ESTABLISHMENT OF A GEOGRAPHIC DATA DICTIONARY : CASE STUDY ON THE BRUSSELS REGIONAL GOVERNMENT GIS

DIMOS PANTAZIS¹, Bernard CORNÉLIS², Roland BILLEN^{2,3}, David SHEEREN²

¹ TEI (Technological Education Institute -TEI) of Athens

² Department of Geomatics, University of Liege Belgium

ABSTRACT

This article focuses on the establishment of a geographical database dictionary. It develops the elaboration process of the *Brussels UrbIS* © data dictionary. It presents the experience gained in the field of urban data management from a practical study. It describes the problems and difficulties encountered as well as the proposed solutions and perspectives on future improvements of this data dictionary. This expertise could be applied to the development of geographic data dictionaries in similar cases.

THE CONTEXT

In 1997, a collaboration between the CIRB (Centre Informatique de la Région Bruxelloise) and the laboratory SURFACES (Service Universitaire de Recherches Fondamentales et Appliquées en Cartographie et en Etudes Spatiales) – Department of Geomatics, University of Liège - was established. The objective of this on-going collaboration is to give scientific support for the improvement, the development and the reengineering of *Brussels UrbIS* © (Brussels Urban Information System).

Some words about Brussels UrbIS ©

Brussels UrbIS © is the geographic information system of the Regional Government of Brussels, a quite complex and advanced one. More specifically, it includes on one hand a set of geographic and non geographic databases related to the Brussels-Capital Region in Belgium and on the other hand software designed to handle the databases in order to sustain the development of different applications. The Brussels UrbIS 2 © is actually composed of four databases:

- the ADM base with principally geographic administrative information;
- the PWN base with the public ways network;
- the TOP base with topographic surveys;
- the FOT base, a collection of aerial photographs.

and of the following software:

- UrbIS SPW or UrbIS WSPW: a software able to search and manage address, taking into account eventual orthographic / spelling errors in the roads names;
- UrbIS MGR: a software specific for the management of the Brussels UrbIS databases in MicroStation Environment:

³ Aspirant au Fonds National de la Recherche Scientifique (FNRS) Belgium

- Extension of ArcView for UrbIS.

Objective of the article

So far the collaboration with the CIRB consisted in the auditing of the organization, in the creation of conceptual data models (CDM), in the creation of a new, up-to-date, complete and analytical geographic database dictionary, in the establishment of quality assessment methodologies.

The interest of this paper lies in the observation that, although there exist nowadays standards (national, European, American, Canadian, of different international organizations,...e.g. the standards of the *Federal Geographic Data Committee-USA*, fig.1) concerning the metadata creation, there is no official / standard indications on the structure of a data dictionary. The problem is even more acute when dealing with geographic databases dictionary, because of the specific nature of these data and of the lack of specific CASE (Computer Aided Systems/Software Engineering) tools. In other words, the alphanumerical databases dictionaries methodologies cannot be applied just as they are.

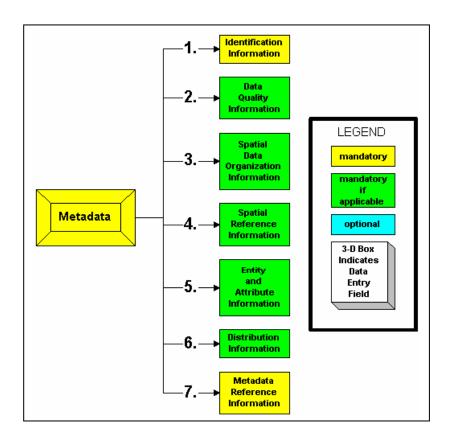


Figure 1. FGDC 1998 Metadata Standards

Structure of the article

In the second paragraph, some concepts about geographic databases metadata, data dictionary and their relationships are presented. The third paragraph exposes the steps followed. The fourth paragraph develops questions which arose and answers given. It goes from assigning a

role to the database dictionary to practical solutions proposed. In paragraph 5, examples extracted from the work are presented. The final paragraph gives the conclusions and explores some perspectives for future work.

METADATA AND DATA DICTIONNARY: DEFINITIONS

The metadata concept is generally defined as data about data (CEN/TC 287, 1998) (Federal Geographic Data Committee, 1998). In regard to geographic information metadata can concern the representation, the quality, the providers, the cost ...etc. of the geographic data.

Many definitions of what is or must be a data dictionary exist (*Antenucci et al.*, 1991), (*Connolly et al.*, 1996), (*Hansen G. & Hansen J.*, 1992). A data dictionary is a subset of the entire metadata set. However, in some cases, modifications of the existing metadata or creations of metadata specific to the dictionary are necessary.

Two main assets of a data dictionary are its flexibility (in comparison to the complete metadata set), and its easiness of access for the end user.

METHODOLOGY FOLLOWED

The methodology adopted for the development of the geographic database dictionary had seven phases :

- I. Bibliography analysis of metadata standards and real case studies of geographic data dictionaries.
- II. Analysis of primitive *Brussels UrbIS* © dictionary and related documents (*Service Communal de Belgique*, 1996), (*Comité*, 1997), (*Donnay & Pantazis 1997*), (*Billen et al. 1998d*), (*Centre d'Information*, 1997).
- III. Development of the dictionary requirements.
- IV. Development of the dictionary based on the list of requirements, with permanent consulting of the Brussels UrbIS development team. This phase also included the evaluation of an alternative dictionary proposal.
- V. Limited distribution of the dictionary to the end users of Brussels UrbIS users for critics, proposals, remarks and validation.
- VI. Final edition of the dictionary after insertion of the remarks of the end users.
- VII.Translation of the dictionary in the other two languages (see next §).

BASIC PROBLEMS, NEEDS, QUESTIONS AND SOLUTIONS

Why a data dictionary? What are exactly the needs to be covered?

While most real world elements have a commonly accepted cognitive meaning, their representation in (geographic) databases require a precise and unambiguous semantic definition. In the case of a street object, does the object correspond to the road axis, to the space between the pedestrian pavements or to something else? The urban environment is quite complex to formalise in terms of objects and to express the objects relations. So, the project objectives was to create a document which neatly describes with a minimal set of metadata the objects of the geographic databases and which could be used as a starting point for future metadata set.

Who will be the user of this database dictionary? For who will it be developed?

This was a critical choice because the answer would largely influence the contents of the dictionary. We choose to design the dictionary mainly for the present and forthcoming end users of *Brussels UrbIS* ©, for the decision makers of its development, and secondarily for its programmers and developers. Thus, we focused on the comprehensive help that a database dictionary can provide to the end user. It was also clear that the dictionary would only cover the databases of the system ADM and PWN – the databases TOP and FOT were not ready yet.

Which information should it hold?

As already stated, the dictionary is a subset of the metadata set, but where should the subset end and start, and why? A quick and easy answer would sound like: the definitions of the "entities". But which entities? The entities of the real world or the representations of these entities in the database? (points, lines, polygons,...). At which level (natural language / non specialist end user level, conceptual, logical, physical) do we have to select the dictionary metadata? How is it possible to balance contradictory requirements of exhaustivity and of minimal information?

The choice of the principal user of the dictionary determined its contents: basic information and description of the databases without complex geoinformatics terms, but which metadata must be selected for the dictionary and how many (too much information would discourage an easy and frequent update)? After analysis of the existing primitive dictionaries, the database objective, the users profile, their wishes and the constraints of the reality, we selected the following elements:

- Base de données: the name of the database that the object belongs to.
- *Type*: type of geographical implementation (point, line, polygon or any combination of this types). This type is described by a text and a pictogram.
- *Name*: name of the object, should be unique.
- *Nom*: translation of the name in French.
- Naam: translation of the name in Dutch.
- Abréviation: abbreviation of the name which allows to identify directly the object.
- Définition de l'objet UrbIS réel: description of the corresponding object in the real world. This textual definition must be illustrated by a scheme, a photo or any kind of visual information that can help to understand the object.
- *Identifiant*: characteristic that allows to identify strictly each instance of the object, identifier. Usually this attribute is a numeric integer.
- *Attributs alphanumériques*: all the alphanumeric (non geometric) attributes describing the object's instances.
- Définition de la représentation graphique: description of the graphical representation of the object. This description is textual and visual using real example of the database.
- *Nombre d'instances*: number of instance of the object in the database.
- Sources: origin of the data (document, property, etc.).
- *Historique*: description of all steps allowing going from the original data source to the final data set of the database.

Indexes and lists were added at the beginning of the dictionary:

- list of the geographic objects;

- an alphabetical index;
- a list with the objects that compose/are composed by others objects;
- a list of the objects that are not composed by others objects;
- a list of the objects by type of implementation (point, line, ...)
- a list of generalised /specialised objects,
- and a thematic index.

Number of dictionaries : one or many dictionaries ?

The dictionary being created was intended for end-users. Was it advisable to develop another one, dedicated to the programmers? Until now, no other document, entitled dictionary, was created, to avoid confusion. Nevertheless complementary documents, covering other aspects of the meta-information of *Brussels UrbIS* © were or will be developed. See also next paragraph for additional information concerning the number of dictionaries.

Language: which language, while three are common in Brussels?

In this highly urban region, the linguistic problem is a sensible one, as three languages (French, Flemish, English) compete. The selected solution was the development of the dictionary in French (working language of our team and of most (>80%) of the inhabitants) until its "final" version would be ready for translation in the other two languages. A delicate work to realise, because of the inadequacy of some terms in all three languages!

Evolution of the dictionary

Should it be an autonomous (or related with something), independent or complementary tool of the database it describes, or of the global set of metadata, or of the database conceptual model? Should it be evolving, or static? Why and how?

The $Brussels\ UrbIS\ \odot$ is not a static system; in the same way, the proposed dictionary is not a static tool too, nor a definitive product, it is an evolving one. The dictionary still has to be improved and updated, as well as the databases.

Having a structured document at hand, active end-users can participate in improving it through critics and modification suggestions. It is well-known that users can better point to features they don't like about an existing system (or indicate when a feature is missing) more easily than describing what they think, what they would like in an imaginative system (dictionary in our case) (*Jenkins*, 1985).

The dictionary built is not complementary to another document e.g. a conceptual data model, or to another metadata set, or to the database itself. In this sense, the dictionary is an autonomous document, in our specific case.

Nevertheless the dictionary is not an independent document, for it can not exist by itself. The database dictionary is related with the database: modifications to the database implies modifications to the data dictionary. This is a feature common to every dictionary.

On what support?

Should it be digital, automatic and in direct connection to the database (that would guarantee the automatic update of the dictionary), or not?

According to the fact that the objective was the establishment and the refinement of the dictionary, paper document accompanied by its file was produced. The possibility to develop a dictionary with a database management system (DBMS) is being studied while a web based approach was discussed and could represent a future extension.

Update of the database dictionary

A database for which the updating process in not exhaustively studied could be very problematic in the near future. This statement can be applied for the database dictionary, and especially for the *Brussels UrbIS* © database dictionary because of its fast and continuous evolution: addition of objects, completion of the dictionary on request of the users,... In a second phase, automated procedures for the updating of the dictionary have been advised (active dictionary).

Cost

Four questions have been analysed:

- I. Who will have to pay the cost of the data dictionary?
 Two options are available: the CIRB could distribute the dictionary freely to the users, and be in charge of the whole cost, or the dictionary could be sold to the users, in order to
 - cover at least part of the costs. No decision has been met till now, as the dictionary is not yet ready for public diffusion.
- II. Will it be of free access?
 - In this case, anyone could have it without charges in digital format or in paper (more expensive option). Another option is to be free on the INTERNET, but at real cost for the printed version.
- III. Will it be handed over with the database only?

 This is the most "strictly policy" option. If adopted the dictionary could not be used by potential future users who want formal information about the contents of the database.
- IV. Will it be possible to buy it without the database?

 The question exists if the dictionary has a price for the end user. In this case, it will possible for anyone to get complete formal information about the databases of *Brussels UrbIS* © at a low individual cost.

EXAMPLES OF BRUSSELS URBIS 2 © DICTIONARY

The different steps of the *Brussel UrbIS* © data dictionary elaboration are presented in the figure 2. At the early beginning, the available information about the object of the database was insufficient, unclear and uncompleted. Therefore, the CIRB wished to elaborate a database dictionary, which would contain the necessary information. Thus, the first convention began and a strong collaboration between partners was established (*Billen et al.*, 1998a). Following the primary advice, the CIRB team gathered all the necessary information available. Then the SURFACES team proposed a first dictionary structure (*Billen et al.* 1998b).

	SURFACES team		CIRB team
	Tasks	Dictionary elaboration	Dictionary elaboration
— 1/2/98 Convention 1 — 1/4/98	Establishment of data dictionary structure Conceptual data model design - version 1		Fig 3
- 1/8/98 Convention 2 - 1/11/98	Conceptual data model design - version 2	1	
— 15/5/99		1	
— 1/2/00 Convention 3 — 1/5/00	Establishment of data quality criterions Establishment of data quality control methodology		— Fig 4 — Fig 5
ļ t			:

: intensive work duration
: elaboration and management duration

Figure 2 : The Brussels UrbIS data dictionary time table.

Figure 3 is an example of an index card presented by the SURFACES team.

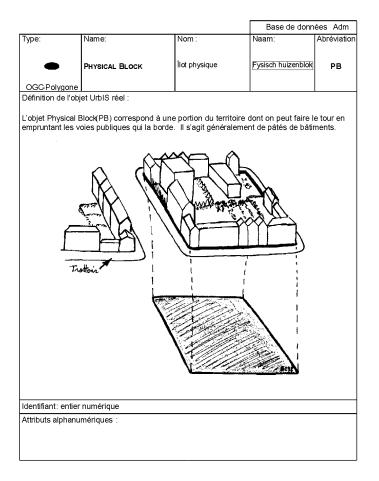


Figure 3a: The proposed index structure for an object – part1

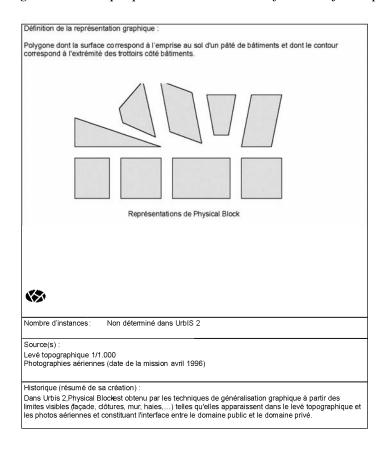


Figure 3b: The proposed index structure for an object – part2

On that basis, the CIRB team began the elaboration of the dictionary. During the convention 2 (*Billen et al.*, 1998c) and later in 1999, different dictionary projects were submitted to the SURFACES team. Some remarks were formulated, especially about the object's definition. The *Brussels UrbIS* 2 © is a complex database that is still in elaboration. New objects were created, others disappeared and the topological relationships had to be updated according to the recent modifications. The figure 4 is an index card of the dictionary in February 2000 just before convention 3.

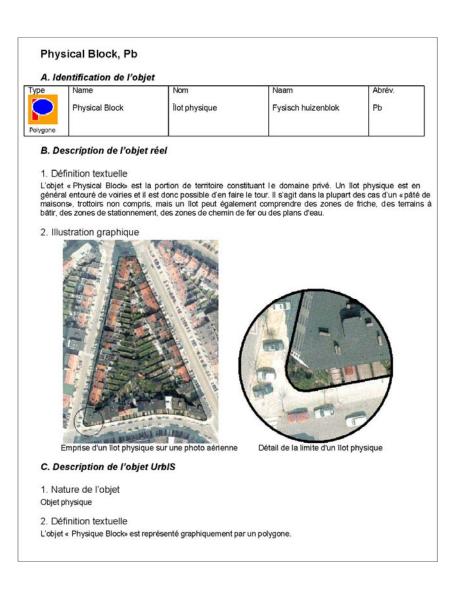


Figure 4a: an index object – part1 (CIRB 1999)

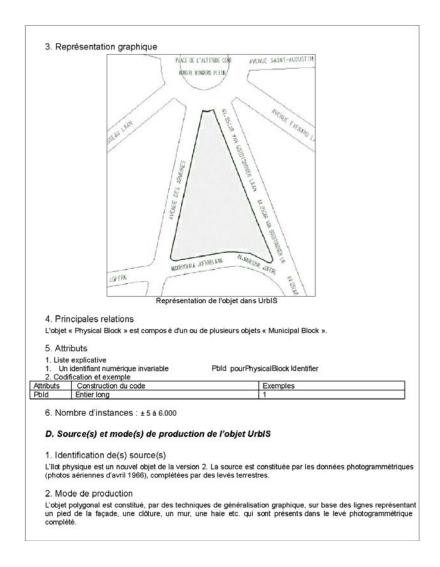


Figure 4b: an index object – part2 (CIRB 1999)

The most of the structure is similar to the one proposed at the end of the convention 1. Except minor vocabulary and sub-structural changes, few differences occur.

Two headings were deleted:

- The base de données. no information about the name of the database is available anymore.
- The *identifiant* was sent to the alphanumerical attributes (*attributs alphanumériques*).

Two news headings were created:

- Principales relations: contains the 'main' relationships of the object with other objects.
- Adaptations successives : presents the different updates.

The information about the 'main' relationship is ambiguous. Actually, it is difficult to say if a relationship is important or not. Above all a distinction between structural and topological relationships should be done. The information about the relationships is an important metadata, but it is certainly not relevant for most of the users of the dictionary. Some inconsistencies (definition, sources) were found in this version. The dictionary is still under construction.

Finally, the convention 3 (*Sheeren & Billen, 2000*) leaded to a new set of metadata about quality. The quality control is not a part of the dictionary, but the CIRB decided users should have access to it.

In order to make a relevant quality control, quality criterions have to be established strictly. On the basis of quality standards, in particular the European pre-standard (*CEN/TC 1998*), quality criterions were hold:

- positional accuracy and precision;
- semantic accuracy and precision,
- completeness;
- logical consistency;
- lineage;
- temporal precision.

Figure 5 is the proposed structure for the quality report.

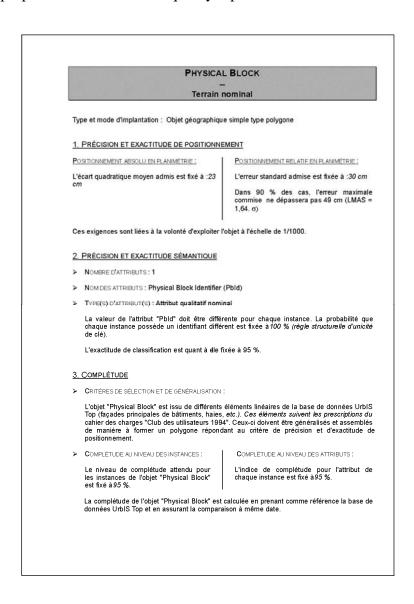


Figure 5a: Quality report structure – part1

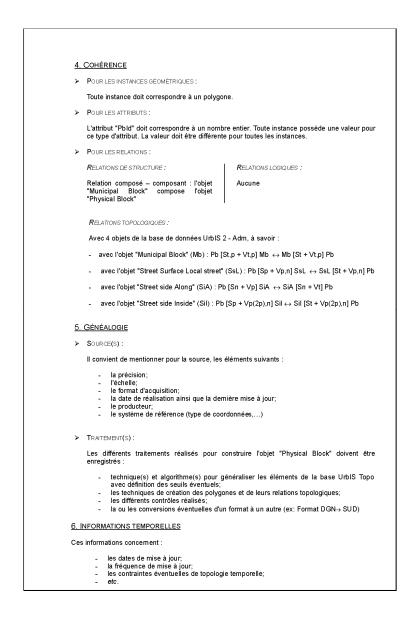


Figure 5b : Quality report structure – part2

CONCLUSIONS AND PROSPECTIVE

The development of a geographic database dictionary may be a very complex project. This was the case with the *Brussels UrbIS* database dictionary. The principal reasons for the difficulties and complexity of such a project arise from the following reasons:

- a) lack of dictionary standards;
- b) complicated and unclear definitions of the geographic objects as well as of the real objects they represent; the object should be strictly defined to prevent confusions and errors, especially in an urban context;
- c) lack of specific CASE-tools that could permit the automatic metadata/dictionary database creation.

In addition to the ongoing phase, that is the development of quality metadata set, the principal prospective of this project are :

- a) a cost / benefit analysis;
- b) study concerning the digital support of the dictionary: word processing software, Data Base Management System, Internet site, combination, other but also the analogic one: dossier with cards that could be replaced, catalogue, etc.;
- c) study concerning the controls and the process that could warrant the simultaneous update of the databases and dictionary;
- d) analytical study of existing software concerning the automatic metadata creation, queries and modifications. Analysis of the possibility to be used in the frame of Brussels UrbIS project.

ACKNOWLEDGMENTS

The authors would like to thank the kind support of the CIRB staff, which closely interacted with us to discuss various aspects of the data models and dictionary. Our recognition goes in particular to Mr Vanderborght, Mr Van Acker, Mrs Himpe, Mrs Roland and Mr Delande. Special thanks to Chantal Donzé for her editorial help and her constructive critics and remarks.

REFERENCES

- Antenucci J., Brown K., Croswell P., Kevany M., (1991), Geographic Information Systems, Chapman et al., USA, 1991
- Billen R., Cornélis B., Muller, F., Pantazis D., Thiam S. & Donnay J.P., (1998a): Élaboration du modèle conceptuel de données de la base UrbIS2 ADM version 1, rapport d'expertise, SURFACES, Université de Liège, 40 p.
- Billen R., Cornélis B., Muller, F., Pantazis D., Thiam S. & Donnay J.P., (1998b): Ébauche du dictionnaire de données de la base UrbIS2 ADM version 0.1α, rapport d'expertise, SURFACES, Université de Liège, 90 p.
- Billen R., Pantazis D. & Donnay J.P., (1998c.): Validation des modèles conceptuels de données de la base Brussels UrbIS 2 ®©, rapport d'expertise, SURFACES, Université de Liège, 90p.
- Billen R., Pantazis D., Donnay J.-P.,(1998d): Validation des modèles conceptuels de données de la base Brussels UrbIS 2, Contrat-Cadre: Bruxelles UrbIS Supervision Scientifique, Convention 2, Rapport final, Université de Liège, Faculté des Sciences, Département de Géomatique, 1998.
- CEN/TC 287 (1998). Geographic information data description metadata (ENV 12657). European prestandard
- CEN/TC 287 (1998): Geographic information data description metadata (ENV 12657). European prestandard
- Centre d'Informatique pour la Région Bruxelloise (1997): Cahier n°6 du CIRB, Catalogue des utilisateurs de Brussels UrbIS, 1997.
- Centre d'Informatique pour la Région Bruxelloise (1999):Brussels UrbIS®© version 2 Base de données Adm Dictionnaire de données version 1.0 novembre 1999, 1999.
- Comité d'Avis pour la cartographie régionale Brussels UrbIS (1997): Historique-Description générale-Objets administratives-linéaire de circulation, 1997.
- Connolly Th., Begg C., Strachan A., (1996) Database Systems, Addison Wesley, 1996
- Conseil National de l'Information Géographique (1994) : Eléments d'un vocabulaire géomatique, Paris : Commission permanente de la Recherche géographique, 57 p.

- Donnay J.-P., Pantazis D., Université de Liège, Faculté des Sciences, Département de Géomatique (1997): Audit de la base de données Brussels UrbIS Région de Bruxelles Capitale, Rapport Final d'Expertise, 1997.
- Federal Geographic Data Commitee. FGDC-STD-001-(1998): Content Standard for Digital Geospatial Metadata (revised June 1998). Federal Geographic Data Commitee. Washington, D.C. <URL: http://fgdc.er.usgs.gov/>, consultation en février 2000.
- Hansen G., Hansen J., (1992) Database Management and design, Prentice-Hall, Englewood Cliffs, New Jersey, USA, 1992.
- *Jenkins*, A.M., (1985): Prototyping: a methodology for the design and development of applications systems, Spectrum, 2 (2), 1985.
- Pantazis D., Donnay J.-P. (1996): Conception des SIG Méthode et formalisme, Editions Hérmès, Paris, 1996.
- Service Communal de Belgique, Marché de Services (1996): Cahier spécial des charges, Dossier n°96F006, Objet de l'Entreprise, Cartographie à grande échelle de la Région de BruxellesCapitale, 1996.
- Sheeren D, Billen R., (2000). Méthodologie de validation de la qualité des données de la base de données UrbIS 2 ADM, rapport d'expertise, SURFACES, Université de Liège, 77 p. + annexes

CO-ORDINATES

Dr., M.Sc., Eng. Dimos PANTAZIS
TEI (Institute of Technological Education) of Athens
Trapezountos 98, Korydallos
18 121
Greece

Phone: +30 (0)1 561 61 54 Fax: +30 (0)1 561 51 54 pantdonz@otenet.gr

Researcher, PhD candidate, Bernard CORNÉLIS University of Liege, Department of Geomatics 7, place du 20-Août, Liege, 4000 Belgium

Phone: +32 (0) 4 366-57-52 Fax: +32 (0) 4 366-56-93 Bernard.Cornelis@ulg.ac.be

Researcher, PhD candidate Roland BILLEN
Aspirant au Fonds national de la Recherche Scientifique (FNRS)
University of Liege, Department of Geomatics
7, place du 20-Août, Liege,
4000
Belgium

Phone: +32 (0) 4 366-57-52 Fax: +32 (0) 4 366-56-93 Roland.Billen@ulg.ac.be Researcher David SHEEREN University of Liege, Department of Geomatics 7, place du 20-Août, Liege, 4000 Belgium

Phone: +32 (0) 4 366-57-52 Fax: +32 (0) 4 366-56-93 sheeren@geo.ulg.ac.be