).

CIHR Canadian Institute of Health Research. n.d. “Knowledge Translation - Definition.” Retrieved (http://www.cihr-irsc.gc.ca/e/29418.html#2).

Clinical Network Development Unit. 2006. *Glossary of Commonly Used Terms in Community and Primary Care*.

Coletti, M. H., and H. L. Bleich. “Medical Subject Headings Used to Search the Biomedical Literature.” *Journal of the American Medical Informatics Association : JAMIA* 8(4):317–23. Retrieved January 17, 2016 (http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=130076&tool=pmcentrez&rendertype=abstract).

DeCS - Descriptores de Saude - Health Descriptors. *PAHO Bireme Sao Paulo*. Retrieved March 11, 2016 (http://decs.bvs.br/I/homepagei.htm).

Denzin, Norman, and Yvonna S. Lincoln. 2000. “The Sage Handbook of Qualitative Research (2nd Ed.).” P. 784 in *Sage Publications*. Thousands Oaks: Sage.

ERIC. 2015. “Education Resources Information Center.” *Institue of Education Sciences, USA*. Retrieved September 4, 2015 (http://eric.ed.gov/?ti=all).

Ernst, E. et al. 1995. “Complementary Medicine — a Definition.” *The British Journal of General Practice* 45(398):506.

Euract. 2014. *The EURACT Performance Agenda of General Practice /Family Medicine. Assessment of General Practitioners’ Performance in Daily Practice*. edited by Stefan Wilm. düsseldorf university press, Düsseldorf 2014. Retrieved (www.euract.eu).

EXPH (EXpert Panel on effective ways of investing in health). 2014. *Preliminary Report on Definition of a Frame of Reference in Relation to Primary Care with a Special Emphasis on Financing Systems and Referral Systems*. European Union. Retrieved (http://ec.europa.eu/health/expert_panel/opinions/docs/001_definitionprimarycare_en.pdf).

FCA. 2014. “FCA Definitions.” *Family Caregiver Alliance*. Retrieved (https://www.caregiver.org/definitions-0).

Flitcraft, Anne, Susan M. Hadley, Marybeth K. Hendricks-Matthews, Susan V McLeer, and Carole Warshaw. 1992. *Diagnostic and Treatment Guidelines on Domestic Violence 2 These Guidelines Were Prepared*. Retrieved July 25, 2016 (http://www.ncdsv.org/images/AMA_Diag&TreatGuideDV_3-1992.pdf).

Genuis, Stephen J., Chris Lipp, Stephen J. Genuis, and Chris Lipp. 2013. “Ethical Diversity and the Role of Conscience in Clinical Medicine.” *International journal of family medicine* 2013:587541. Retrieved July 25, 2016 (http://www.ncbi.nlm.nih.gov/pubmed/24455248).

Given, Lisa M. 2013. *The SAGE Encyclopedia of Qualitative Research Methods Codes and Coding*. SAGE Publications Ltd;

Govender, Veloshnee, and Loveday Penn-Kekana. 2007. “Gender Biases and Discrimination: A Review of Health Care Interpersonal Interactions.”

GPRD - General Practice Research Database. *https://www.herc.ox.ac.uk*. Retrieved (https://www.herc.ox.ac.uk/downloads/health_datasets/browse-data-sets/general-practice-research-database-gprd).

Gruber, Thomas R. 1993. *A Translation Approach to Portable Ontology Specifications by A Translation Approach to Portable Ontology Specifications*. Retrieved (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.117.3273&rep=rep1&type=pdf).

Haggerty, Jeannie et al. 2007. "Operational Definitions of Attributes of Primary Health Care: Consensus among Canadian Experts." *Annals of family medicine* 5(4):336–44. Retrieved February 9, 2016 (http://www.annfammed.org/content/5/4/336.abstract).

Harnad, Stevan. 2005. "To Cognize Is to Categorize: Cognition Is Categorization." Pp. 19–43 in *Handbook of Categorization in Cognitive Science*. Elsevier.

Hattis, RP, and Law M. n.d. "The Five Stages of Prevention: A Practical New Classification for Health Planning and Clinical Practice." Retrieved November 9, 2014 (http://www.capreventivemed.org/wp-content/themes/eventbrite-child/documents/Stages of Prevention 1-12.pdf).

Hausmann Muela, Susanna, Joan Muela Ribera, Elizabeth Toomer, and Koen Peeters Grietens. 2012. "The PASS-Model: A Model for Guiding Health-Seeking Behavior and Access to Care Research." *Malaria Reports* 2(1):3. Retrieved (http://www.pagepressjournals.org/index.php/malaria/article/view/malaria.2012.e3).

Heyvaert, M., ·. B. Maes, P. Onghena, and B. Maes. n.d. "Mixed Methods Research Synthesis: Definition, Framework, and Potential."

Horton, Richard, and Selina Lo. 2015. "Planetary Health: A New Science for Exceptional Action." *Lancet (London, England)* 386(10007):1921–22. Retrieved May 11, 2016 (http://www.ncbi.nlm.nih.gov/pubmed/26188746).

Howie, J. G. R., D. Heaney, and M. Maxwell. 2004. "Quality, Core Values and the General Practice Consultation: Issues of Definition, Measurement and Delivery." *Family practice* 21(4):458–68. Retrieved March 15, 2012 (http://www.ncbi.nlm.nih.gov/pubmed/15249538).

IHTSDO. n.d. "SNOMED-CT." Retrieved February 13, 2014 (http://www.ihtsdo.org/snomed-ct/).

ISCO- International Labour Organization. 2007. "International Standard Classification of Occupations Draft ISCO-08 Group Definitions: Occupations in Health." *ILO Bureau of statistics* 56. Retrieved September 19, 2015 (http://www.ilo.org/public/english/bureau/stat/isco/docs/health.pdf).

Jupp, Victor. 2006. *The SAGE Dictionary of Social Research Methods :* SAGE Publications. Retrieved August 16, 2014 (http://srmo.sagepub.com/view/the-sage-dictionary-of-social-research-methods/SAGE.xml).

Kaba, R., and P. Sooriakumaran. 2007. "The Evolution of the Doctor-Patient Relationship." *Int J Surg.* 5(1).

Kaur, Manjit, and Sarah Hall. 2002. *ECHO. Medical Supplies and Equipment for Primary Health Care. A Practical Resource for Procurement and Management*. edited by Kathy Attawell.

Kruschinski, Carsten, Maaike Lange, Christos Lionis, Chris van Weel, and Eva Hummers-Pradier. 2010. "Themes and Methods of Research Presented at European General Practice Research Network Conferences." *Family practice* 27(4):459–67. Retrieved January 28, 2016 (http://www.ncbi.nlm.nih.gov/pubmed/20427325).

Labaree, Robert V. n.d. "Research Guides: Organizing Your Social Sciences Research Paper: Types of Research Designs."

Lavrakas, Paul J. 2008. "ESRM Entries a-Z." in *Encyclopedia of Survey Research Methods. Sage online publications*. Retrieved (http://srmo.sagepub.com/view/encyclopedia-of-survey-research-methods/SAGE.xml).

Le Reste, Jean Yves et al. 2013. "The European General Practice Research Network Presents a Comprehensive Definition of Multimorbidity in Family Medicine and Long Term Care, Following a Systematic Review of Relevant Literature." *Journal of the American Medical Directors*

Association 14(5):319–25. Retrieved November 17, 2014 (http://www.ncbi.nlm.nih.gov/pubmed/23411065).

Légaré, France et al. 2010. "Effective Continuing Professional Development for Translating Shared Decision Making in Primary Care: A Study Protocol." *Implementation science : IS* 5:83. Retrieved November 15, 2014 (http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2988066&tool=pmcentrez&rendertype=abstract).

Macmillan Dictionnary Online.Retrieved (http://www.macmillandictionary.com/).

Mariño, María Ana. 2015. "Incidentalomas: Concept, Relevance and Challenges for Medical Practice." *Revista Brasileira de Medicina de Família e Comunidade* 10(35):1. Retrieved July 25, 2016 (http://www.rbmfc.org.br/rbmfc/article/view/1053).

McCloskey, Laura A. et al. 2007. "Abused Women Disclose Partner Interference with Health Care: An Unrecognized Form of Battering." *Journal of general internal medicine* 22(8):1067–72. Retrieved July 25, 2016 (http://www.ncbi.nlm.nih.gov/pubmed/17530312).

MeSH. MeSH Search Facilities. Retrieved (http://www.ncbi.nlm.nih.gov/mesh/).

Morse, J. M., and Lyn Richards. 2002. *Readme First for a User's Guide to Qualitative Methods.* Thousand Oaks, CA: Sage Publications.

Moynihan, Ray, Evan Doran, and David Henry. 2008. "Disease Mongering Is Now Part of the Global Health Debate." *PLoS medicine* 5(5):e106. Retrieved November 26, 2014 (http://dx.plos.org/10.1371/journal.pmed.0050106).

Moynihan, Ray, Jenny Doust, and David Henry. 2012. "Preventing Overdiagnosis: How to Stop Harming the Healthy." *BMJ (Clinical research ed.)* 344:e3502. Retrieved August 12, 2012 (http://www.ncbi.nlm.nih.gov/pubmed/22645185).

NAWHME 1996, and GD Donohue. 1996. *WOMEN ' S HEALTH in the CURRICULUM RESOURCE GUIDE FOR FACULTY*. Philadelphia,: National Academy on Women's Health Medical Education (NAWHME).

Niven, Daniel J. et al. 2015. "Towards Understanding the de-Adoption of Low-Value Clinical Practices: A Scoping Review." *BMC medicine* 13:255. Retrieved July 25, 2016 (http://www.ncbi.nlm.nih.gov/pubmed/26444862).

Nutbeam, D. 1998. *Health Promotion Glossary*. Geneva: World Health Organization. Retrieved (http://www.heapro.oxfordjournals.org/cgi/doi/10.1093/heapro/13.4.349).

Pizzanelli, Miguel. 2015. "¿Overscreening O Prevención a Escala Humana? Tamizaje Excesivo." *Revista Brasileira de Medicina de Família e Comunidade* 10(35):1. Retrieved August 5, 2016 (http://www.rbmfc.org.br/rbmfc/article/view/1068).

Porta, Miguel. 2014. *A Dictionary of Epidemiology*. Sixth edit. edited by International epidemiological association; Oxford University Press. Retrieved September 8, 2015 (http://www.irea.ir/files/site1/pages/dictionary.pdf).

Rask, Mette T. et al. 2014. "Towards a Clinically Useful Diagnosis for Mild-to-Moderate Conditions of Medically Unexplained Symptoms in General Practice: A Mixed Methods Study." *BMC Family Practice* 15(1):118. Retrieved July 25, 2016 (http://bmcfampract.biomedcentral.com/articles/10.1186/1471-2296-15-118).

Reitz, Joan. 2013. "ODLIS: Online Dictionary of Library and Information Science." Retrieved July 23, 2015 (http://www.library.ucsb.edu/research/db/1182).

ReliefWeb. 2008. *RELIEFWEB GLOSSARY OF HUMANITARIAN TERMS*. Retrieved (http://reliefweb.int/report/world/reliefweb-glossary-humanitarian-terms).

Resnik, David B. 2015. “Glossary of Commonly Used Terms in Research Ethics.” *National Institute of Environmental Health Science, National Institutes of Health*.

Rodríguez, Charo et al. 2014. “Family Physicians’ Professional Identity Formation: A Study Protocol to Explore Impression Management Processes in Institutional Academic Contexts.” *BMC medical education* 14(1):184. Retrieved October 11, 2015 (http://www.biomedcentral.com/1472-6920/14/184).

Royal College of General Practitioners. 1973. “A General-Practice Glossary.” *Journal of the Royal College of General Practitioners* 3(Suppl n°3):15. Retrieved November 5, 2014 (http://trove.nla.gov.au/work/21584373?selectedversion=NBD332424).

Ryu, Seewon, Seung Hee Ho, and Ingoo Han. n.d. “Knowledge Sharing Behavior of Physicians in Hospitals.” Pp. 1–20 in *Expert Systems with applications*. Elsevier. Retrieved (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.104.1314&rep=rep1&type=pdf).

SNOMED-CT. n.d. “The IHTSDO SNOMED CT Browser.” Retrieved (http://browser.ihtsdotools.org/?).

Švab, Igor, and Milica Katić. 2014. “International Textbook of Family Medicine: The Application of EURACT Teaching Agenda.” *Acta medica academica* 43(1):30–34. Retrieved January 6, 2015 (http://www.ncbi.nlm.nih.gov/pubmed/24893636).

Tashakkori, A., and C. Teddlie. 2003. “Handbook of Mixed Methods in Social & Behavioral Research.” *Sage Publication* 14(1):893.

The society of homeopaths. n.d. “ABOUT HOMEOPATHY.” Retrieved (http://www.homeopathy-soh.org/about-homeopathy).

UN WOMEN 2001. “Concepts and Definitions.” *United Nations Entity for Gender Equality* . Retrieved (http://www.un.org/womenwatch/osagi/conceptsandefinitions.htm).

UNESCO. 2006. “DIGITAL LIBRARIES IN EDUCATION GLOSSARY OF TERMS.” *UNESCO INSTITUTE FOR INFORMATION TECHNOLOGIES IN EDUCATION* 116–18.

United Nations. 1985. *Compendium of United Nations Standards and Norms in Crime Prevention and Criminal Justice II. Victims*. Retrieved (https://www.unodc.org/pdf/compendium/compendium_2006_part_03_02.pdf).

van den Akker, M., F. Buntinx, J. F. Metsemakers, and J. A. Knottnerus. 1998. “Morbidity in Responders and Non-Responders in a Register-Based Population Survey.” *Fam.Pract.* 15(3):261–63.

van Ravenzwaaij, J. et al. 2010. “Explanatory Models of Medically Unexplained Symptoms: A Qualitative Analysis of the Literature.” *Mental health in family medicine* 7(4):223–31. Retrieved December 10, 2014 (http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3083258&tool=pmcentrez&rendertype=abstract).

Vaz Cardoso, Raquel. 2014. *Medicalization and Health Care in the Family Strategy (Medicalização E O Cuidado Em Saúde Na Estratégia de Saúde Da Família)*. edited by Thesis. Universidade Estadual de Campinas, Brazil. Retrieved April 28, 2014 (http://www.ph3c.org/PH3C/docs/27/000281/0000605.pdf).

Webster. Retrieved (http://www.merriam-webster.com/).

White, Alan et al. 2011. *EU 2011 . The State of Men's Health in Europe Report*. Retrieved (doi:10.2772/60721).

WHO Centre for Health Development. 2004. "WHO Glossary of Terms for Community Health Care and Services for Older Persons." *Ageing and Health Technical Report. WHO/WKC/Tech.Ser./04.2* 5:109.

WHO Upssala. 2015. "WHO UPssala Glossary of Terms in Pharmacovigilance." *WHO Upssala monitoring centtre*. Retrieved (http://who-umc.org/DynPage.aspx?id=97224&mn1=7347&mn2=7252&mn3=7257).

WHO. 1984. *WHO Glossary of Terms Used in Health for All Series.* Geneva: World Health Organization.

Wilson, Robert A., and Frank C. Keil. 2001. *The MIT Encyclopedia of the Cognitive Sciences*. MIT Press. Retrieved June 6, 2012 (http://books.google.com/books?id=-wt1aZrGXLYC&pgis=1).

Wojtczak, A. 2003. "IIME. Glossary of Medical Education Terms Online." *The Institute for International Medical Education. New York*. Retrieved (http://www.iime.org/glossary.htm).

Woncadic. 2003. *Wonca Dictionary of General/Family Practice*. Maanedsskr. edited by Niels Bentzen. Copenhagen. Retrieved (http://www.ph3c.org/PH3C/docs/27/000092/0000052.pdf).

World Health Organization. 2016. "WHO | Health Systems Strengthening Glossary." *WHO web pages*. Retrieved February 10, 2016 (http://www.who.int/healthsystems/hss_glossary/en/index8.html).

BIBLIOGRAPHY

AAFP (2011). *Primary Care definitions – AAFP Policies*. URL: http://www.aafp.org/about/policies/all/primary-care.html (visited on 02/07/2016).

Abele, Andrejs et al. (2015). *Linking Open Data cloud diagram 2017*. URL: http://lod-cloud.net/.

Adcock, Robert and David Collier (2001). "A Shared Standard for Qualitative and Quantitative Research". In: *American Political Science Review* 95.3, pp. 529–546. ISSN: 00030554. DOI: 10.1017/S0003055401003100. arXiv: arXiv:1011.1669v3.

Allemang, Dean and James Hendler (2008). *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. Morgan Kaufmann Publishers, p6–p6. ISBN: ISBN-13: 978-0123735560.

Allen, Bradley P (2016). "The role of metadata in the second machine age". In: *Second International Conference Establishment Surveys*. URL: http://www.amstat.org/meetings/ices/2000/proceedings/S57.pdf.

Allen, Justin et al. (2005). "The European definition of general practice/family medicine". In: *European Academy of Teachers in General Practice*, pp. 1–11. ISSN: 08954356. DOI: 10.1016/S0895-4356(00)00274-2.

Allen, Justin et al. (2011). "The European Definition Of General Practice / Family Medicine". In: URL: http://www.woncaeurope.org.

Alpaydin, Ethem (2016). *Machine Learning: The New AI*. MIT Press.

Angell, Marcia (2017). "Drug Companies & Doctors: A Story of Corruption". In:

Antoniou, G and Frank van Harmelen (2004). *Handbook on Ontologies in Information Systems, chapter Web Ontology Language: OWL*. Springer Verlag., pp. 67–92.

Baca, Murtha (2008). *Introduction to metadata*. Getty Publications. URL: http://www.getty.edu/publications/intrometadata.

Bachman, CW (1969). "Data structure diagrams". In: *SIGMIS Newsletter* 1.2, pp. 4–10. DOI: 10.1145/1017466.1017467. URL: http://dl.acm.org/citation.cfm?id=1017467.

Balint, Michael and J. P. Valabrega (1960). "Le Médecin, Son Malade Et la Maladie". In: *Les Etudes Philosophiques* 15.4, pp. 526–526.

Bartholomeeusen, Stefaan, Frank Buntinx, and Jan Heyrman (2002). "Ziekten in de huisartspraktijk: methode en eerste resultaten van het Intego-netwerk". In: *Tijdschrift voor Geneeskunde* 58.863-871.

Bauer, Florian and Martin Kaltenbock (2012). *Linked Open Data : The Essentials*, pp. 3–56. URL: http://www.semantic-web.at/LOD-TheEssentials.pdf.

Becker, HW et al. (2005). *ICPC2-ICD10 Thesaurus. A diagnostic terminology for semi-automatic double coding in Electronic Patient Records. ICPC in the Amsterdam Transition Project. CD-Rom*. Tech. rep. Academic Medical Center/University of Amsterdam, Department of Family Medicine,

Benjamins, V Richard et al. (2002). "Six Challenges for the Semantic Web". In: *Challenges* 1.i, pp. 1–15.

Bentsen, B G (1986). "International classification of primary care". In: *Scandinavian journal of primary health care* 4.1, pp. 43–50.

Bentsen, Bent (1976). "An international classification of the health problems of primary care". In: *The Journal of the Royal College of General Practitioners. Occasional Paper* 1, pp. 1–41. ISSN: 0309-6300. URL: http://www.ncbi.nlm.nih.gov/pubmed/1053266.

Bentzen, Niels (1995). "An international glossary for general/family practice". In: *Family Practice* 12.3, pp. 341–369. ISSN: 0263-2136. DOI: 10.1093/fampra/12.3.341.

Bentzen N.(ed) (2003). *Wonca Dictionary of general/family practice*. Ed. by Niels Bentzen. Maanedsskr. Copenhagen. URL: http://www.ph3c.org/PH3C/docs/27/000092/0000052.pdf.

Bernaras, Amaia, Iñaki Laresgoiti, and Jose Corera (1996). "Building and Reusing Ontologies for Electrical Network Applications'". In: *ECAI*. PITMAN, pp. 298–302.

Berners-Lee, Tim (1998). *An attempt to give a high-level plan of the architecture of the Semantic WWW*. URL: http://www.w3.org/DesignIssues/Semantic.html.

— (2002). *What do HTTP URIs Identify?* URL: http://www.w3.org/DesignIssues/HTTP-URI.

Berners-Lee, Tim, J Hendler, and O. Lassila (2001). "The SemanticWeb: A New Form of Web Content that is Meaningful to Computers Will Unleash a Revolution of New Possibilities." In: *The Scientific American* 284.5, 34–43.

Binsbergen, J J van and A J Drenthen (1999). "ICPC-code approach of nutritional questions in general practice: a look at the future." In: *European journal of clinical nutrition* 53 Suppl 2, S22–4. ISSN: 0954-3007. URL: http://www.ncbi.nlm.nih.gov/pubmed/10406432.

Bizer, C, T Heath, and Tim Berners-Lee (2009). *Linked Data - The Story So Far. Special Issue on Linked Data*. URL: http://linkeddata.org/docs/ijswis-special-issue.

Blake, Gary and Robert W Bly (1993). *The elements of technical writing*. Macmillan New York, NY.

Blashki, Grant et al. (2014). *Family doctors in the field. Environmental stories from across the globe*. World Organization of Family Doctors. URL: https://tinyurl.com/WONCA-QH.

Blaum, Wolf E et al. (2013). "Towards Web 3.0: taxonomies and ontologies for medical education – a systematic review." In: *GMS Zeitschrift für medizinische Ausbildung* 30.1, Doc13. DOI: 10.3205/zma000856.

Bonnet, Doris (1999). "La taxinomie des maladies en anthropologie : aperçu historique et critique". In: *Sciences Sociales et Santé* 17.2, pp. 5–21. URL: http://www.documentation.ird.fr/hor/fdi:010018593.

Boole, Georges (1854). "The Laws of Thought 1854/2003. Facsimile of 1854 edition, with an introduction by J. Corcoran. Buffalo: Prometheus Books (2003). Reviewed by James van Evra ". In: *Philosophy in Review* 24, pp. 167–169.

Boot, Cécile R L and Frans J Meijman (2010). "Classifying health questions asked by the public using the ICPC-2 classification and a taxonomy of generic clinical questions: an empirical exploration of the feasibility." In: *Health Communication* 25.2, pp. 175–181. URL: http://www.ncbi.nlm.nih.gov/pubmed/20390683.

Bowker, GC and SL Star (1999). *Sorting Things Out: Classification and Its Consequences.* Cambridge, MA: MIT Press.

Bradley, Elizabeth H, Leslie A Curry, and Kelly J Devers (2007). "Qualitative data analysis for health services research: developing taxonomy, themes, and theory." In: *Health services research* 42.4, pp. 1758–72. DOI: 10.1111/j.1475-6773.2006.00684.x.

Braslavski, Pavel et al. (2015). *Introduction to Formal Concept Analysis and Its Applications in Information Retrieval and Related Fields*. Vol. 505. December 2015. ISBN: 9783319254845. DOI: 10.1007/978-3-319-25485-2.

Bratková, Eva and Helena Kučerová (2014). "Knowledge Organization Systems and Their Typology". In: *Knihovna [online]* 25, supplementum 2. URL: http://full.nkp.cz/nkkr/knihovna142_suppl/1402sup01.htm.

Braun, R N (1965). "ORGANIZING DIAGNOSIS IN GENERAL PRACTICE." In: *World medical journal* 12, pp. 50–3. ISSN: 0049-8122. URL: http://www.ncbi.nlm.nih.gov/pubmed/14286699.

Braun, Robert Nikolaus (1970). *Lehrbuch der ärztlichen Allgemeinpraxis [Gebundene Ausgabe]*. Urban u. Schwarzenberg.

Britt, H et al. (2003). "Bettering the Evaluation And Care of Health 2001-2002 (summary of results)". In: *Australian Family Physician* 32.1/2, pp. 59–63. URL: http://www.racgp.org.au/document.asp?id=8998.

Britt, Helena et al. (2016). *A decade of Australian general practice activity 2006–07 to 2015–16. Bettering the Evaluation and Care of Health (BEACH)*. Report. Family Medicine Research Centre. Sydney School of Public Health. University of Sydney. URL: https://ses.library.usyd.edu.au/bitstream/2123/15482/5/9781743325162{_}ONLINE.pdf.

Brown, Lynsey et al. (2014). "Facilitating access to evidence: Primary Health Care Search Filter." In: *Health information and libraries journal* 31.4, pp. 293–302. DOI: 10.1111/hir.12087.

Buono, Nicola et al. (2013). "40 years of biannual family medicine research meetings – The European General Practice Research Network (EGPRN)". EN. In: *Scandinavian Journal of Primary Health Care*. URL: http://www.tandfonline.com/doi/full/10.3109/02813432.2013.847594.

Bushman, B., D. Anderson, and G. Fu (2015). "Transforming the Medical Subject Headings into Linked Data: Creating the Authorized Version of

MeSH in RDF". In: *J Libr Metadata* 15.3-4, pp. 157–176. ISSN: 1938-6389 (Print)1937-5034. DOI: 10.1080/19386389.2015.1099967.

Cabot, Chloé et al. (2017a). "Evaluation of the Terminology Coverage in the French Corpus LiSSa". In: pp. 2–6.

Cabot, Chloé et al. (2017b). "Evaluation of the Terminology Coverage in the French Corpus LiSSa." In: *Studies in health technology and informatics* 235, pp. 126–130.

Cardillo, Elena (2011). "A lexi-ontological resource for Consumer healthcare.The italian consumer medical vocabulary". PhD thesis. Trento, p. 192. URL: http://eprints-phd.biblio.unitn.it/570/1/PhD-Thesis-Elena-Cardillo.pdf.

— (2015a). "Mapping between international medical terminologies to SHN Work Package 3". In: *SemanticHealthNet*. Chap. Deliverable 3.3, 18p.

— (2015b). "Medical terminologies for patients Annex 1 to SHN Work Package 3". In: *SemanticHealthNet*. Chap. Deliverable 3.3, p. 16.

Cardillo, Elena et al. (2008). "Logical Analysis of Mappings between Medical Classification Systems". In: *International Conference on Artificial Intelligence: Methodology, Systems, and Applications AIMSA 2008*. Vol. 5253. Springer Berlin Heidelberg, pp. 311–321. ISBN: 978-3-540-85775-4. URL: http://www.springerlink.com/content/b225335809376016/.

Carey, Iain M et al. (2004). "Developing a large electronic primary care database (Doctors' Independent Network) for research". In: *International Journal of Medical Informatics* 5, pp. 443–453. ISSN: 1386-5056. DOI: 10.1016/j.ijmedinf.2004.02.002.

Casado Vicente, Verónica (2012). *Tratado de medicina de familia y comunitaria*. Ed. by Verónica Casado Vicente. Médica Panamericana, p. 2563. ISBN: 8498355850.

Cavadas, L F, T Villanueva, and J Gervas (2010). "General practice innovation: a Portuguese virtual conference". In: *Med Educ* 44.5, pp. 514–515. ISSN: 0308-0110. DOI: 10.1111/j.1365-2923.2010.03649.x.

CFPC (1980-1992). *FAMLI: Family Medicine Literature Index*. vol.1-12. World Organization of National Colleges, Academies, and Academic Associations of General Practitioners/Family Physicians (WONCA). URL: https://books.google.be/books?id=InlRAQAAIAAJ.

Challis, EB (1981). "FAMLI: A Tribute". In: *Canadian family physician Médecin de famille canadien* 26.November, p. 1470. ISSN: 0008-350X. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2383786/.

Charity, Melanie J et al. (2013). "Extending ICPC-2 PLUS terminology to develop a classification system specific for the study of chiropractic encounters." En. In: *Chiropractic & manual therapies* 21.1, p. 4. ISSN: 2045-709X. DOI: 10.1186/2045-709X-21-4.

Charlton, Rachel A et al. (2010). "Identifying major congenital malformations in the UK General Practice Research Database (GPRD)". In: *Drug Safety: An International Journal of Medical Toxicology and Drug Experience* 9, pp. 741–750. ISSN: 0114-5916. DOI: 10.2165/11536820-000000000-00000.

Cheung, Kei-Hoi et al. (2009). "Semantic Web for Health Care and Life Sciences: a review of the state of the art." In: *Briefings in bioinformatics* 10.2, pp. 111–3. ISSN: 1477-4054. DOI: 10.1093/bib/bbp015.

Chiaro, Natalia and Pablo Damonte (2005). "Generación de descripciones de Web Semántica en una red de ONG". PhD thesis. Tesis de grado. Universidad de la República (Uruguay). Facultad de Ingeniería. Instituto de Computación. URL: https://www.colibri.udelar.edu.uy/handle/123456789/3081?locale=es.

Chinitz, David P and Victor G Rodwin (2014). "Perspective On Health Policy and Management (HPAM): mind the theory-policy-practice gap". In: *International Journal of Health Policy Management* 3.x, pp. 1–3. DOI: 10.15171/ijhpm.2014.122.

Chute, C G, S P Cohn, and J R Campbell (1998). "A Framework for Comprehensive Health Terminology Systems in the United States: Development Guidelines, Criteria for Selection, and Public Policy Implications". In: *Journal of the American Medical Informatics Association* 6, pp. 503–510. DOI: 10.1136/jamia.1998.0050503.

Cimino, J J (1996). "Review paper: coding systems in health care." In: *Methods of information in medicine* 35.4-5, pp. 273–84. ISSN: 0026-1270. URL: http://www.ncbi.nlm.nih.gov/pubmed/9019091.

— (1998). "Desiderata for controlled medical vocabularies in the twenty-first century". In: *Methods of Information in Medicine* 37.4-5, pp. 394–403.

Clark, Malcolm et al. (2012). "Automatically structuring domain knowledge from text: An overview of current research". In: 48.3, 552–568. DOI: 10.1016/j.ipm.2011.07.002.

Colliers, Annelies et al. (2016). "Improving Care And Research Electronic Data Trust Antwerp (iCAREdata): a research database of linked data on out-of-hours primary care". In: *BMC Research Notes* 9.1, p. 259. DOI: 10.1186/s13104-016-2055-x.

congress, Library of (2017). *MARC STANDARDS (Network Development and MARC Standards Office, Library of Congress)*. URL: https://www.loc.gov/marc/ (visited on 10/21/2017).

Corbin, Juliet and Anselm Strauss (1990). "Grounded theory research: Procedures, canons and evaluative criteria". In: *Zeitschrift für Soziologie* 19.6, pp. 418–427.

Corcho, Oscar et al. (2005). "Building legal ontologies with METHONTOLOGY and WebODE". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 3369 LNAI, pp. 142–157. ISSN: 3540250638. DOI: 10.1007/978-3-540-32253-5_9.

Darmoni, S and M Joubert (2000). "CISMeF: a structured health resource guide." In: *Methods of information in medicine* 39.1, pp. 30–35.

Darmoni, SJ et al. (2012). "Improving information retrieval using MeSH Concepts: a test case on rare and chronic diseases". In: *J Med Libr Assoc* 100.3, pp. 176–83. DOI: 10.3163/1536-5050.100.3.007.

Darmoni, Stéfan Jacques et al. (2001). "The use of Dublin Core metadata in a structured health resource guide on the Internet". In: *Bulletin of the Medical Library Association* 89.3, p. 297.

David, A.K. et al. (2013). *Family Medicine: Principles and Practice*. Springer Science & Business, p. 1240. ISBN: 0387217444.

Davis, D. A. (2004). "CME and the pharmaceutical industry: two worlds, three views, four steps". In: *CMAJ*. Vol. 171, pp. 149–50. ISBN: 0820-3946 (Print)1488-2329 (Electronic). DOI: 10.1503/cmaj.1040361.

Dekkers, Makx (2009). "History, objectives and approaches of the Dublin Core Metadata Initiative". In: pp. 1–24. URL: http://dublincore.org/resources/training/frd_20091217/Tutorial_FRD_dekkers-1.pdf.

Demner-Fushman, Dina et al. (2013). "YAGO2 : A Spatially and Temporally Enhanced Knowledge Base from Wikipedia". In: *Journal of Biomedical Informatics* 46.SUPPL. Pp. 129–132. DOI: 10.1016/j.jbi.2013.09.007.

Denzin, Norman and Yvonna S Lincoln (2000). "The Sage handbook of qualitative research (2nd ed.)" In: *Sage Publications*. Thousands Oaks: Sage., p. 784. ISBN: 9781412974172. DOI: Doi10.1177/1354067x07080505.

Derks, D, A E Bos, and J von Grumbkow (2008). "Emoticons in computer-mediated communication: social motives and social context". eng. In: *Cyberpsychol Behav* 1, pp. 99–101. DOI: 10.1089/cpb.2007.9926.

Dias, Rodrigo Da Silva et al. (2015). "Telemental health in Brazil: past, present and integration into primary care". In: *Archives of Clinical Psychiatry (São Paulo)* 42.2, pp. 41–44.

Dick, Bob (2014). *Grounded theory: a thumbnail sketch*. URL: http://www.aral.com.au/resources/grounded.html (visited on 04/14/2016).

dictionary, Dorlands medical (1981). "Medical dictionary". In: *Twenty-sixth Edition. Philadelphia. WB saunders company*.

Dictionary, The Merriam-Webster (2006). *The Merriam-Webster Dictionary*. Merriam-Webster, Incorporated.

Doherty William J., Sussman Marvin B. (1987). *Family Medicine: The Maturing of a Discipline*. Psychology Press, p. 236. URL: https://books.google.com/books/about/Family_Medicine.html?hl=fr&id=Q_GsIIMRG4wC.

Doorn, Peter (2017). "Monitoring the FAIRness of data sets - Introducing the DANS approach to FAIR metrics". In: *OSFAIR2017*. Athens, Greece: National Library of Greece. URL: http://www.opensciencefair.eu/speakers/peter-doorn.

Dowden, John (2015). "Conflict of interest in medical journals". In: *Australian Prescriber* 38.1, pp. 2–3. DOI: 10.18773/austprescr.2015.001.

Druais, PL et al. (2009). *Médecine générale*. Ed. by D Pouchain. Masson SA, p. 460.

Duclos, C et al. (2014). "Medical Vocabulary, Terminological Resources and Information Coding in the Health Domain". In: *Medical Informatics, e-Health*. Springer, pp. 11–41.

Dunikowski, L G and W W Weston (1986). "A library for family physicians." In: *Canadian family physician Médecin de famille canadien* 32, pp. 2226–

40. ISSN: 0008-350X. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2328226/.

Dunikowsky (1992). "FAMLI Family Medicine Literature Index". In: *The College of Family Physician, Canada. The National Library of Medicine* 13. Ed. by Lynn Dunikowsky, p. 12.

Eberbach, Andreas et al. (2016). "A simple heuristic for Internet-based evidence search in primary care: a randomized controlled trial". In: *Advances in Medical Education and Practice*, pp. 433–441. ISSN: 1179-7258. DOI: 10.2147/AMEP.S78385.

Eco, Umberto (1979). *The Role of the Reader: Explorations in the Semiotics of Texts*. Idiana University press.

Eco, Umberto. and Alastair. McEwen (2001). *Experiences in translation*. University of Toronto Press, p. 135. ISBN: 9780802096142. URL: http://www.utppublishing.com/Experiences-in-Translation.html.

Elish, M. C. and Danah Boyd (2017). "Situating methods in the magic of Big Data and AI". In: *Communication Monographs*, pp. 1–24. DOI: 10.1080/03637751.2017.1375130.

Elo, Satu and Helvi Kyngäs (2008). "The qualitative content analysis process." In: *Journal of advanced nursing* 62.1, pp. 107–15. ISSN: 1365-2648. DOI: 10.1111/j.1365-2648.2007.04569.x.

Ermilov, Ivan et al. (2016). "LODStats: The data web census dataset". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9982 LNCS, pp. 38–46. ISSN: 16113349. DOI: 10.1007/978-3-319-46547-0_5.

Eysenbach, Gunther (2006). "The open access advantage." In: *Journal of medical Internet research [electronic resource].* 8.2. ISSN: 14388871. DOI: 10.2196/jmir.8.2.e8.

Faggiolani, Chiara (2011). *Perceived Identity: applying Grounded Theory in Libraries*. it. DOI: 10.4403/jlis.it-4592.

Fernandez, Mariano, Asunción Gomez-Perez, and Natalia Juristo (1997). "METHONTOLOGY: From Ontological Art Towards Ontological Engineering". In: *Proceedings of the AAAI97 Spring Symposium. Series on Ontological Engineering*. Stanford, USA, pp. 33–40. URL: http://oa.upm.es/5484/1/METHONTOLOGY_.pdf.

Fernández-Breis, Jesualdo Tomás et al. (2013). "Leveraging electronic healthcare record standards and semantic web technologies for the identification of patient cohorts." In: *Journal of the American Medical Informatics Association : JAMIA* 20.e2, e288–96. ISSN: 1527-974X. DOI: 10.1136/amiajnl-2013-001923.

Fernandez-Lopez, M. et al. (1999). "Building a Chemical Ontology Using Methonology and the Ontology Design Environment". In: *IEEE Intelligent Systems* 14.1, pp. 37–46.

Fernández-López, Mariano and Asunción Gómez-Pérez (2002). "Overview and analysis of methodologies for building ontologies". In: *The Knowledge Engineering Review* 17.02, pp. 129–156.

Ferreras Fernández, Tránsito (2016). "Visibilidad e impacto de la literatura gris científica en repositorios institucionales de acceso abierto. Estudio de caso bibliométrico del repositorio Gredos de la Universidad de Salamanca". PhD thesis. URL: https://gredos.usal.es/jspui/handle/10366/132444.

Fickweiler, Freek, Ward Fickweiler, and Ewout Urbach (2017). "Interactions between physicians and the pharmaceutical industry generally and sales representatives specifically and their association with physicians' attitudes and prescribing habits: a systematic review". In: *BMJ open* 7.9, e016408.

Fitzgerald, Dorothy (1980). "Searching The Medical Literature". In: *Can Fam Physician* 26.October, pp. 1386–1391. URL: http://www.ncbi.nlm.nih.gov/pubmed/21293706.

Friedman, Carol et al. (1999). "Representing Information in Patient Reports Using Natural Language Processing and the Extensible Markup Language". In: *Journal of the American Medical Informatics Association* 6.1, pp. 76–87. URL: http://www.jamia.org/cgi/content/abstract/6/1/76.

Funk, M E and C A Reid (1983). "Indexing consistency in MEDLINE." In: *Bulletin of the Medical Library Association* 71.2, pp. 176–83. ISSN: 0025-7338. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC227138/.

Garshol, Lars Marius (2004). "Metadata? Thesauri? Taxonomies? Topic Maps! Making sense of it all". In: *Journal of Information Science* 30, pp. 378–391.

Gelly, J. et al. (2014). "Profile of French general practitioners providing opportunistic primary preventive care–an observational cross-sectional multicentre study". In: *Family Practice* 31.4, pp. 445–452. URL: http://www.ncbi.nlm.nih.gov/pubmed/24925928.

Gill, P. J. et al. (2014). "Development of a search filter for identifying studies completed in primary care". In: *Fam Pract* 31.6, pp. 739–45. ISSN: 0263-2136. DOI: 10.1093/fampra/cmu066.

Giunchiglia, Fausto and Pavel Shvaiko (2003). "Semantic matching". In: *The Knowledge Engineering Review* 18.3, pp. 265–280.

Giunchiglia, Fausto and Ilya Zaihrayeu (2009). "Lightweight ontologies". In: *Encyclopedia of Database Systems*. Springer, pp. 1613–1619.

Giustini, Dean (2007). "Web 3.0 and medicine". In: *BMJ (Clinical research ed.)* 335.7633, pp. 1273–4. DOI: 10.1136/bmj.39428.494236.BE.

Glaser, Barney G. and Anselm L. Strauss (1999). *Discovery of Grounded Theory: Strategies for Qualitative Research - - Livres*. AldineTransaction, p. 284. ISBN: 0202302601.

Goggi, Sara et al. (2016). "A semantic engine for grey literature retrieval in the oceanography domain." In: *Grey Journal (TGJ)* 12.3.

González-González, A I et al. (2007). "Information Needs and Information-Seeking Behavior of Primary Care Physicians". In: *Ann Fam Med* 4, pp. 345–352. ISSN: 1544-1709 (Print)1544-1717 (Electronic). DOI: 10.1370/afm.681.

Gotzsche, Peter G (2013). *Deadly Medicines and Organised Crime: How Big Pharma Has Corrupted Healthcare*. Radcliffe Publishing Ltd.

Greenberg, S. J. and P. E. Gallagher (2009). "The great contribution: Index Medicus, Index-Catalogue, and IndexCat". In: *J Med Libr Assoc* 97.2, pp. 108–13. ISSN: 1536-5050 (Print)1558-9439 (Electronic). DOI: 10.3163/1536-5050.97.2.007.

Greenway, Tyler and Joseph S Ross (2017). "US drug marketing: how does promotion correspond with health value?" In: DOI: https://doi.org/10.1136/bmj.j1855.

Greshake, Bastian (2017). "Looking into Pandora's Box: The Content of Sci-Hub and its Usage". In: *F1000Research* 6.

GreyNet (2014). *Pisa Declaration on Policy Development for Grey Literature Resources*. URL: http://greyguiderep.isti.cnr.it/Pisadeclapdf/Pisa-Declaration-May-2014.pdf.

Griffon, N et al. (2014). "Evaluating alignment quality between iconic language and reference terminologies using similarity metrics". eng. In: *BMC Med Inform Decis Mak* 14, p. 17. DOI: 10.1186/1472-6947-14-17.

Griffon, Nicolas et al. (2011). "Evaluation of multi-terminology super-concepts for information retrieval". In: *Studies in Health Technology and Informatics* 169.January, pp. 492–496. ISSN: 09269630. DOI: 10.3233/978-1-60750-806-9-492.

Grosjean, Julien (2014). "Modélisation, réalisation et évaluation d'un portail dans le cadre de la Plateforme d'Indexation Régionale (PlaIR)". PhD thesis. Université de Rouen. URL: http://www.chu-rouen.fr/tibs/wp-content/uploads/th{\'{e}}se{_}Julien{_}Grosjean{_}final.pdf.

Grosjean, Julien et al. (2012). "Teaching medicine with a terminology/ontology portal." eng. In: *Studies in health technology and informatics* 180, pp. 949–53. URL: http://europepmc.org/abstract/MED/22874333.

Gruber, Thomas R (1993). *A Translation Approach to Portable Ontology Specifications by A Translation Approach to Portable Ontology Specifications*. Tech. rep. April, pp. 199–220. DOI: http://dx.doi.org/10.1006/knac.1993.1008.

Grüninger, Michael and Mark S Fox (1995). "Methodology for the design and evaluation of ontologies". In:

Gunnarsson, B L, Per Linell, and Bengt Nordberg (2014). *The Construction of Professional Discourse*. Routledge.

Gunter, Tracy D and Nicolas P Terry (2005). "The emergence of national electronic health record architectures in the United States and Australia: models, costs, and questions". In: *Journal of medical Internet research* 7.1.

Gusso, Gustavo and Marc Jamoulle (2016). "From Dictionary to Terminologies / Ontologies in General Practice/Family Medicine". In: *The World Book of Family Medicine – Iberoamericana Edition 2016*, Chapter 54 p 1–4. URL: http://hdl.handle.net/2268/202788.

Gusso, Gustavo and José Mauro Ceratti Lopes (2012). *Tratado de Medicina de Família e Comunidade: 2 Volumes: Princípios, Formação e Prática*, p. 2180. ISBN: 8536327979.

Gutierrez, Cecilia and Peter Scheid (2002). “The History of Family Medicine and Its Impact in US Health Care Delivery”. In: *AAFP Foundation*, pp. 1–31.

Gómez-Pérez, A., M. Fernández-López, and O. Corcho (2003). “Ontological Engineering and the Semantic Web”. URL: http://www.exa.unicen.edu.ar/escuelapav/cursos/corcho/01_introduction.pdf.

Hastings, Jane et al. (2017). *Emotion-ontology*. URL: https://github.com/jannahastings/emotion-ontology.

Heilman, J (2015). “Point-of-Care Information in Open Access: A Time to Sow?” In: *PLoS Medicine* 12.8, e1001870. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1001870.

Helman, Cecil G (2008). *Medical Anthropology*. Ashgate, p. 580. ISBN: 978-0-7546-2655-8.

Hepp, Martin and Jos de Bruijn (2007). “GenTax: A generic methodology for deriving OWL and RDF-S ontologies from hierarchical classifications, thesauri, and inconsistent taxonomies”. In: *The Semantic Web: Research and Applications*, pp. 129–144. ISSN: 3540726667. DOI: 10.1007/978-3-540-72667-8_11.

Hess, Dean (2004). “How to Write an Effective Discussion”. In: *RESPIRATORY CARE* 49.10, pp. 1238–1241. ISSN: 1438-471X. DOI: 10.1365/s35131-011-0274-y.

Heyrman, J (ed.) (2005). *EURACT Educational Agenda, European Academy of Teachers in General Practice EURACT,* Leuven.

Hina, Saman, Eric Atwell, and Johnson Owell (2011). “Semantic Tagging of Medical Narratives with Top Level Concepts from SNOMED CT Healthcare Data Standard”. In: *International Journal of Intelligent Computing Research* 2.204-210.

Hoffmann, Kathryn et al. (2011). “Antibiotic resistance in primary care in Austria - a systematic review of scientific and grey literature”. In: *BMC Infectious Diseases* 11.1, p. 330. ISSN: 1471-2334. DOI: 10.1186/1471-2334-11-330.

Holden, Ronald B. (2010). “Face validity”. In: *The Corsini encyclopedia of psychology*. Ed. by Irving B. Weiner and W. Edward. Craighead. Wiley, pp. 637–638. ISBN: 9780470170267.

Hong, Y. et al. (2016). “Knowledge structure and theme trends analysis on general practitioner research: A Co-word perspective”. In: *BMC Fam Pract* 17, p. 10. ISSN: 1471-2296. DOI: 10.1186/s12875-016-0403-5.

Hopewell, S et al. (2007). “Grey literature in meta-analyses of randomized trials of health care interventions.” In: *Cochrane database of systematic reviews (Online)* 2, MR000010. DOI: 10.1002/14651858.MR000010.pub3.

Horton, Richard and Selina Lo (2015). “Planetary health: a new science for exceptional action.” In: *Lancet (London, England)* 386.10007, pp. 1921–2. ISSN: 1474-547X. DOI: 10.1016/S0140-6736(15)61038-8.

Huang, Minlie, Aurélie Névéol, and Zhiyong Lu (2011). "Recommending MeSH terms for annotating biomedical articles." In: *Journal of the American Medical Informatics Association : JAMIA* 18.5, pp. 660–7. ISSN: 1527-974X. DOI: 10.1136/amiajnl-2010-000055.

Hubbard, Derek (2008). *How to Find Clinical Information Quickly at the Point of Care*. Vol. 15. 6. American Academy of Family Physicians, p. 23.

Hummers-pradier, Eva (2007). "Which Abstracts Do Get Published ? – Output Of German Gp Research 1999-2003". In: *Wonca Europe Paris 2007*.

Ioannidis, John et al. (2017). "How to survive the medical misinformation mess". In: *European Journal of Clinical Investigation*.

ISO (2011).

Ittoo, Ashwin and Gosse Bouma (2013a). "Minimally-supervised learning of domain-specific causal relations using an open-domain corpus as knowledge base". In: *Data & Knowledge Engineering* 88, pp. 142–163.

— (2013b). "Term extraction from sparse, ungrammatical domain-specific documents". In: *Expert Systems with Applications* 40, pp. 2530–2540.

James, Jack E (2016). "Free-to-publish, free-to-read, or both? Cost, equality of access, and integrity in science publishing". In: *Journal of the Association for Information Science and Technology* 68.6, pp. 1584–1589. ISSN: 2330-1643. DOI: 10.1002/asi.23757.

Jamoulle, M et al. (2014). "Mapping French terms in a Belgian guideline on heart failure to international classifications and nomenclatures: the devil is in the detail". eng. In: *Inform Prim Care* 21.4, pp. 189–198. DOI: 10.14236/jhi.v21i4.66.

Jamoulle, Marc (1986). "Information et informatisation en médecine générale [Computer and computerisation in general practice]". In: *Les informa-g-iciens*. Namur, Belgium: Presses Universitaires de Namur, pp. 193–209. URL: http://orbi.ulg.ac.be/handle/2268/170822.

— (1998). "ICPC-2. The International Classification of Primary Care, an introduction". In: URL: http://orbi.ulg.ac.be/handle/2268/195813.

— (2005). "Attempt to classify main descriptors of GP/FM job. Proposal for a metaclinical classification". In: *WICC meeting. Dunedin. New Zealand*, pdf. URL: http://orbi.ulg.ac.be/handle/2268/195853.

— (2010). *Some views about SNOMED-CT by a General Practitioner*. URL: http://hdl.handle.net/2268/173908.

— (2013). "Using the International Classification for Primary Care (ICPC) and the Core Content Classification for General Practice (3CGP) to classify conference abstracts. Letter." en. In: *Rev Port Med Geral Fam* 29.5, pp. 66–67. URL: http://orbi.ulg.ac.be/handle/2268/171601.

— (2015a). "Quaternary prevention, an answer of family doctors to overmedicalization". In: *International Journal of Health Policy and Management* 4.2, pp. 61–64. ISSN: 2322-5939. DOI: 10.15171/ijhpm.2015.24. URL: http://ijhpm.com/article{_}2950{_}0.html.

— (2015b). "Une étude bibliographique avec 3CGP (ICPC+Q-Codes)". fr. In: *Cisp club Rouen 2015*. URL: http://orbi.ulg.ac.be/handle/2268/186926.

Jamoulle, Marc (2016). "New 2016 MeSH addressing information gap, poverty, violence and danger of medicine set the tone for policy makers in patient care". en. In: *International Journal of Health Policy and Management* 5.6, pp. 397–398. URL: http://orbi.ulg.ac.be/handle/2268/195481.

Jamoulle, Marc and Christophe Berkhout (2014). "Analyse de contenu et indexation des résumés de communications des congrès CNGE Clermont 2013 et Lille 2014". In: *14eme congres de la Confederation Nationale des Generalistes Enseignant*, (pdf 14 slides). URL: http://hdl.handle.net/2268/179603.

Jamoulle, Marc and Dominique Dekeuster (2007). "Workshop : Towards an e-archive for Wonca documents". en. In: *WONCA Europe Conference 2007, Paris, France*. URL: http://orbi.ulg.ac.be/handle/2268/187135.

Jamoulle, Marc and Melissa P Resnick (2016). *General Practice / Family Medicine Multilingual Terminology*. Care Editions, Jumet, Belgique, p. 66.

Jamoulle, Marc et al. (1999). "Origine de la CISP et mise en application actuelle dans les pays francophones". In: *L'informatisation du cabinet médical du futur*. Ed. by A Venot and Hector Falcoff. Paris: Informatique et Santé. Springer-Verlag, pp. 201–212. URL: http://hdl.handle.net/2268/175668.

Jamoulle, Marc et al. (2015a). "International Classification of Primary Care in a cross-lingual terminology portal". In: *4° Congreso Iberoamericano de Medicina Familiar y Comunitaria. Montevideo March 19-21 2015*. Montevideo, Poster. URL: http://hdl.handle.net/2268/179499.

Jamoulle, Marc et al. (2015b). "Semantic Web and the Future of Health Care Data in Family Practice". In: *Merit Research Journal of Medicine and Medical Sciences* 3.12, pp. 586–594. URL: http://orbi.ulg.ac.be/handle/2268/189292.

Jamoulle, Marc et al. (2015c). "The words of prevention, part II: ten terms in the realm of quaternary prevention". In: *Revista Brasileira de Medicina de Família e Comunidade* 10.35, p. 1. ISSN: 2179-7994. DOI: 10.5712/rbmfc10(35)1063.

Jamoulle, Marc et al. (2017a). "A terminology in General Practice / Family Medicine to represent non-clinical aspects for various usages : the Q-Codes". In: *Medical Informatics Europe (MIE2017) Informatics for Health 2017* / April, pp. 1–5. URL: http://orbi.ulg.ac.be/handle/2268/206527.

Jamoulle, Marc et al. (2017c). "Analysis of definitions of General Practice/Family Medicine and Primary Health Care". In: *British Journal of General Practice - Open*, accepted. ISSN: 0960-1643. URL: http://orbi.ulg.ac.be/handle/2268/210049.

— (2017b). "Analysis of definitions of General Practice/Family Medicine and Primary Health Care". In: *British Journal of General Practice* accepted. URL: http://orbi.ulg.ac.be/handle/2268/210049.

Jamoulle, Marc et al. (2017d). "Indexing grey multilingual literature in General Practice in the era of Semantic Web". In: *The Grey Journal* accepted. URL: http://orbi.ulg.ac.be/handle/2268/210490.

Janamian, T et al. (2016). “Quality tools and resources to support organisational improvement integral to high-quality primary care: a systematic review of published and grey literature”. In: *Med J Aust* 204.7 Suppl, S22–8. ISSN: 0025-729x.

Jelercic, S. et al. (2010). “Assessment of publication output in the field of general practice and family medicine and by general practitioners and general practice institutions”. In: *Fam Pract* 27.5, pp. 582–9. ISSN: 0263-2136. DOI: 10.1093/fampra/cmq032.

Jenuwine, Elizabeth S and Judith A Floyd (2004). “Comparison of Medical Subject Headings and text-word searches in MEDLINE to retrieve studies on sleep in healthy individuals.” In: *Journal of the Medical Library Association : JMLA* 92.3, pp. 349–53. ISSN: 1536-5050. URL: http://www.ncbi.nlm.nih.gov/pubmed/15243641.

Jilka, S R et al. (2015). “"Nothing About Me Without Me": An Interpretative Review of Patient Accessible Electronic Health Records”. In: *J Med Internet Res* 17.6, e161. ISSN: 1438-8871. DOI: 10.2196/jmir.4446.

Johnson, R Burke (1997). “Examining the validity structure of qualitative research”. In: *Education* 118.2, p. 282.

Jonquet, Clement et al. (2016). “SIFR BioPortal : Un portail ouvert et générique d ’ ontologies et de terminologies biomédicales françaises au service de l ’ annotation sémantique”. In: *16e Journées Francophones d’Informatique Médicale(JFIM)* June.

Jupp, Simon et al. (2014). “The EBI RDF platform: linked open data for the life sciences.” In: *Bioinformatics (Oxford, England)* 30.9, pp. 1338–9. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btt765.

Khan, Nada F, Sian E Harrison, and Peter W Rose (2010). “Validity of diagnostic coding within the General Practice Research Database: a systematic review”. In: 60.572, e128–e136. ISSN: 0960-1643. DOI: 10.3399/bjgp10X483562.

Kochen, Michael M. (2012). *Allgemeinmedizin und Familienmedizin*. Thieme; Auflage: 4., vollständig überarbeitete und erweiterte Auflage, p. 652. ISBN: 3131413840.

Kruschinski, Carsten et al. (2010). “Themes and methods of research presented at European General Practice Research Network conferences.” In: *Family practice* 27.4, pp. 459–67. ISSN: 1460-2229. DOI: 10.1093/fampra/cmq023.

Kumar, Anand et al. (2004). “An ontological framework for the implementation of clinical guidelines in health care organizations”. In: *Studies in health technology and informatics* 102, pp. 95–107. URL: http://www.ncbi.nlm.nih.gov/pubmed/15853266.

Labonté, Ronald et al. (2014). “Is the Alma Ata vision of comprehensive primary health care viable? Findings from an international project.” In: *Global health action* 7, p. 24997. ISSN: 1654-9880. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4141965/.

Lagasse, Raphael et al. (2001). *European situation of the routine medical data collection and their use for the health monitoring (EURO-MED-DATA)*. Tech.

rep. European Commission Directorate General SanCo. URL: http://www.ulb.ac.be/esp/emd/Emd{_}rep.pdf.

Lakhani, Mayur K. (2003). *A Celebration of General Practice*. Radcliffe Publishing, p. 203. ISBN: 1857759230.

Lambe, Patrick (2007). *Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness*. Chandos Knowledge Management.

Lamberts, H (1991). "Episode-oriented epidemiology in family practice. The practical use of the international classification of primary care (ICPC) as illustrated in patients with headache". In: *Primary Care Research*. Ed. by P G Norton et al. Newbury Park, CA: Sage.

Lamberts, H and I Hofmans-Okkes (1996). "The core of computer based patient records in family practice. Episodes of care classified with ICPC". In: *International journal of biomedical computing* Special Is.1-2. Ed. by A Hasman, pp. 35–41.

Lamberts, H and M Wood (1987). *ICPC: International Classification of Primary Care*. Oxford: Oxford University Press. ISBN: 0 19 261633 1.

Lawrence, Amanda et al. (2014). "Where is the evidence: realising the value of grey literature for public policy and practice". In: *Australian Policy Online*. DOI: 10.4225/50/5580B1E02DAF9.

Leeuwenhorst European Working Party (1980). *Continuing Education and General Practitioners*, p. 20. ISBN: 8091016883.

Lehmann, Jens et al. (2015). "DBpedia - A large-scale, multilingual knowledge base extracted from Wikipedia". In: *Semantic Web* 6.2, pp. 167–195. ISSN: 22104968. DOI: 10.3233/SW-140134.

Lei Zeng, Marcia (2008). "Knowledge organization systems (KOS)". In: *Knowledge organization* 35.2-3, pp. 160–182.

Lelong, R et al. (2016). "Semantic Search Engine to Query into Electronic Health Records with a Multiple-Layer Query Language". In: *MEDIR workshop*. URL: http://medir2016.imag.fr/data/MEDIR_2016_paper_8.pdf.

Liang, S. F. et al. (2014). "Semi Automated Transformation to OWL Formatted Files as an Approach to Data Integration". In: *Methods of Information in Medicine* 54.1, pp. 32–40. ISSN: 0026-1270. DOI: 10.3414/ME13-02-0029.

Library and Archives Canada (2017). *Canadian Subject Headings*. URL: http://www.bac-lac.gc.ca/.

Lin, Jennifer and Carly Strasser (2014). "Recommendations for the role of publishers in access to data." In: *PLoS biology* 12.10, e1001975. ISSN: 1545-7885. DOI: 10.1371/journal.pbio.1001975.

Lipscomb, C E (2000). "Medical Subject Headings (MeSH)." In: *Bulletin of the Medical Library Association* 88.3, pp. 265–6. ISSN: 0025-7338. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC35238/.

Liu, Kaihong, William R Hogan, and Rebecca S Crowley (2011). "Natural Language Processing methods and systems for biomedical ontology learning." In: *Journal of biomedical informatics* 44.1, pp. 163–79. DOI: 10.1016/j.jbi.2010.07.006.

Liyanage, H. et al. (2013). "The Evidence-base for Using Ontologies and Semantic Integration Methodologies to Support Integrated Chronic Disease Management in Primary and Ambulatory Care: Realist Review. Contribution of the IMIA Primary Health Care Informatics WG". In: *Yearbook of medical informatics* 8.1, pp. 147–54. URL: http://www.ncbi.nlm.nih.gov/pubmed/23974562.

Liyanage, Harshana, Siaw Teng Liaw, and Simon De Lusignan (2012). "Accelerating the development of an information ecosystem in health care, by stimulating the growth of safe intermediate processing of health information (IPHI)". In: *Informatics in Primary Care* 20.2, pp. 81–86. ISSN: 14760320.

López, Fernández (1999). "Overview Of Methodologies For Building Ontologies". In: *Proceedings of the IJCAI99 Workshop on Ontologies and ProblemSolving Methods Lessons Learned and Future Trends CEUR Publications* 1999.2, pp. 1–13. ISSN: 02698889. DOI: 10.1.1.39.6002.

Lopez, Vanessa et al. (2013). "Evaluating question answering over linked data". In: *Web Semantics: Science, Services and Agents on the World Wide Web* 21, pp. 3–13. ISSN: 15708268. DOI: 10.1016/j.websem.2013.05.006. URL: http://linkinghub.elsevier.com/retrieve/pii/S157082681300022X.

LOV (2017). *Linked Open Vocabularies (LOV)*. URL: http://lov.okfn.org/dataset/lov/ (visited on 08/13/2017).

Lowe, H J and G O Barnett (1994). "Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches." In: *JAMA* 271.14, pp. 1103–8. ISSN: 0098-7484. URL: http://www.ncbi.nlm.nih.gov/pubmed/8151853.

Lusignan, Simon de (2005). "Codes, classifications, terminologies and nomenclatures: definition, development and application in practice". In: *Informatics in primary care* 13.1, pp. 65–70. ISSN: 1476-0320. URL: http://www.ncbi.nlm.nih.gov/pubmed/15949178.

Madkour, Mohcine, Driss Benhaddou, and Cui Tao (2016). "Temporal data representation, normalization, extraction, and reasoning: A review from clinical domain". In: *Computer Methods and Programs in Biomedicine*, pp. 52–68. DOI: 10.1016/j.cmpb.2016.02.007.

Mahood, Quenby, Dwayne Van Eerd, and Emma Irvin (2014). "Searching for grey literature for systematic reviews: challenges and benefits". In: *Research synthesis methods* 5.3, pp. 221–34. DOI: 10.1002/jrsm.1106.

Marc, D. T. et al. (2015). "Indexing Publicly Available Health Data with Medical Subject Headings (MeSH): An Evaluation of Term Coverage". In: *Stud Health Technol Inform* 216, pp. 529–33. URL: https://www.ncbi.nlm.nih.gov/pubmed/26262107.

Martin, Patricia and Barry A. Turner (2016). "Grounded Theory and Organizational Research". In: *http://dx.doi.org/10.1177/002188638602200207*. DOI: 10.1177_002188638602200207. URL: http://journals.sagepub.com/doi/abs/10.1177/002188638602200207.

McEntyre, J et al. (2002). "The NCBI Handbook [Internet]". In: *Bethesda, MD: National Center for Biotechnology Information. See http://www. ncbi. nlm. nih. gov/books/NBK21101*.

McGuinness, D. L. and Frank van Harmelen (2004). *OWL Web Ontology Language - W3C Recommendation*. Report. URL: http://www.w3.org/TR/owl-features/.

McIntyre, Ellen et al. (2016). "The contribution of a knowledge exchange organisation in primary healthcare." In: *Australian family physician* 45.9, pp. 684–7. ISSN: 0300-8495. URL: http://www.ncbi.nlm.nih.gov/pubmed/27606374.

McKenzie, Lindsay (2017). "Sci-Hub's cache of pirated papers is so big, subscription journals are doomed, data analyst suggests". In: *Science*. ISSN: 0036-8075. DOI: 10.1126/science.aan7164.

McMillan, Wendy J (2009). "Finding a method to analyze qualitative data: using a study of conceptual learning." In: *Journal of dental education* 73.1, pp. 53–64. ISSN: 0022-0337. URL: http://www.ncbi.nlm.nih.gov/pubmed/19126767.

McWhinney, Ian R. (1997). *A Textbook of Family Medicine*. Oxford University Press, p. 448. ISBN: 019511518X.

Mendis, Kumara and Indragit Solangaarachchi (2005). "PubMed perspective of family medicine research: Where does it stand?" In: *Family Practice* 22.5, pp. 570–575. ISSN: 02632136. DOI: 10.1093/fampra/cmi085.

Meneghini, Rogerio and Abel L Packer (2007). "Is there science beyond English? Initiatives to increase the quality and visibility of non-English publications might help to break down language barriers in scientific communication." en. In: *EMBO reports* 8.2, pp. 112–6. ISSN: 1469-221X. DOI: 10.1038/sj.embor.7400906.

Merabti, T, J Grosjean, and SJ Darmoni (2017). "MT@HeTOP: The Mapping and Translation tool based on the Health multi-terminology portal HeTOP". In: *ISHIC2017 2nd International Saudi Health Informatics Conference*. Riyadh, Kingdom of Saudi of Arabia.

Merabti, T, R Lelong, and SJ Darmoni (2015). "InfoRoute: the CISMeF Context-specific Search Algorithm". In: *Studies in Health Technology and Informatics MEDINFO*. Vol. 216, pp. 544–8.

Microsoft (2017). *Introduction to managed metadata in SharePoint Server 2010 - SharePoint*. URL: https://support.office.com.

Mihindukulasooriya, Nandana, Raúl García-Castro, and Asunción Gómez-Pérez (2017). "LD Sniffer: A Quality Assessment Tool for Measuring the Accessibility of Linked Data". In: Springer, Cham, pp. 149–152. DOI: 10.1007/978-3-319-58694-6_20.

Mil, J W van, R Brenninkmeijer, and T F Tromp (1998). "The ICPC coding system in pharmacy: developing a subset, ICPC-Ph." In: *Pharmacy world & science : PWS* 20.1, pp. 38–42. ISSN: 0928-1231. URL: http://www.ncbi.nlm.nih.gov/pubmed/9536470.

Miller, Eric (1998). "An Introduction to the Resource Description Framework". In: *Bulletin of the Association for Information Science and Technology* 25.1, pp. 15–19. ISSN: 1550-8366. DOI: 10.1002/bult.105.
Miller, William L et al. (2005a). "New Knowledge for and About Primary Care : A View Through the Looking Glass". In: 3.3, supll.data. DOI: 10.1370/afm.330.he.
Miller, William L et al. (2005b). "New knowledge for and about primary care: a view through the looking glass of the Annals of Family Medicine." In: *Annals of family medicine* 3.3, p. 197. ISSN: 1544-1717. DOI: 10.1370/AFM.330.
Missie, Thierry et al. (2015). *Développement D'une Application De Gestion Des Congrès Scientifiques. Rapport De Projet Technique.* Tech. rep. Groupe ESAIP – Enseignement supérieur – Projet Technique Chef de Projet Informatique 2014/2015. URL: http://3cgp.docpatient.net/q-codes-and-icpc-for-congress-organisation/.
Mittal, Sparsh and Ankush Mittal (2011). "Versatile question answering systems: seeing in synthesis". In: *International journal of intelligent information and database systems* 5.2, pp. 119–142.
Moffett, James (1967). *Teaching the Universe of Discourse.* en. Houghton Mifflin Company. URL: http://eric.ed.gov/?id=ED030664.
Moher, David et al. (2000). "Mejora de la calidad de los informes de los metaanálisis de los ensayos clínicos controlados: el acuerdo QUOROM". In: *Rev. Esp. Salud Publica* 74.2. ISSN: 1135-5727. URL: http://scielo.isciii.es/pdf/resp/v74n2/mejora.pdf.
Moynihan, R (2003). "Who pays for the pizza? Redefining the relationships between doctors and drug companies. 1: entanglement". In: *Bmj* 7400, pp. 1189–1192. ISSN: 0959-535x. DOI: 10.1136/bmj.326.7400.1189.
Moynihan, R and L Bero (2017). "Toward a Healthier Patient Voice: More Independence, Less Industry Funding". In: *JAMA Intern Med* 177.3, pp. 350–351. ISSN: 2168-6106. DOI: 10.1001/jamainternmed.2016.9179.
Murtagh, John (2011). *John Murtagh's General Practice*. McGraw-Hill Medical Publishing Division, p. 1508. ISBN: 0070285381.
Musen, Mark A. et al. (2012). "The National Center for Biomedical Ontology". In: *Journal of the American Medical Informatics Association : JAMIA* 19.2, pp. 190–5. DOI: 10.1136/amiajnl-2011-000523.
Myška, Matěj and Jaromír Šavelka (2013). "A Model Framework for Publishing Grey Literature in Open Access Defining Grey Literature". In: *Open Access* 4.JIPITEC 2, para 104, pp. 1–12.
Navigli, Roberto and Simone Paolo Ponzetto (2012). "BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network". In: *Artificial Intelligence*, pp. 217–250. ISSN: 00043702. DOI: 10.1016/j.artint.2012.07.001.
Neghme, A. (1975). "Operations of the Biblioteca Regional de Medicina (BIREME)". In: *Bull Med Libr Assoc* 63.2, pp. 173–9. ISSN: 0025-7338 (Print).
Nelson, Stuart J. (2009). "Medical terminologies that work: The example of MeSH". In: *I-SPAN 2009 - The 10th International Symposium on Pervasive*

Systems, Algorithms, and Networks, pp. 380–384. DOI: 10.1109/I-SPAN.2009.84.

Neveol, A, K Zeng, and O Bodenreider (2006). "Besides precision & recall: exploring alternative approaches to evaluating an automatic indexing tool for MEDLINE." In: *AMIA Annu Symp Proc*, pp. 589–593. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1839480.

Névéol, Aurélie, Rezarta Islamaj Doğan, and Zhiyong Lu (2010). "Author keywords in biomedical journal articles". In: *AMIA Annual Symposium Proceedings*. Vol. 2010. American Medical Informatics Association, p. 537.

Newbold, Elizabeth and Jennie Grimshaw (2010). "Collection building with special regards to Report Literature". In: *Grey Literature in Library and Information Studies*. Ed. by Dominic John Farace and Joachim Schöpfel. Walter de Gruyter, p. 63.

NISO (2005). *ANSI/NISO Z39.19-2005: Guidelines for the Construction , Format , and Management of Monolingual Controlled Vocabularies*. Vol. 2003, pp. 1–184. ISBN: 1880124653. arXiv: 1007.5481v1.

NLM (2017). "UMLS Glossary". In: URL: https://www.nlm.nih.gov/research/umls/new_users/glossary.html.

Noble, Helen and Joanna Smith (2015). "Issues of validity and reliability in qualitative research". In: *Evidence-Based Nursing* 18.2, pp. 34–35. ISSN: 1367-6539. DOI: 10.1136/eb-2015-102054. URL: http://ebn.bmj.com/cgi/doi/10.1136/eb-2015-102054.

Oandasan, Ivy F., Danielle Saucier, and (eds) (2013). *Triple C Competency-based Curriculum Report – Part 2 Advancing Implementation*. College of Family Physicians of Canada, p. 164. URL: www.cfpc.ca/uploadedFiles/Education/{_}PDFs/TripleC{_}Report{_}pt2.pdf..

Okkes, I. et al. (2000). "ICPC-2-E: the electronic version of ICPC-2. Differences from the printed version and the consequences." en. In: *Family practice* 17.2, pp. 101–7. ISSN: 0263-2136. URL: http://www.ncbi.nlm.nih.gov/pubmed/10758069.

PAHO Bireme Sao Paulo (2016). *Virtual Health Library Search Portal of the Regional library of Medicine*. URL: http://regional.bvsalud.org/http://www.paho.org/bireme/.

Parkerson, G R et al. (1996). "Classification of severity of health problems in family/general practice: an international field trial." In: *Family practice* 13.3, pp. 303–9. ISSN: 0263-2136. URL: http://www.ncbi.nlm.nih.gov/pubmed/8671140.

Pellizzon, Rosely de Fátima (2004). "Pesquisa na área da saúde: 1. Base de dados DeCS (Descritores em Ciências da Saúde)". In: *Acta Cirurgica Brasileira* 19, pp. 153–163. URL: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-86502004000200013&nrm=iso.

Pereira, Teresa Avalos et al. (2012). "Terminologia DeCS e as novas regras ortográficas da língua portuguesa: orientações para uma atualização DeCS terminology and the new rules on orthography of Portuguese language: guidelines for an update". In: *Acta Cirúrgica Brasileira* 27.7, p. 509. URL: http://www.scielo.br/pdf/acb/v27n7/a14v27n7.

Pierce, Christopher (2014). "SemanticDB: A Semantic Web Infrastructure for Clinical Research and Quality Reporting". In: *Current Bioinformatics* 7.3. URL: http://benthamscience.com/journal/abstracts.php?journalID=cbio&articleID=101604.

Pingitore, D and R A Sansone (1998). "Using DSM-IV primary care version: a guide to psychiatric diagnosis in primary care." In: *American family physician* 58.6, pp. 1347–52. ISSN: 0002-838X. URL: http://www.ncbi.nlm.nih.gov/pubmed/9803199.

Pizzanelli, Miguel (2015). "¿Overscreening o prevención a escala humana? Tamizaje excesivo". In: *Revista Brasileira de Medicina de Família e Comunidade* 10.35, p. 1. ISSN: 2179-7994. DOI: 10.5712/rbmfc10(35)1068. URL: http://www.rbmfc.org.br/rbmfc/article/view/1068.

PLoS Medicine Editors (2015). "Point-of-Care Information in Open Access: A Time to Sow?" In: *PLoS medicine* 12.8, e1001870. ISSN: 1549-1676. DOI: 10.1371/journal.pmed.1001870.

Pols, David H J et al. (2015). "Development and Validation of Search Filters to Identify Articles on Family Medicine in Online Medical Databases." In: *Annals of family medicine* 13.4, pp. 364–6. ISSN: 1544-1717. DOI: 10.1370/afm.1780.

Pope, C, S Ziebland, and N Mays (2000). "Qualitative research in health care. Analysing qualitative data." In: *BMJ (Clinical research ed.)* 320.7227, pp. 114–6. ISSN: 0959-8138. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1117368/.

Qamar, Rahil (2008). "Semantic mapping of clinical model data to biomedical terminologies to facilitate interoperability". Thesis. URL: http://www.openehr.org/publications/archetypes/RQamar-PhD-Thesis-Mar2008.pdf.

Quan, Wei, Bikun Chen, and Fei Shu (2017). "Publish or impoverish: An investigation of the monetary reward system of science in China (1999-2016)". In: *ArXiv e-prints*. arXiv: 1707.01162. URL: http://arxiv.org/abs/1707.01162.

Rafaeli, A and M Worline (1999). "Symbol in organizational culture". In: *Handbook of organizational culture and climate*. iem.technion.ac.il.

Ralph, N., M. Birks, and Y. Chapman (2014). "Contextual Positioning: Using Documents as Extant Data in Grounded Theory Research". In: *SAGE Open* 4.3. DOI: 10.1177/2158244014552425.

Rayner, Jennifer (2009). *Review of ENCODE-FM Nomenclature and Terminology System use by Ontario Community Health Centres*. Tech. rep. March. Ontario CHCs' Performance Measurement Committee, pp. 1–6.

RCGP (1973). "A general-practice glossary". In: *Journal of the Royal College of General Practitioners* 3.Suppl n°3, pp. 15–15. URL: http://trove.nla.gov.au/work/21584373?selectedversion=NBD332424.

Resnick, Melissa P et al. (2013). "Representing second opinion requests from primary care within the Brazilian tele-health program: international classification of primary care, second edition." In: *Studies in health technology* *and informatics* 192, p. 1190. ISSN: 0926-9630. URL: http://www.ncbi.nlm.nih.gov/pubmed/23920964.

Richards, IA and CK Ogden (1930). *The meaning of meaning*. Kegan Paul Trench Trubner And Co Ltd.

Rigaux, Sébastien (2015). "Classification automatisée de résumés médicaux – Belgique 2015 [Multi-Label Text Classification of Medical Abstracts]". Master thesis.

Roland, Michel, Marianne Prevost, and Marc Jamoulle (2001). "L'assurance de qualité et le médecin de famille." In: *Archives of Public Health* 59, pp. 1–28. URL: http://hdl.handle.net/2268/173932.

Roland, Michel, Marianne Prevot, and Marc Jamoulle (2002). "L'assurance de qualité : une méthode, des concepts, une pratique - Michel Roland, Marianne Prévost et Marc Jamoulle - 15p". In: *Santé conjuguée* 21, p. 15.

Rosemann, Thomas et al. (2006). "Referrals from general practice to consultants in Germany: If the GP is the initiator, patients' experiences are more positive". In: *BMC Health Services Research* 6.1, p. 5. ISSN: 1472-6963. DOI: 10.1186/1472-6963-6-5.

Rosenbloom, S Trent et al. (2006). "Interface terminologies: facilitating direct entry of clinical data into electronic health record systems". In: *Journal of the American Medical Informatics Association* 13.3, pp. 277–288.

Roski, Joachim, George W. Bo-Linn, and Timothy A. Andrews (2014). "Creating Value In Health Care Through Big Data: Opportunities And Policy Implications". In: DOI: 10.1377/hlthaff.2014.0147.

Roumier, Joseph et al. (2011). "Approach to the Creation of a Multilingual, Medical Interface Terminology". In: *Proceedings of the 9th International Conference on Terminology and Artificial Intelligence*, pp. 13–15. URL: http://hal.inria.fr/hal-00646223_v1/.

Sadegh-Zadeh, Kazem (2015). *Handbook of Analytic Philosophy of Medicine*. 2nd. Dordrecht: Springer, p. 777.

Salman Bin Naeem, Salman, Ahmed Shamshad, and Khan Amjid (2013). "Information seeking in primary care: a survey of doctors working in remote government health facilities in Pakistan". In: *Library Philosophy and Practice (e-journal)* Paper 1009. URL: http://digitalcommons.unl.edu/libphilprac/1009.

Salvadores, M. and M. Horridge (2012). "Using SPARQL to query bioportal ontologies and metadata". In: *The Semantic Web–...* URL: http://link.springer.com/chapter/10.1007/978-3-642-35173-0_12.

Samwald, Matthias et al. (2011). "Linked open drug data for pharmaceutical research and development." In: *Journal of cheminformatics* 3.1, p. 19. DOI: 10.1186/1758-2946-3-19.

Schmidt, M et al. (2015). "The Danish National Patient Registry: a review of content, data quality, and research potential". In: *Clin Epidemiol* 7, pp. 449–490. ISSN: 1179-1349. DOI: 10.2147/clep.s91125.

Schneeweiss, R et al. (1977). "A conversion code from the RCGP to the ICH-PPC classification system". In: *J.Fam Pract.* 5.3, pp. 415–424.

Schöpfel, Joachim (2015). “Littérature « grise » : de l’ombre à la lumière”. In: *I2D – Information, données & documents*. Vol. Volume 52. 1. Chap. Introduction, pp. 28–29. URL: https://www.cairn.info/revue-i2d-information-donnees-et-documents-2015-1-page-28.htm.

Schöpfel, Joachim and Dominic J Farace (2010). “Grey literature”. In:

Schrans, D. et al. (2016). “The search for person-related information in general practice: a qualitative study”. In: *Fam Pract* 33.1, pp. 95–9. ISSN: 0263-2136. DOI: 10.1093/fampra/cmv099.

Schuers, Matthieu et al. (2015). “Mise en ligne de la CISP en près de 20 langues au sein d’un portail terminologique de santé”. In: *Congrès de la Médecine Générale, Paris, 26 et 27 mars 2015*, poster. URL: http://orbi.ulg.ac.be/handle/2268/207421.

Schwitzer, Gary (2017). *Conflicts of interest in health care journalism. Who’s watching the watchdogs? We are. Part 1 of 3 - HealthNewsReview.org*. URL: https://www.healthnewsreview.org/2017/06/conflicts-of-interest-in-health-care-journalism-1-of-3/.

Sfakianaki, Pepi et al. (2015). “Semantic biomedical resource discovery: a Natural Language Processing framework.” en. In: *BMC medical informatics and decision making* 15.1, p. 77. ISSN: 1472-6947. DOI: 10.1186/s12911-015-0200-4.

Shannon, C E (1958). “A Mathematical Theory of Communication”. In: *The Bell Technical Journal* 27.1, pp. 379–423, 623–656. URL: http://lanethames.com/dataStore/ECE/InfoTheory/shannon.pdf.

Shannon, Claude E (1951). “Prediction and entropy of printed English”. In: *Bell Labs Technical Journal* 30.1, pp. 50–64.

Shen, Cenyu and Bo-Christer Björk (2015). “‘Predatory’ open access: a longitudinal study of article volumes and market characteristics”. In: *BMC Medicine* 13.1, p. 230. ISSN: 1741-7015. DOI: 10.1186/s12916-015-0469-2.

Shukor, Ali Rafik (2017). “An Alternative Paradigm for Evidence-Based Medicine : Revisiting Lawrence Weed , MD ’ s Systems Approach”. In: *The Permanente Journal* 21.16, pp. 147–147.

Shultz, M. (2007). “Comparing test searches in PubMed and Google Scholar”. In: *J Med Libr Assoc* 95.4, pp. 442–5. ISSN: 1536-5050 (Print)1558-9439 (Electronic). DOI: 10.3163/1536-5050.95.4.442.

Silva, Caio da, Regina Garcia, and Rita Bonadio Inacio de Cássia (2009). “Literatura Cinzenta : teses , eventos e relatórios”. Thesis. URL: http://rabci.org.

Silver, Christina and Ann Lewins (2014). *Using Software in Qualitative Research*. SAGE Companion. URL: https://study.sagepub.com/using-software-in-qualitative-research.

Simon, C. (2009). “From generalism to specialty–a short history of General Practice”. In: *InnovAiT* 2.1, pp. 2–9. ISSN: 1755-7380. DOI: 10.1093/innovait/inn171.

Singh, Punam Mony et al. (2007). “Language preferences on websites and in Google searches for human health and food information”. In: *Journal of medical Internet research* 9.2, e18–e18. DOI: 10.2196/jmir.9.2.e18.

Sladek, Ruth et al. (2006). "Development of a subject search filter to find information relevant to palliative care in the general medical literature." In: *Journal of the Medical Library Association : JMLA* 94.4, pp. 394–401. ISSN: 1558-9439. URL: http://www.ncbi.nlm.nih.gov/pubmed/17082830.

Smith, Barry and Christiane Fellbaum (2004). "Medical WordNet : A New Methodology for the Construction and Validation of Information Resources for Consumer Health". In: *Proceedings of Coling: The 20th International Conference on Computational Linguistics, Geneva*. August, pp. 31–38.

Smith, Barry et al. (2006). "Towards a reference terminology for ontology research and development in the biomedical domain." In: *KR-MED*. Vol. 2006, pp. 57–66.

Smith, Barry et al. (2007a). "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration." In: *Nature biotechnology* 25.11, pp. 1251–5. ISSN: 1087-0156. DOI: 10.1038/nbt1346. URL: http://dx.doi.org/10.1038/nbt1346.

Smith, Barry et al. (2007b). "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration". In: *Nature biotechnology* 25.11, pp. 1251–5. DOI: 10.1038/nbt1346.

Smith, Robert Rutherford (1980). "Semiotics and Communication Theory". en. In: *Journal of Communication* 30.1, pp. 205–210. DOI: 10.1111/j.1460-2466.1980.tb01788.x.

Soiferman, L. Karen (2010). *Compare and Contrast Inductive and Deductive Research Approaches*. Tech. rep. University of Manitoba. URL: http://files.eric.ed.gov/fulltext/ED542066.pdf.

Soler, Jean karl, Marc Jamoulle, and Peter Schattner (2015). "The International Classification of Primary Care". en. In: *The World Book of Family Medicine – European Edition 2015*. Ljubljana. ISBN: 978-961-281-983-5. URL: http://orbi.ulg.ac.be/handle/2268/187050.

Soler, Jean K. et al. (2012). "The interpretation of the reasons for encounter 'cough' and 'sadness' in four international family medicine populations". In: *Informatics in Primary Care* 20, pp. 25–39. ISSN: 14760320.

Soler, Jean-Karl K et al. (2008). "The coming of age of ICPC: celebrating the 21st birthday of the International Classification of Primary Care". In: *Family Practice* 25.4, pp. 312–317. URL: http://www.ncbi.nlm.nih.gov/pubmed/18562335.

Sommerhalder, K et al. (2009). "Internet information and medical consultations: experiences from patients' and physicians' perspectives". In: *Patient Educ Couns* 2, pp. 266–271. ISSN: 0738-3991. DOI: 10.1016/j.pec.2009.03.028.

Starfield, B. (2008). "Access, primary care, and the medical home: rights of passage". In: *Med Care* 46.10, pp. 1015–6. ISSN: 0025-7079. DOI: 10.1097/MLR.0b013e31817fae3c.

Starfield, B., L. Shi, and J. Macinko (2005). "Contribution of primary care to health systems and health". In: *Milbank Q* 83.3, pp. 457–502. ISSN: 0887-378X (Print) 0887-378x. DOI: 10.1111/j.1468-0009.2005.00409.x.

Starfield, Barbara (2010). "Reinventing Primary Care: Lessons From Canada For The United States". In: DOI: 10.1377/hlthaff.2010.0002.

Sutton, Stuart A (2007). *Tutorial 1: Basic Semantics*. URL: http://dublincore.org/resources/training/dc-2007/T1-BasicSemantics.pdf.

Swan, Alma. (2012). *Policy guidelines for the development and promotion of open access*. United Nations Educational, Scientific, and Cultural Organization, p. 76. ISBN: 9789230010522.

Swartout, B. et al. (1997). "Toward distributed use of large-scale ontologies". In: ed. by A. Farquhar et al. California: Stanford University, p. 138–48.

Tabatabaei-Malazy, O, S Nedjat, and R Majdzadeh (2012). "Which information resources are used by general practitioners for updating knowledge regarding diabetes?" In: *Arch Iran Med* 4, pp. 223–227. ISSN: 1029-2977. DOI: 012154/aim.0010.

Tan, Sharon Swee-Lin and Nadee Goonawardene (2017). "Internet Health Information Seeking and the Patient-Physician Relationship: A Systematic Review". In: *Journal of Medical Internet Research* 19.1, e9. ISSN: 1438-8871. DOI: 10.2196/jmir.5729.

Tao, Cui et al. (2013). "Terminology representation guidelines for biomedical ontologies in the semantic web notations." In: *Journal of biomedical informatics* 46.1, pp. 128–38. ISSN: 1532-0480. DOI: 10.1016/j.jbi.2012.09.003.

Testern, Cecilie D, Lise Hestbæk, and Simon D French (2015). "The use of diagnostic coding in chiropractic practice." In: *Chiropractic & manual therapies*, p. 8. ISSN: 2045-709X. DOI: 10.1186/s12998-015-0051-1.

Tharmalingam, Sukirtha, Simon Hagens, and Jennifer Zelmer (2016). "The value of connected health information: perceptions of electronic health record users in Canada". In: *BMC Medical Informatics and Decision Making* 16.1, p. 93. ISSN: 1472-6947. DOI: 10.1186/s12911-016-0330-3.

Thirion, B, S Pereira, and A Névéol (2006). "French MeSH Browser: a cross-language tool to access MEDLINE/PubMed." In: *AMIA... Annual ...* URL: http://europepmc.org/abstract/med/18694229.

Thirion, B, I Robu, and SJ Darmoni (2009). "Optimization of the PubMed Automatic Term Mapping". In: *MIE2009*. Vol. 150, pp. 238–42. DOI: 10.3233/978-1-60750-044-5-238.

Thompson, C. A. et al. (2014). "Patient and provider characteristics associated with colorectal, breast, and cervical cancer screening among Asian Americans". In: *Cancer Epidemiol Biomarkers Prev*. Vol. 23. United States: (c)2014 American Association for Cancer Research., pp. 2208–17. ISBN: 1538-7755 (Electronic)1055-9965 (Linking). DOI: 10.1158/1055-9965.epi-14-0487.

Troeung, L. et al. (2016). "A new electronic screening tool for identifying risk of familial hypercholesterolaemia in general practice". In: *Heart* 102.11, pp. 855–61. ISSN: 1355-6037. DOI: 10.1136/heartjnl-2015-308824.

Tudorache, Tania et al. (2010). "Supporting the Collaborative Authoring of ICD-11 with WebProtégé". In: *AMIA ... Annual Symposium Proceedings /*

AMIA Symposium. AMIA Symposium 2010, pp. 802–806. ISSN: 1942-597X. URL: http://www.ncbi.nlm.nih.gov/pubmed/21347089.

Tudorache, Tania et al. (2013). "WebProtégé: A collaborative ontology editor and knowledge acquisition tool for the web". In: *Semantic web* 4.1, pp. 89–99.

Unger, C, A Freitas, and P Cimiano (2014). "An Introduction to Question Answering over Linked Data". In: *Reasoning Web: Reasoning on the Web in the Big Data Era* 8714, pp. 100–140. ISSN: 16113349. DOI: 10.1007/978-3-319-10587-1_2.

Uschold, Michael (2004). "Ontologies and Semantics for Seamless Connectivity ‡ 1 Introduction". In: *SIGMOD Record*, 33.4, pp. 58–64.

Uschold, Mike (2006). "Ontologies Ontologies Everywhere – but Who Knows What to Think?" In: *Protégé Users Conference*, Keynote address / power–point. URL: http://protege.stanford.edu/conference/2006/submissions/slides/1.2{_}Uschold.pdf.

Uschold, Mike and Michael Gruninger (1996). "Ontologies: Principles, methods and applications". In: *Knowledge Engineering Review* 11.2, pp. 93–136. DOI: 10.1.1.111.5903.

Uschold, Mike and Martin King (1995). "Towards a Methodology for Building Ontologies". In: *Methodology* 80.July, pp. 275–280. ISSN: 02698889. DOI: 10.1.1.55.5357.

Ustün, T B et al. (1995). "New classification for mental disorders with management guidelines for use in primary care: ICD-10 PHC chapter five." In: *The British journal of general practice : the journal of the Royal College of General Practitioners* 45.393, pp. 211–5. ISSN: 0960-1643. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1239204/.

Van Dormael, M. (2001). "Identités professionnelles en médecine générale et en soins de santé primaires.[Professional identity in GP/FM and in Primary Care]". In: ed. by B. Hours. Karthala Médecines du Monde, Chap VI–Chap VI.

Van Royen, Paul et al. (2010). "Are presentations of abstracts at EGPRN meetings followed by publication?" EN. In: *The European journal of general practice* 16.2, pp. 100–5. ISSN: 1751-1402. DOI: 10.3109/13814788.2010.482582.

vanBoven, K et al. (2011). "Do unexplained symptoms predict anxiety or depression? Ten-year data from a practice-based research network". In: *Br J Gen Pract* 587, e316–25. DOI: 10.3399/bjgp11X577981.

Vandenbussche, Pierre-Yves and Bernard Vatant (2011). "Metadata recommendations for linked open data vocabularies". In: *Version* 1, pp. 2011–12.

Vander Stichele, Robert H et al. (1991). "Attitude of the Public toward Technical Package Inserts for Medication Information in Belgium". en. In: *Annals of Pharmacotherapy*. DOI: 10.1177_106002809102500916.

Vanmeerbeek, M, F Felgueroso-Bueno, and JB Lafontaine (2016). "Guide de rédaction du TFE en médecine générale". In: URL: http://orbi.ulg.ac.be/handle/2268/202592.

Vanmeerbeek, Marc et al. (2015). "Prévention et promotion de la santé: quels modèles pour soutenir l'évolution des pratiques en soins de santé primaires?" In: *La Presse Médicale* 44.6, e211–e220.

VanNieuwenborg, L et al. (2016). "Continuing medical education for general practitioners: a practice format". In: *Postgrad Med J* 92.1086, pp. 217–222. ISSN: 0032-5473. DOI: 10.1136/postgradmedj-2015-133662.

Vanopstal, Klaar et al. (2011). "Vocabularies and retrieval tools in biomedicine: disentangling the terminological knot." In: *Journal of medical systems* 35.4, pp. 527–43. ISSN: 0148-5598. DOI: 10.1007/s10916-009-9389-z.

Vellinga, A et al. (2016). "Intervention to improve the quality of antimicrobial prescribing for urinary tract infection: a cluster randomized trial". In: *Cmaj* 188.2, pp. 108–115. ISSN: 0820-3946. DOI: 10.1503/cmaj.150601. URL: http://dx.doi.org/10.1503/cmaj.150601.

Verbeke, Marc et al. (2006). "The International Classification of Primary Care (ICPC-2): an essential tool in the EPR of the GP." In: *Studies in health technology and informatics* 124, pp. 809–14. ISSN: 0926-9630. URL: http://www.ncbi.nlm.nih.gov/pubmed/17108613.

Verhoeven, a a, E J Boerma, and B Meyboom-de Jong (1995). "Use of information sources by family physicians: a literature survey." In: *Bulletin of the Medical Library Association* 83.1, pp. 85–90. ISSN: 0025-7338. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC226003/.

Veuillotte, I et al. (2015). "General practice and the Internet revolution. Use of an Internet social network to communicate information on prevention in France". In: *Health Informatics J* 1, pp. 3–9. ISSN: 1460-4582. DOI: 10.1177/1460458213494905.

W3C OWL Working Group (2012). "OWL 2 Web Ontology Language Document Overview". In: *OWL 2 Web Ontology Language* December, pp. 1–7. URL: http://www.w3.org/TR/owl2-overview/.

Walter, Sebastian (2017). *Question Answering over Linked Data*. URL: https://qald.sebastianwalter.org/ (visited on 08/13/2017).

Wang, Yefeng et al. (2008). "A computational linguistics motivated mapping of ICPC-2 PLUS to SNOMED CT". In: *BMC medical informatics and decision making* 8.1, S5.

Wells, K (1995). "The strategy of grounded theory: possibilities and problems." In: *Social work research* 19.1, pp. 33–7. ISSN: 1070-5309. URL: http://www.ncbi.nlm.nih.gov/pubmed/10140997.

Westen, Drew (1998). "The scientific legacy of Sigmund Freud: Toward a psychodynamically informed psychological science." In: *Psychological Bulletin* 124.3, p. 333.

Weston, W W and L G Dunikowski (1992). "A library for family physicians: Recommended list of books and journals." In: *Canadian family physician Médecin de famille canadien* 38, pp. 1051–3. ISSN: 0008-350X. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2145827/.

Whetzel, P. L. et al. (2011). "BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications". In: *Nucleic Acids Res* 39.Web

Server issue, W541–5. ISSN: 0305-1048. DOI: 10.1093/nar/gkr469. URL: http://dx.doi.org/10.1093/nar/gkr469.

Whitmee, Sarah et al. (2015). "Safeguarding human health in the Anthropocene epoch: report of The Rockefeller Foundation–Lancet Commission on planetary health". In: *The Lancet* 386.10007, pp. 1973–2028. ISSN: 01406736. DOI: 10.1016/S0140-6736(15)60901-1.

Wilkinson, Mark D et al. (2016). "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific data* 3, p. 160018.

Wittgenstein, Ludwig (1922). *Tractatus logico-philosophicus*. London: Kegan Paul, p. 108. ISBN: 1602064512. URL: http://people.umass.edu/klement/tlp/.

WONCA (2005). *ICPC-2-R: International Classification of Primary Care (Oxford Medical Publications)*. Oxford University Press, USA, p. 204. ISBN: 019262802X.

Yamamoto, Yasunori, Atsuko Yamaguchi, and Akinori Yonezawa (2013). "Building Linked Open Data towards integration of biomedical scientific literature with DBpedia." In: *Journal of biomedical semantics* 4.1, p. 8. ISSN: 2041-1480. DOI: 10.1186/2041-1480-4-8.

Younes, Nora (2016). *Introduction à Wittgenstein*. Paris: Editions La Découverte, p.11.

GLOSSARY

Abstract [résumé] ; A brief summary of a research article, thesis, review, conference proceeding, or any in-depth analysis of a particular subject and is often used to help the reader quickly ascertain the paper's purpose (Blake and Bly, 1993).

Annotator [annotateur] ; extract concepts, words, phrases, classifications, and named entities from unstructured content and mark these extractions as annotations. (https://www.ibm.com).

Bioontology portal ; the world's most comprehensive repository of biomedical ontologies (https://bioportal.bioontology.org/).

Brazilian Telehealth Network Program ; Brazilian national telemedicine system (Dias et al., 2015)

Category [Catégorie] ; Synonym of the term Class (Sadegh-Zadeh, 2015).

Category label [Etiquette de catégorie] ; often identical to Preferred term, is for human use.

Class [Classe] ; mutually exclusive, hierarchical (generic or partitive) and exhaustive designation of the concepts of a domain (Duclos et al., 2014).

Classification [Classification] ; is a mental process which aims at the ordering of the universe and experience and which is based on a reference system (Bonnet, 1999)).

Classification scheme [Shéma classificatoire]; is an official terminological system, recognized and endorsed by a national or international body, that is used to classify data. (http://meteor.aihw.gov.au).

Classifiers [classificateur]; In the terminology of machine learning, classification is considered an instance of supervised learning, i.e. learning where a training set of correctly identified observations is available. The corresponding unsupervised procedure is known as clustering, and involves grouping data into categories based on some measure of inherent similarity or distance. (Alpaydin, 2016).

Code [Code]; Unique Identifier of a concept (Duclos et al., 2014).

Coding [Codage] ; to bring similar data according to themes, concepts, etc. Generate code from the data level (inductively) or according to existing ideas (deductively) as necessary (Silver and Lewins, 2014).

Coding system [Système de codage] ; is a combination of a set of concepts, a set of code values, and at least one coding scheme mapping code values to coded concepts (Duclos et al., 2014).

Computational linguistics [Linguistique computationnelle] ; The branch of linguistics in which the techniques of computer science are applied to the analysis and synthesis of language and speech. (https://en.oxforddictionaries.com).

Concept [Concept] ; is a unit of thought, a specific meaning which refers to an object, the referent (Nelson, 2009).

Concept (in MeSH) [Concept] : the bearer of linguistic meaning in MeSH as defined by the Scope Note and consisting of Term elements. A single record will have one or more concepts, one of which is the preferred concept of the record. A Concept will have one or more terms, one of which is the preferred term for the concept. (https://www.nlm.nih.gov/mesh/xml_data_elements.html)

Conceptual data model; An abstract model or representation of data for a particular domain, business enterprise, or field of study, independent of any specific software or information system. Usually expressed in terms of entities and relationships (Baca, 2008).

Conceptualization [Conceptualisation] ; a world view often conceived as a set of concepts, their definitions and their interrelationships (Uschold and Gruninger, 1996).

Contextual [Contextuel] ; relates to the context i. e. the whole situation, background, or environment relevant to a particular event, personality, creation, etc (Dictionary, 2006).

Controlled vocabulary [Vocabulaire contrôlé] ; An organized arrangement of words and phrases used to index content and/or to retrieve content through browsing or searching. A controlled vocabulary typically includes preferred and variant terms and has a limited scope or describes a specific domain (Baca, 2008).

Crosslingual [Interlingue] move from one language to another by keeping the meaning in a concept centered system

Data; In common usage in computer science, this term is used as a singular noun to refer to information that exists in a form that may be used by a computer, excluding the program code (Baca, 2008).

Dereferencing [Dereferencement] ; Obtaining the address of a data item held in another location (https://en.oxforddictionaries.com).

Descriptor [Descripteur] ; concepts whose meaning does not overlap that of a concept in any other descriptor, and whose application will achieve a partition of the literature, expressed by the preferred term of the preferred concept assigned and attached to the citation record (https://www.nlm.nih.gov).

Distributed data [Données distribuées] ; is a database in which storage devices are not all attached to a common processor. It may be stored in multiple computers, located in the same physical location; or may be dispersed over a network of interconnected computers (https://www.its.bldrdoc.gov).

Dublin Core Metadata Element Set ; A set of fifteen metadata elements optimized for resource discovery on the web that can be assigned to information resources. Also often used as a "lowest common denominator" in metadata mapping. (http://dublincore.org/).

Dublin Core Schema ; is a small set of vocabulary terms that can be used to describe web resources (video, images, web pages, etc.), as well as physical resources such as books or CDs, and objects like artworks (http://dublincore.org/).

Electronic Medical Record (EMR) [Dossier medical électronique] (DMI)] ; Is the systematized collection of patient and population electronically-stored health information in a digital format (Gunter and Terry, 2005).

Entry term [Terme d'entree]; sometimes called "See cross-references" in printed listings, are synonyms, alternate forms, and other closely related terms in a given MeSH record that are generally used interchangeably with the preferred term for the purposes of indexing and retrieval, thus increasing the access points to MeSH-indexed data. The entry terms are equivalent to the preferred term for purposes of indexing and retrieval (https://www.nlm.nih.gov/).

Formal language [Langage formel]; is a language with mathematically precise construction rules applicable by computers.

Family Physician (FP) [Medecin de famille] (MF) ; A medical specialist who plans and provides the comprehensive primary health care of all members of a family, regardless of age or sex, on a continuous basis (dictionary, 1981).

Grey literature [literature grise] ; materials and research produced by organizations outside of the traditional commercial or academic publishing and distribution channels. (Schöpfel and Farace, 2010).

Health Sciences Descriptors (DeCS) [Descripteurs des sciences de la Sante]; is a controlled vocabulary developed by BIREME in the four languages (en, fr, pt, sp) of the Pan American Health Organization (PAHO/WHO). It is an extension of MeSH - MeSH Subject Headings of the U.S. National Library of Medicine and is used to represent the contents of indexed documents in major health databases, such as LILACS and

MEDLINE, providing a consistent means for retrieving information regardless of language. (http://decs.bvs.br).

Health Terminology/Ontology Portal(HeTOP) [Portail Terminologique de Santé]; The Health Terminology/Ontology Portal, made by the CISMeF team, Rouen University Hospital, Rouen (Normandy, France), provides access to more than 70 terminologies/ontologies (T/O) in Health (http://hetop.eu).

Health Sciences Descriptors (DeCs) [Descripteur de Sante] ; Quadrilingual (en, fr, es, pt) and structured vocabulary created by BIREME to serve as a unique language in indexing articles from scientific journals, books, congress proceedings, technical reports, and other types of materials, as well as for searching and retrieving subjects from scientific literature from information sources available on the Virtual Health Library (VHL) such as LILACS, MEDLINE, and others (http://decs.bvs.br).

Hypernym [Hypernyme] ; A word that is more generic than a given word, representing a class, a superordinate word (https://everything2.com/).

Hyponym [Hyponyme] ; A word that is more specific than a given word. Hyponymy is a sense relation in semantics that serves to relate word-concepts in a hierarchical fashion. Ex; apple:fruit (https://everything2.com/).

Index [Index] ; An index is a list of words or phrases ('headings') and associated pointers ('locators') to where useful material relating to that heading can be found in a document or collection of documents (https://www.nlm.nih.gov/).

Indexing [Indexation] ; choose of selected term in a text, selected from a controlled vocabulary according to a specific protocol in the larger scope of a collection (Névéol, Doğan, and Lu, 2010).

Interface terminology [Terminologie d'interface]; terminology used to facilitate data entry into electronic medical records. It link user's descriptions to structured data elements in a reference terminology (Rosenbloom et al., 2006).

International Classification of Primary Care (ICPC) [Classification Internationale des Soins Primaires] (CISP) ; is a coding system used in primary care to classify data about three elements of the health care encounter: reasons for encounter (RFE), diagnosis or problem, and process of care (https://www.nlm.nih.gov).

Keyword [mot clef] ; word representing what an author consider as important to describe the content of an article (Névéol, Doğan, and Lu, 2010).

Lightweight ontology [ontologie légère]; could be considered as informal, basic ontologies consisting of backbone taxonomies only (Giunchiglia and Zaihrayeu, 2009).

Linked data ; Data that is semantically linked by following a set of best practices for publishing and interlinking structured data that uses RDF syntaxes and HTTP URIs. (Baca, 2008)).

Linked Open Data (LOD) ; Linked data that is made available for use, reuse, and redistribution on the visible web (Baca, 2008)).

Linked Open Vocabularies (LOV) ; could be considered as a kind of informal Light Weigh Ontology (LOV, 2017).

Mapping [alignement] ; A set of correspondences between terms, fields, or element names used for translating data from one standard or vocabulary into another, or as a means of combining terms or data for search and retrieval. (Baca, 2008).

Matrix navigation [Navigation matricielle]; is the navigation between concepts of different terminologies via the semantic network, always with a notion of semantics (exact, BTNT, NTBT, related, etc.)

Medical Subject Heading (MeSH) : refers to the controlled vocabulary of NLM used for indexing articles in PubMed. MeSH terminology provides a consistent way to retrieve information that may use different terminology for the same concepts (McEntyre et al., 2002).

Medline ; is National Library of Medicine database of indexed journal citations and abstracts in the fields of biomedicine and healthcare. It encompasses nearly 4,500 journals published in the United States and more than 70 other countries (McEntyre et al., 2002).

Meronym [Meronyme] ; is a semantic relation specific to linguistics. A meronym denotes a constituent part of, or a member of something. (`https://everything2.com/`).

Meta-term (in CISMeF) [Méta-terme] ; CISMeF Meta-terms correspond to medical specialties (e.g. cardiology), types of medical procedures (e.g. surgery) or health topics (e.g. diagnosis, therapy), which has semantic links with one or more MeSH terms and subheadings.

Metadata ; are basic description mechanism for digital information that, can be used in all domains, for any type of resource, simple, yet powerful, can be extended and can work with specific solutions, making it easier to find information on the Web as it develops (`http://dublincore.org/`).

METHONTOLOGY ; Methodology for ontology development (Gómez-Pérez, Fernández-López, and Corcho, 2003).

Natural Language Processing (NLP) [Traitement Automatique de la Langue (TAL)] ; Natural language processing (NLP) is a field of computer science, artificial intelligence and computational linguistics concerned with the interactions between computers and human (natural) languages, and, in particular, concerned with programming computers to fruitfully process large natural language corpora. (https://ec.europa.eu/cefdigital).

Nomenclature [Nomenclature]; is a set of vocabularies that supports a singular classification system (Bentzen N.(ed), 2003).

Nomenclature [Nomenclature]; is an inventory of terms used to designate objects in a particular field, mostly when the system is based on user-specific rules rather than concepts (Duclos et al., 2014).

Ontology [Ontologie] ; in Computer Science, it is defined as a (formal), explicit specification of a (shared) conceptualization (Gruber, 1993). It can also be viewed as a metadata scheme that provide a controlled vocabulary, with each term defined explicitly by a machine computerizable semantics Chiaro and Damonte, 2005.

Preferred Term (PF) ; is a descriptor term used to represent a concept when indexing. A preferred term is usually a noun or a noun phrases (ISO, 2011).

PubMed; National Library of Sciences retrieval system containing citations, abstracts, and indexing terms for journal articles in the biomedical sciences (McEntyre et al., 2002).

Q-Codes ; Contextual classification in General Practice Family Medicine (GP/FM°.

Query [interrogation] ; In the context of retrieval, a command to look in a database and find records or other information that meet a specified set of criteria. The most precise queries are those that return the fewest false hits (Baca, 2008).

Question answering (Q & A) [Question - Réponse]; Given a human-language question, determine its answer. Typical questions have a specific right answer (such as "What is the capital of Canada?"), but sometimes open-ended questions are also considered (such as "What is the meaning of life?") (Mittal and Mittal, 2011).

Redundancy; is related to the extent to which it is possible to compress the language (Shannon, 1958).

Reference terminology [Terminologie de réfrence]; is a set of concepts and relationships that provides a common reference point for comparison and aggregation of data about the entire health care process, recorded by multiple different individuals, systems, or institutions (Smith et al., 2006).

Resource Description Framework (RDF) ; A standard model for data interchange on the web that extends the linking structure of the web to use URIs to name relationships between things. RDF enables structured and semistructured data to be exposed and shared across different applications. (http://www.w3.org).

Semantic Interoperability [Interoperabilité sémantique] ; The ability of different agents, services, and applications to communicate data while ensuring accuracy and preserving the meaning of the data. (Baca, 2008).

Semantic triple [Triplet] ; A triple is a set of three entities that codifies a statement about semantic data in the form of subject–predicate–object expressions (e.g. "Bob is 35", or "Bob knows John"). (http://www.w3.org).

Semantic Web [Web Sémantique] ; An evolving, collaborative effort led by the World Wide Web Consortium (W3C) whose goal is to provide a common framework that will allow data to be shared and reused across various applications and enterprise and community boundaries. It derives from W3C director and inventor of the World Wide Web Tim Berners-Lee's vision of the web as a universal medium for data, information, and knowledge exchange (Baca, 2008).

Simple Knowledge Organization System (SKOS); An endeavor of the World Wide Web Consortium that develops specifications and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading lists, and taxonomies within the framework of the Semantic Web (Baca, 2008).

SPARQL (Protocol and RDF Query Language); is an RDF query language, that is, a semantic query language for databases, able to retrieve and manipulate data stored in Resource Description Framework (http://www.w3.org).

Structured vocabulary ; synonym of controlled vocabulary.

Synonym [Synonyme] ; Expression semantically related and meaning the same thing (Duclos et al., 2014).

Taxonomy [Taxonomie] ; A subject-based classification that arranges the terms of a controlled vocabulary into a hierarchy, based on essentially one relationship: the broader/narrower relationship used to build the hierarchy (Garshol, 2004).

Term [Terme] ; is the verbal designation of a general concept in a specific subject field (Duclos et al., 2014).

Term (in MeSH) [Terme] : Alpha-numeric string which comprises the basic unit of the MeSH vocabulary. (https://www.nlm.nih.gov/mesh/xml_data_elements.html).

Terminology [Terminologie] ; is a set of designations belonging to a special language related to the concepts of a specific domain (Duclos et al., 2014).

Thesaurus [Thesaurus] ; controlled and structured vocabulary in which concepts are represented by terms, organized so that relationships between concepts are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms (https://www.iso.org/obp/ui/).

Uniform resource identifier (URI) ; short strings that identify resources in the web: documents, images, downloadable files, services, electronic mailboxes, and other resources. (http://www.w3.org).

Uniform resource locator (URL) ; The address of a resource on the Internet (McEntyre et al., 2002).

Unique identifier [Identifiant unique]; A number or other string that is associated with a record or piece of data, exists only once in a database, and is used to uniquely identify and disambiguate that record or piece of data from all others in the database (Baca, 2008).

Universe of discourse [Univers du discours] ; whatever may be the extent of the field within which all the objects of our discourse are found, that field may properly be termed the universe of discourse (Boole, 1854).

Virtual repository [dépôt virtuel] ; collection of local, remote and other repositories accessed through a single logical URL which hides the access details of the underlying repositories letting users work with a single, well-known URL. (www.jfrog.com).

Vocabulary coding scheme ; synonym of Controlled medical vocabulary

Web ontology language (OWL) ; is a Semantic Web language designed to represent rich and complex knowledge about things, groups of things, and relations between things. OWL is a computational logic-based language such that knowledge expressed in OWL can be exploited by computer programs, e.g., to verify the consistency of that knowledge or to make implicit knowledge explicit. (https://www.w3.org/OWL/).

World of reference [Monde de référence]; The world to which the reader is invited to refer (Eco and McEwen, 2001).

World Wide Web Consortium (W3C); an international community that develops open standards to ensure the long-term growth of the Web. (http://www.w3.org).

INDEX