



Stéphane Polis (F.R.S.-FNRS – ULg)

Vincent Razanajao (ULg)

Nathalie Sojic (ULg)

Towards a data model for (inter)textual relationships

Connecting Ancient Egyptian texts and
understanding scribal practices

+ Outline of the talk

- Background information about the *Ramses Project* & *Ramses Online*
- Evolution of the data model between 2006 and 2016
- Data models: state-of-the-art in digital editing
- Towards a data model for (inter)textual relationships in Ancient Egyptian



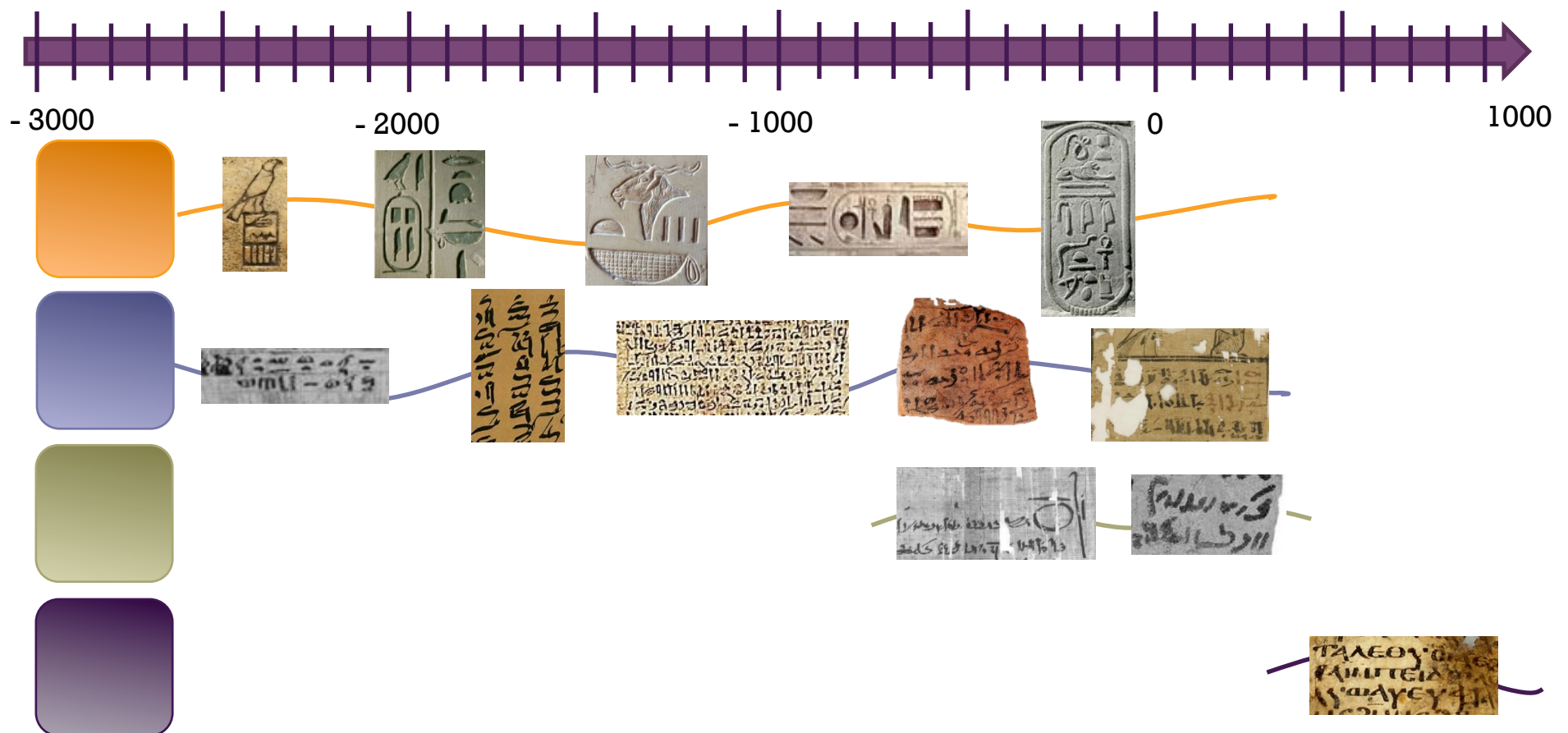
+ The Ramses Project & Ramses Online

Background information

+ The Ramses Project

Goal

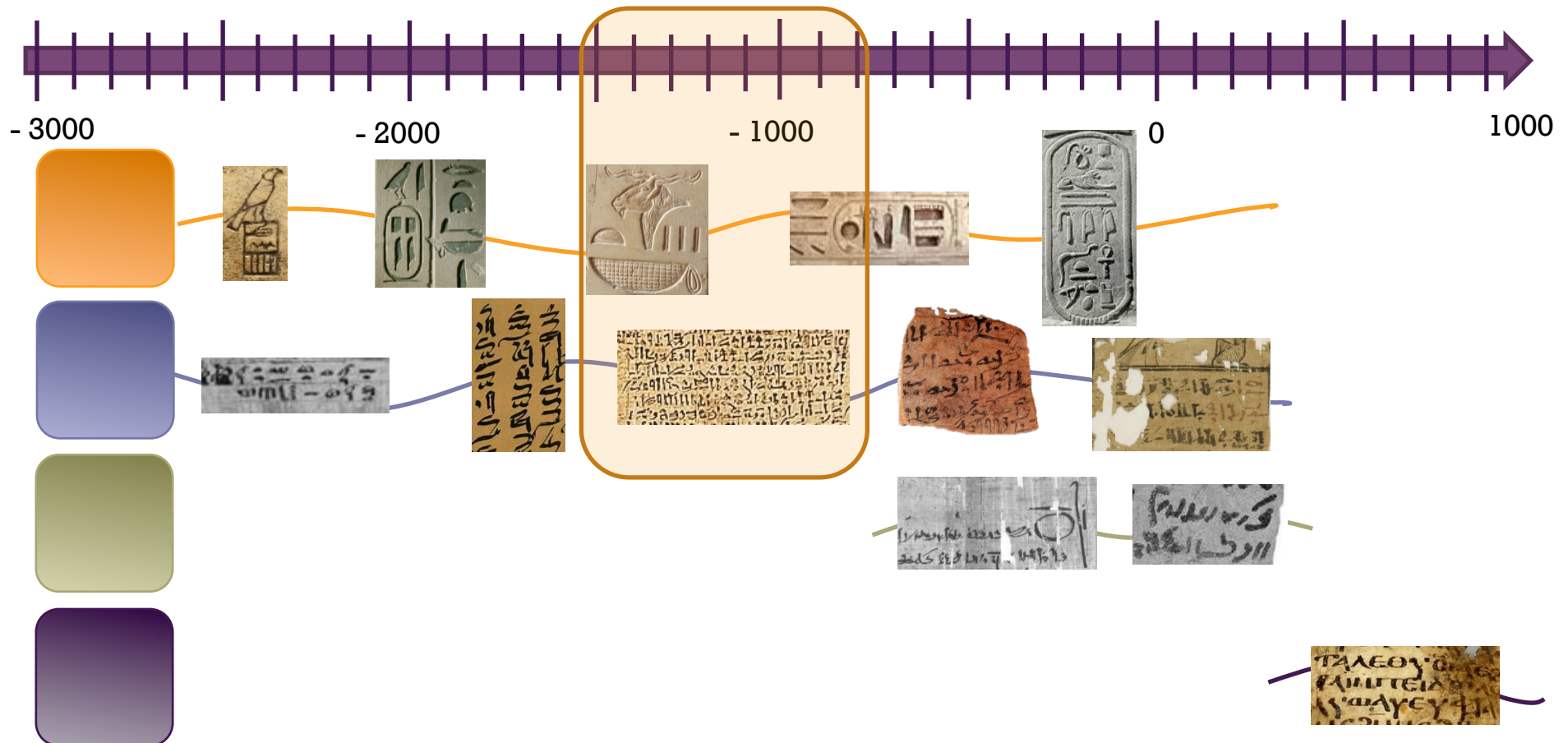
- Build a richly annotated corpus of Late Egyptian texts



+ The Ramses Project

Goal

- Build a richly annotated corpus of Late Egyptian texts



+ The Ramses Project

Goal

- Build a richly annotated corpus of Late Egyptian texts
- Useful both for philologists and linguists

+ The Ramses Project

JAVA software (MySQL – texts stored in XML)

- LexiconEditor
- TextEditor

Ramses 20140317 / P. BM 10440 (= LRLC 3, P. Salt 1821, 9 A. 106, 8 A. 132) (= P. BM 10440); id: 224 / [2/pl. 9]

Fichier Menu textes Éditer Insérer Chercher Exportation Bibliographie Statistiques

LRLC P. BM 10440 pl. 9 1

substantive scribe

substantive +tTry+I

n default connectors de

article/demonstrative le

substantive tombe, nécropole

spelling ? n default connectors à, pour

substantive chef

n default connectors de

substantive équipe

substantive +tbAk-n-mw.t+I

n default connectors de

article/demonstrative le

substantive +tbw-thA-imm+I

substantive gardien

substantive +tkr+I

m default connectors dans, en tant que,de

substantive V,S,F

spelling ? m default connectors dans, en tant que,de

substantive éloge, faveur

substantive Amon-Rê

substantive roi

Insérer mot Mettre à jour mot Statut des éléments Noir Rouge Autre Normal Lacune Complément de l'éditeur Ajouter lacune Ajouter Ponctuation Ajouter espace

Traduction Le scribe de la Tombe +tTry+I au chef d'équipe de la Tombe +tbAk-n-mw.t+I, au scribe +tbw-thA-imm+I et au gardien +tkr+I, en vie, prospérité, santé et < dans > la faveur d'Amon-Rê, roi des dieux.

Commentaire

Filterer Signes

Lemme

- m-bAH (en présence de) connecto
- in (par) connectors
- iwi (venir) verb
- n (à, pour) connectors
- n(j) (de) connectors
- rdi (donner) verb
- mi (comme) connectors
- m-sA (derrière) connectors
- mtw (mtw (conj.)) independant gra
- m ((négation)) independant grammatic
- m (dans, en tant que,de) connectors

Flexion

Graphie

Analyse

Graphie:
















Lemmes et flexions

- m (dans, en tant que,de) connectors
- default m

+ The Ramses Project

What kind of data?

- Hieroglyphic spellings

 ir	 p3y	 4	 hrd.w	P. Ashmolean Museum 1945.97 r° 4:8	 ink	spelling ? bn	 iw	 w	 r	 c q	 r
default quant à, si	ce, cet	default 4	enfant		je	default négation	default iw (F3)	ils	default r (F III)	infinitive entrer	default vers, pour
 pš	 m	 h.wt	 i	 nb.wt	END						
infinitive diviser	default dans, en tant que, de	chose, biens	je	default tout							

+ The Ramses Project

What kind of data?

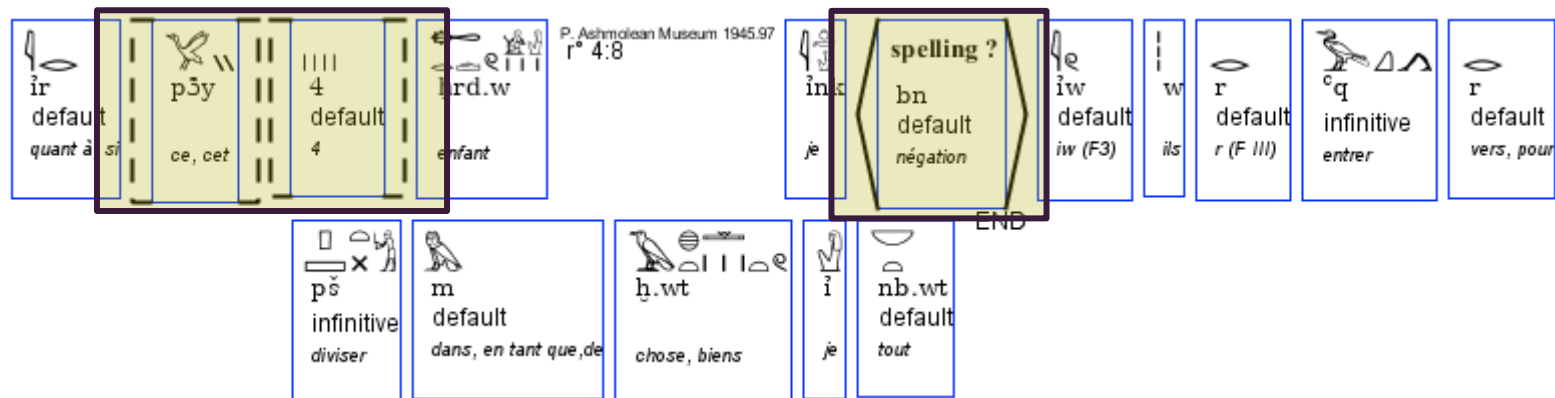
- Hieroglyphic spellings
- Lemmatization and morphological annotation

default quant à, si	ce, cet	default 4	enfant	je	spelling ? bn default négation	default iw (F3)	ils	default r (F III)	infinitive entrer	default vers, pour
					END					
infinitive diviser	default dans, en tant que, de	chose, biens	je	default tout						

+ The Ramses Project

What kind of data?

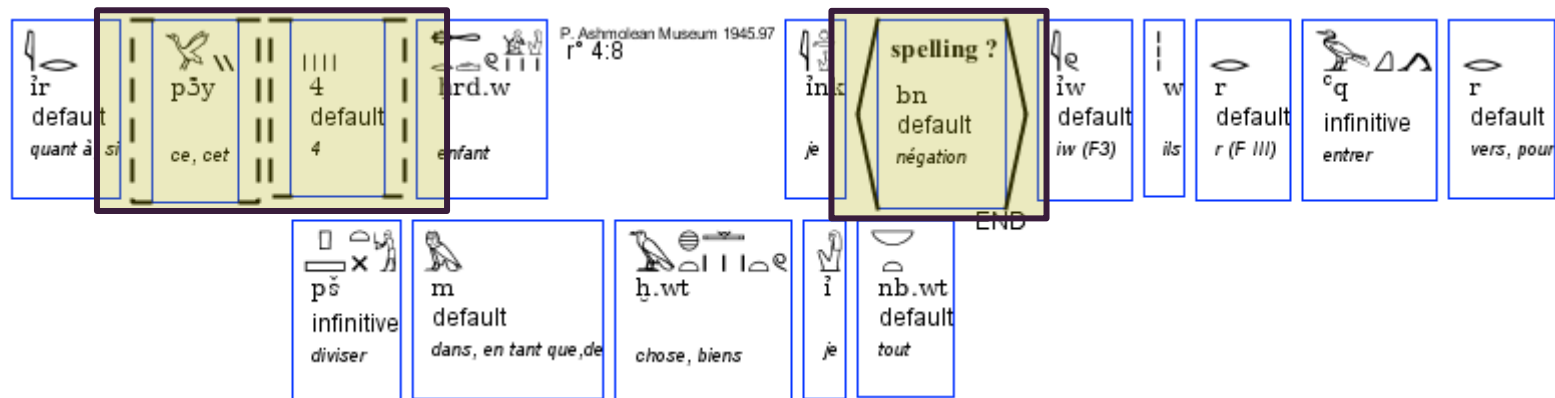
- Hieroglyphic spellings
- Lemmatization and morphological annotation
- Textual criticism



+ The Ramses Project

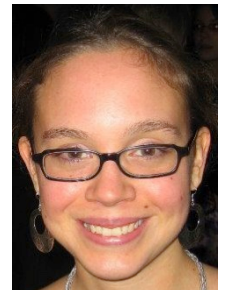
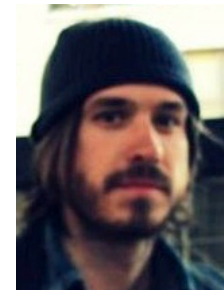
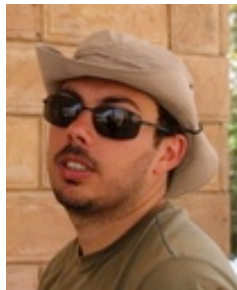
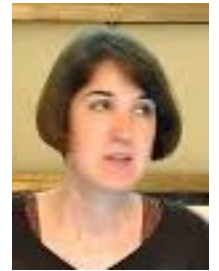
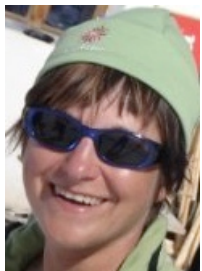
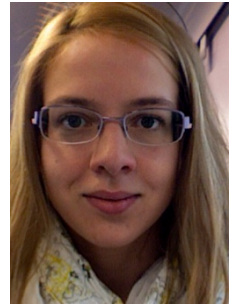
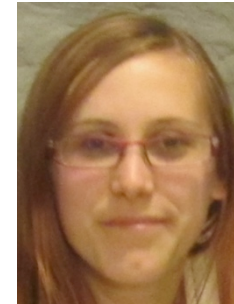
What kind of data?

- Hieroglyphic spellings
- Lemmatization and morphological annotation
- Textual criticism
- Translation (French / English)



+ The Ramses Project

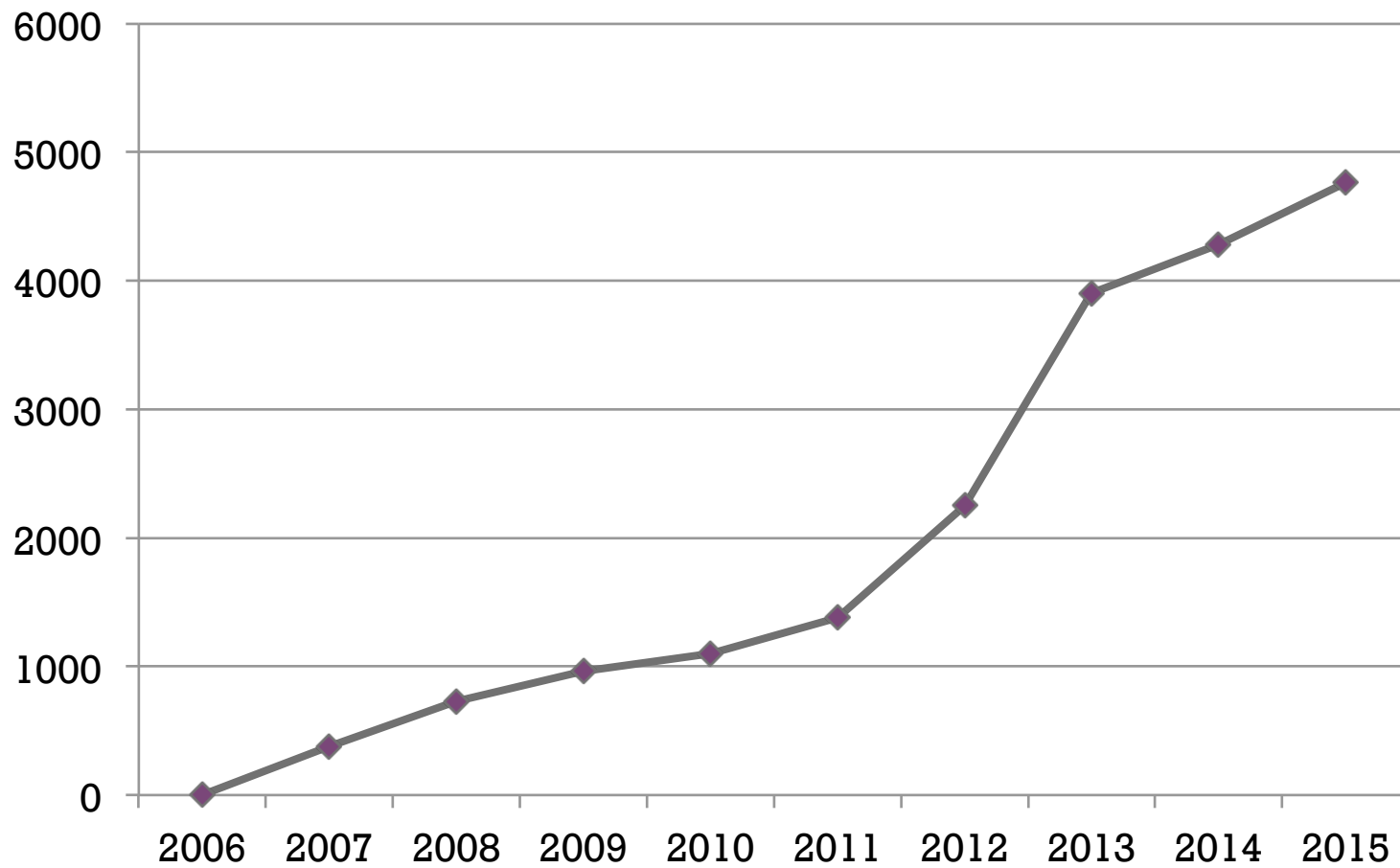
History (2006-2016)



+ The Ramses Project

The corpus

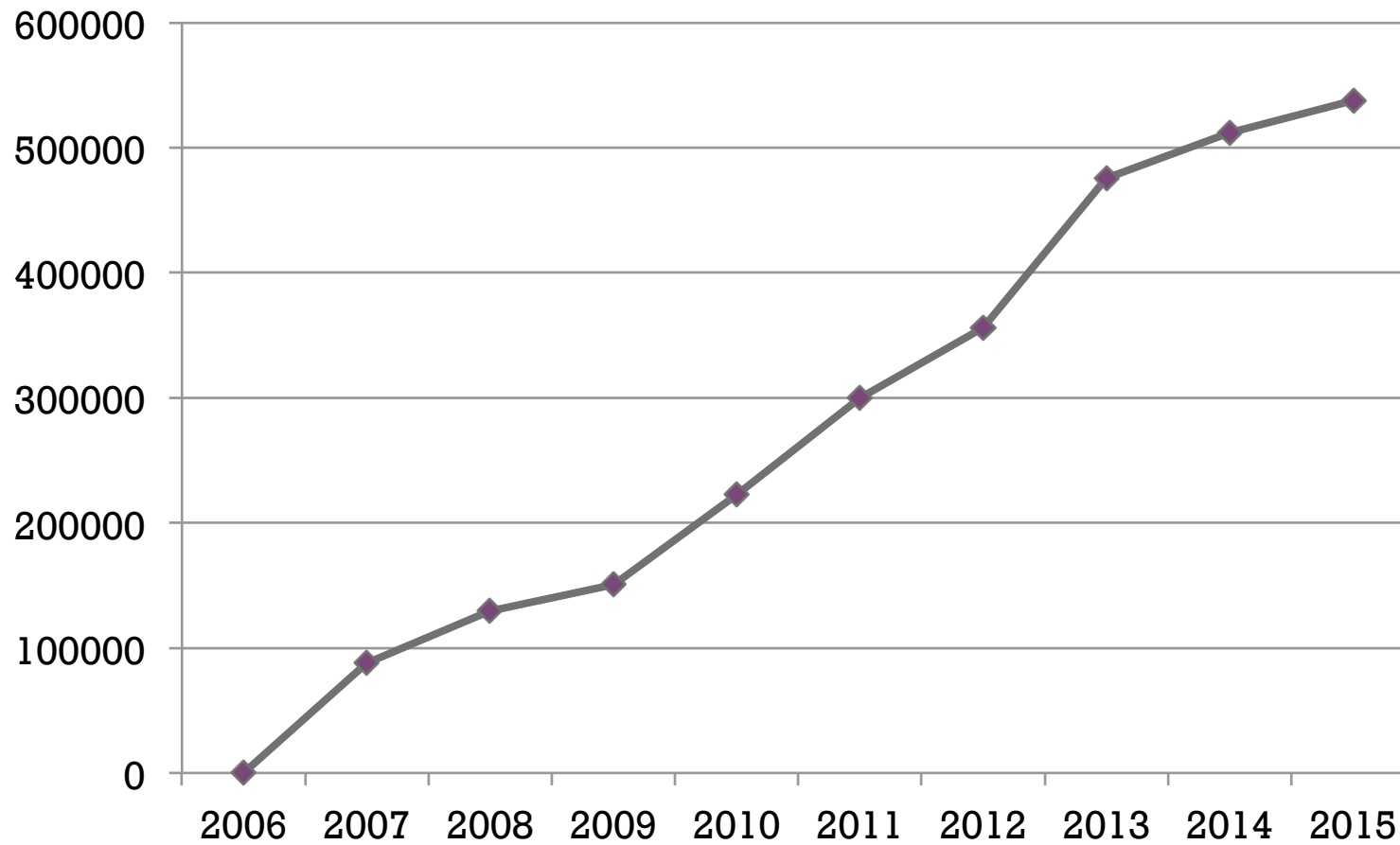
■ Number of witnesses



+ The Ramses Project

The corpus

■ Number of occurrences



+ *Ramses Online* (ramses.ulg.ac.be)

Ramses Online un corpus annoté du néo-égyptien

À propos ▾

Corpus ▾

Recherches ▾

Aide

Version bêta

Enregistrement

Mot de passe oublié

login

Recherche simple dans les textes

Lemmes

Graphies

Traductions

Translittération ou traduction

Filtrer sur

Translittération

Traduction

Les deux



3



(PTCL)

mais, vraiment

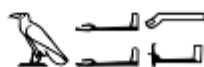
3.t



(SUBST)

moment

3^{cc}



(VB)

enduire, crépir

3^{cc}

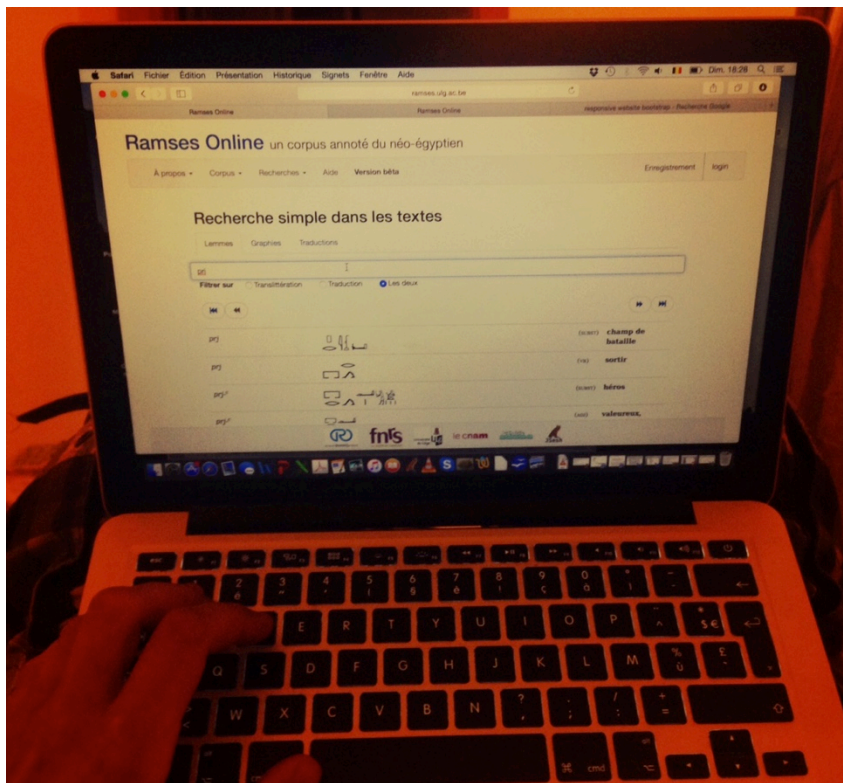


(VB)

faire du tort

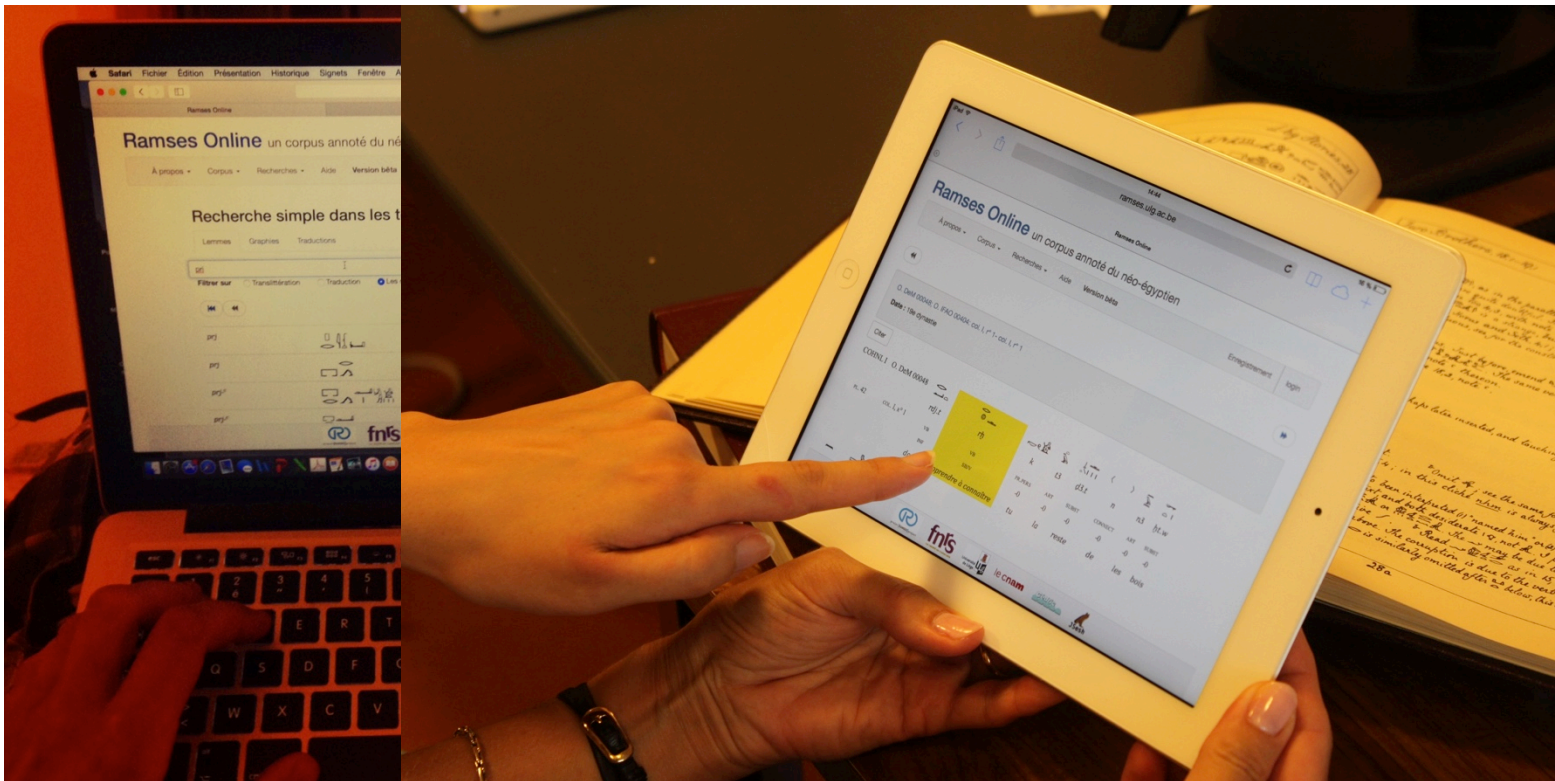
+ *Ramses Online* (ramses.ulg.ac.be)

- Responsive website based on *Bootstrap*



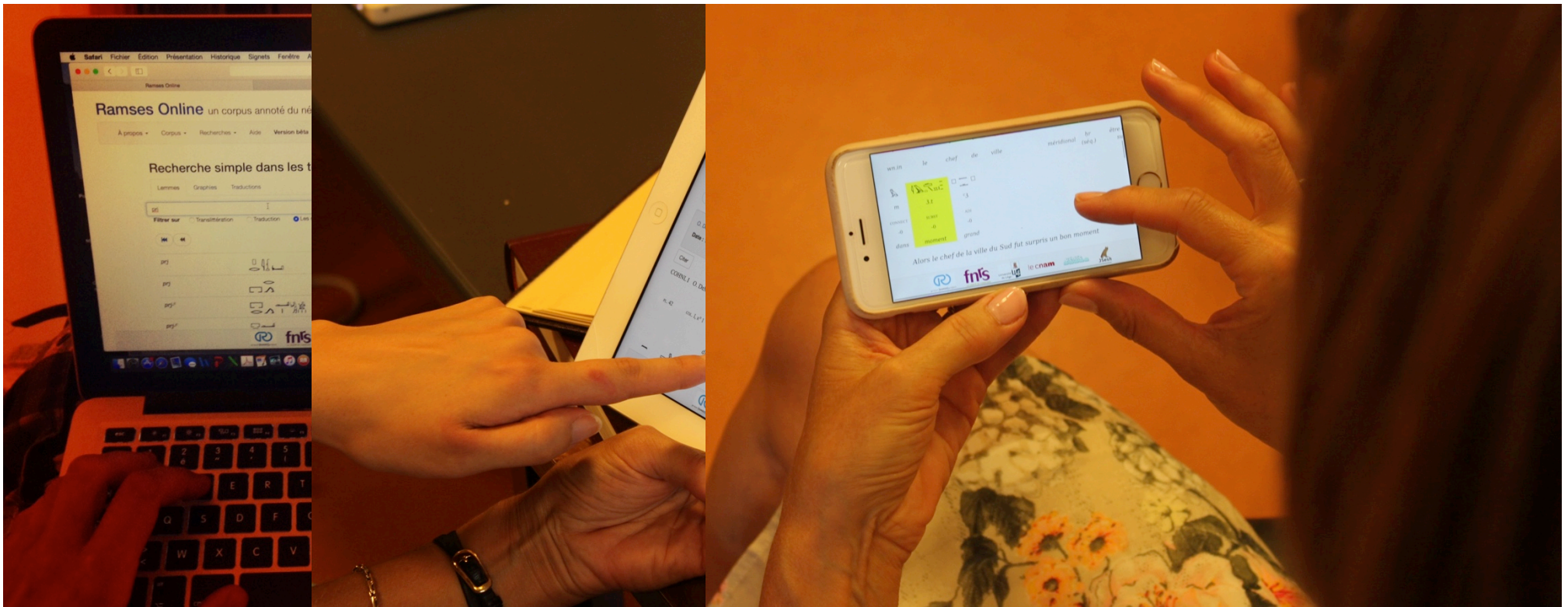
+ *Ramses Online* (ramses.ulg.ac.be)

- Responsive website based on *Bootstrap*



+ *Ramses Online* (ramses.ulg.ac.be)

- Responsive website based on *Bootstrap*



+ *Ramses Online* (ramses.ulg.ac.be)

- Responsive website based on *Bootstrap*
- With powerful linguistic searching capabilities

The screenshot shows a search query in the Ramses Online interface. The query is composed of several elements, each enclosed in a red box:

- A box containing the text "article/demonstratif (Quelconque)".
- A box containing the text "substantif (Quelconque)".
- A box containing an asterisk (*).
- A box containing a wavy line icon, a triangle icon, and the text "qui".
- A box containing an asterisk (*).
- A box containing a lightbulb icon, a vertical line, and the text "Hr (P1)".
- A box containing the text "Et", "verbe (Quelconque)", "Signes", and two icons (a foot and a mountain).
- A box containing a plus sign (+).

Ajouter Lemme

Ajouter Graphie

Ajouter Grammaire

Ajouter Et logique

Ajouter Saut

Effacer

Chercher

+ *Ramses Online* (ramses.ulg.ac.be)

- Responsive website based on *Bootstrap*
- With powerful linguistic searching capabilities
- In interaction with its users

P. Sallier 1 (r° 1,1-3,3); P. BM EA 10185 (r° 1,1-3,3); La querelle d'Apophis et Séqenrenre

Date : Mérenptah

Pour citer ce passage

Aurore Motte, dans *Ramses Online*; référence [legacy/1642/quote/673/78]

Citer



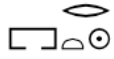


				LES				
<i>jw</i>	<i>f</i>	<i>hr</i>	<i>tm</i>	<i>b3k</i>	<i>n</i>	<i>ntr</i>	<i>nb</i>	<i>nty</i>
GRAM	PR.PERS	GRAM	GRAM	85, 10	CONNECT	SUBST	QUANT	CONNECT
				VB INF				
<i>jw (séq.)</i>	<i>il</i>	<i>hr (séq.)</i>	<i>m (négation)</i>	<i>travailler</i>	<i>à, pour</i>	<i>dieu</i>	<i>tout</i>	<i>qui</i>

+ *Ramses Online* (ramses.ulg.ac.be)

- Responsive website based on *Bootstrap*
- With powerful linguistic searching capabilities
- In interaction with its users



KRI 1 O. DeM 00010

					
	<i>3bd</i>	<i>3</i>	<i>pr.t</i>	<i>sw</i>	<i>10</i>
365, 11	R° 1	SUBST	QUANT	SUBST	QUANT
	<i>mois</i>	<i>3</i>	<i>Péret</i>	<i>jour (du mois), moment</i>	<i>10</i>

3e mois de Péret, 10e jour :



+ Collaborations

■ TEI interchange format

- Laurent Coulon (EPHE – Paris)
- Frederik Elwert (CERES – Bochum)
- Emmanuelle Morlock (HiSoMA – CNRS)
- Stéphane Polis (F.R.S.-FNRS – Liège)
- Vincent Razanajao (ULg – Liège)
- Serge Rosmorduc (CNAM – Paris)
- Simon Schweitzer (BBAW – Berlin)
- Daniel A. Werning (EXC Topoi – Berlin)



TEXT ENCODING INITIATIVE
CONFERENCE AND MEMBERS' MEETING 2015
OCTOBER 28-31, LYON, FRANCE

+ Collaborations

- TEI interchange format
 - Laurent Coulon (EPHE – Paris)
 - Frederik Elwert (CERES – Bochum)
 - Emmanuelle Morlock (HiSoMA – CNRS)
 - Stéphane Polis (F.R.S.-FNRS – Liège)
 - Vincent Razanajao (ULg – Liège)
 - Serge Rosmorduc (CNAM – Paris)
 - Simon Schweitzer (BBAW – Berlin)
 - Daniel A. Werning (EXC Topoi – Berlin)

- Linked metadata
 - Thot (thot.philo.ulg.ac.be)



THOT - THESAURI & ONTOLOGY
for documenting Ancient Egyptian Resources BETA

HOME THESAURI TEI INTERCHANGE FORMAT ABOUT THIS WEBSITE

Thanks to the [Anneliese Maier Research Award](#) granted to Prof. Jean Winand by the Humboldt Foundation, the [Department of Egyptology](#) of the University of Liege, in collaboration with the Berlin-Brandenburg Academy of Sciences and the Saxon Academy of Sciences in Leipzig ([Thesaurus Linguae Aegyptiae](#)), is now developing Thot, a set of resources for documenting and encoding ancient Egyptian resources in a shared, interoperable approach. Moving towards Web Semantic and Linked Open Data, the resources that will be pushed online progressively will consist in a set of thesauri covering most of the metadata relating to ancient Egyptian texts and monuments, as well as a proposal for a TEI interchange format that enables exchange and sharing of textual data.

[Access Thot Thesauri](#)

Thesauri
A first set of resources will consist in a wide range of multilingual thesauri related to documentary and textual metadata, aiming at providing controlled vocabularies and terms so that Egyptological projects with text and object focus will interoperate with each other in an easier way ([read more](#)).

TEI exchange format
Sharing digital textual resources is an actual challenge for scholars working on Ancient Egyptian-Coptic (3000 BC-1350 AD), mainly because of the different writing systems that have been used throughout the history of this language, and because of the diverging aims and scopes of the projects involved in creating annotated corpora of Ancient Egyptian-Coptic. In collaboration with several leading projects in this field, Thot is coordinating the definition of a XML/TEI interchange format in order to enable this data sharing ([read more](#)).

powered by

version: 0.6-20160905

+ Collaborations

- TEI interchange format
 - Laurent Coulon (EPHE – Paris)
 - Frederik Elwert (CERES – Bochum)
 - Emmanuelle Morlock (HiSoMA – CNRS)
 - Stéphane Polis (F.R.S.-FNRS – Liège)
 - Vincent Razanajao (ULg – Liège)
 - Serge Rosmorduc (CNAM – Paris)
 - Simon Schweitzer (BBAW – Berlin)
 - Daniel A. Werning (EXC Topoi – Berlin)

- Linked metadata
 - Thot (thot.philo.ulg.ac.be)
 - Trismegistos (<http://trismegistos.org/>)



THOT - THESAURI & ONTOLOGY
for documenting Ancient Egyptian Resources BETA

HOME THESAURI TEI INTERCHANGE FORMAT ABOUT THIS WEBSITE

Thanks to the [Anneliese Maier Research Award](#) granted to Prof. Jean Winand by the Humboldt Foundation, the Department of Egyptology of the University of Liège, in collaboration with the Berlin-Brandenburg Academy of Sciences and the Saxon Academy of Sciences in Leipzig ([Thesaurus Linguae Aegyptiae](#)), is now developing Thot, a set of resources for documenting and encoding ancient Egyptian resources in a shared, interoperable approach. Moving towards Web Semantic and Linked Open Data, the resources that will be pushed online progressively will consist in a set of thesauri covering most of the metadata relating to ancient Egyptian texts and monuments, as well as a proposal for a TEI interchange format that enables exchange and sharing of textual data.

[Access Thot Thesauri](#)

Thesauri

A first set of resources will consist in a wide range of multilingual thesauri related to documentary and textual metadata, aiming at providing controlled vocabularies and terms so that Egyptological projects with text and object focus will interoperate with each other in an easier way ([read more](#)).

TEI exchange format

Sharing digital textual resources is an actual challenge for scholars working on Ancient Egyptian-Coptic (3000 BC-1350 AD), mainly because of the different writing systems that have been used throughout the history of this language, and because of the diverging aims and scopes of the projects involved in creating annotated corpora of Ancient Egyptian-Coptic. In collaboration with several leading projects in this field, Thot is coordinating the definition of a XML/TEI interchange format in order to enable this data sharing ([read more](#)).



version:
0.6-
20160905

TRISMEGISTOS

An interdisciplinary portal of papyrological and epigraphical resources
formerly Egypt and the Nile valley (800 BC-AD 800), now **expanding** to the Ancient World in general

[How to cite TM](#)

TEXTS	COLLECTIONS	ARCHIVES	PEOPLE	PLACES	AUTHORS
-------	-------------	----------	--------	--------	---------

+ Collaborations

- TEI interchange format
 - Laurent Coulon (EPHE – Paris)
 - Frederik Elwert (CERES – Bochum)
 - Emmanuelle Morlock (HiSoMA – CNRS)
 - Stéphane Polis (F.R.S.-FNRS – Liège)
 - Vincent Razanajao (ULg – Liège)
 - Serge Rosmorduc (CNAM – Paris)
 - Simon Schweitzer (BBAW – Berlin)
 - Daniel A. Werning (EXC Topoi – Berlin)

- Linked metadata
 - Thot (thot.philo.ulg.ac.be)
 - Trismegistos (<http://trismegistos.org/>)
 - TopBib (<http://topbib.griffith.ox.ac.uk/>)



THOT - THESAURI & ONTOLOGY
for documenting Ancient Egyptian Resources BETA

HOME THESAURI TEI INTERCHANGE FORMAT ABOUT THIS WEBSITE

Thanks to the [Anneliese Maier Research Award](#) granted to Prof. Jean Winand by the Humboldt Foundation, the Department of Egyptology of the University of Liège, in collaboration with the Berlin-Brandenburg Academy of Sciences and the Saxon Academy of Sciences in Leipzig (Thesaurus Linguae Aegyptiae), is now developing Thot, a set of resources for documenting and encoding ancient Egyptian resources in a shared, interoperable approach. Moving towards Web Semantic and Linked Open Data, the resources that will be pushed online progressively will consist in a set of thesauri covering most of the metadata relating to ancient Egyptian texts and monuments, as well as a proposal for a TEI interchange format that enables exchange and sharing of textual data. [Access Thot Thesauri](#)

Thesauri
A first set of resources will consist in a wide range of multilingual thesauri related to documentary and textual metadata, aiming at providing controlled vocabularies and terms so that Egyptological projects with text and object focus will interoperate with each other in an easier way ([read more](#)).

TEI exchange format
Sharing digital textual resources is an actual challenge for scholars working on Ancient Egyptian-Coptic (3000 BC-1350 AD), mainly because of the different writing systems that have been used throughout the history of this language, and because of the diverging aims and scopes of the projects involved in creating annotated corpora of Ancient Egyptian-Coptic. In collaboration with several leading projects in this field, Thot is coordinating the definition of a XML/TEI interchange format in order to enable this data sharing ([read more](#)).

powered by version: 0.6-20160905

TRISMEGISTOS
An interdisciplinary portal of papyrological and epigraphical resources
formerly Egypt and the Nile valley (800 BC-AD 800), now **expanding** to the Ancient World in general

[How to cite TM](#)

TEXTS COLLECTIONS ARCHIVES PEOPLE PLACES AUTHORS



Evolution of the data model

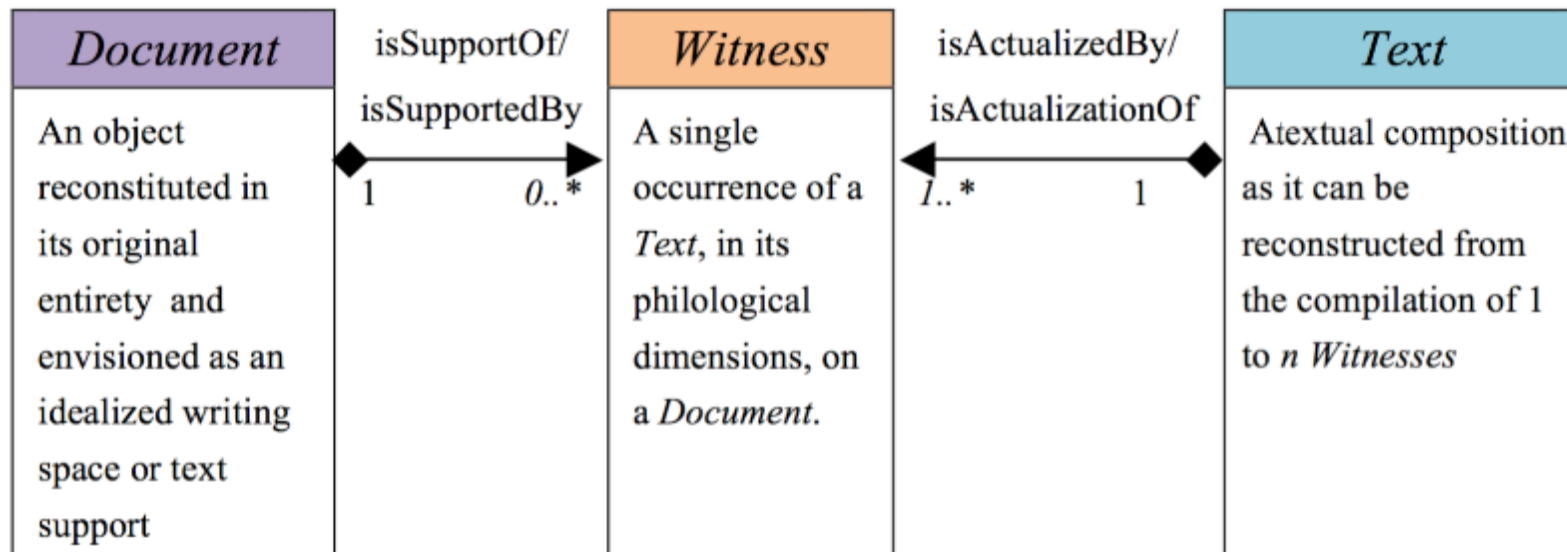
Between 2006 and 2016

+ Evolution of the data model

- The original model (2006)
- The Thot Data Model (TDM – 2016)
- Towards TDM 2.0

+ Evolution of the data model

- The original model (2006)



Following the egyptological practice, the decision was made to encode in hieroglyphs (and to annotate) every single witness of each text (envisioned as an abstraction). The document, on the other hand, was seen as the object on which multiple witnesses could occur.

⇒ **No ideal/intended text, rather individual scribes as agents**

+ Evolution of the data model



- The original model (2006)

ፊደላት ለግንባር ለግንባር ለግንባር ለግንባር ለግንባር	ፊደላት ለግንባር ለግንባር ለግንባር ለግንባር ለግንባር	R
— — —	— — —	An
ፊደላት ለግንባር ለግንባር ለግንባር ለግንባር ለግንባር	ፊደላት ለግንባር ለግንባር ለግንባር ለግንባር ለግንባር	AOS
— — —	— — —	G
ፊደላት ለግንባር ለግንባር ለግንባር ለግንባር ለግንባር	ፊደላት ለግንባር ለግንባር ለግንባር ለግንባር ለግንባር	C
ፊደላት ለግንባር ለግንባር ለግንባር ለግንባር ለግንባር	ፊደላት ለግንባር ለግንባር ለግንባር ለግንባር ለግንባር	B3

Synoptic edition of *Sinuhe* (Koch 1990: 8,2-7)

+ Evolution of the data model

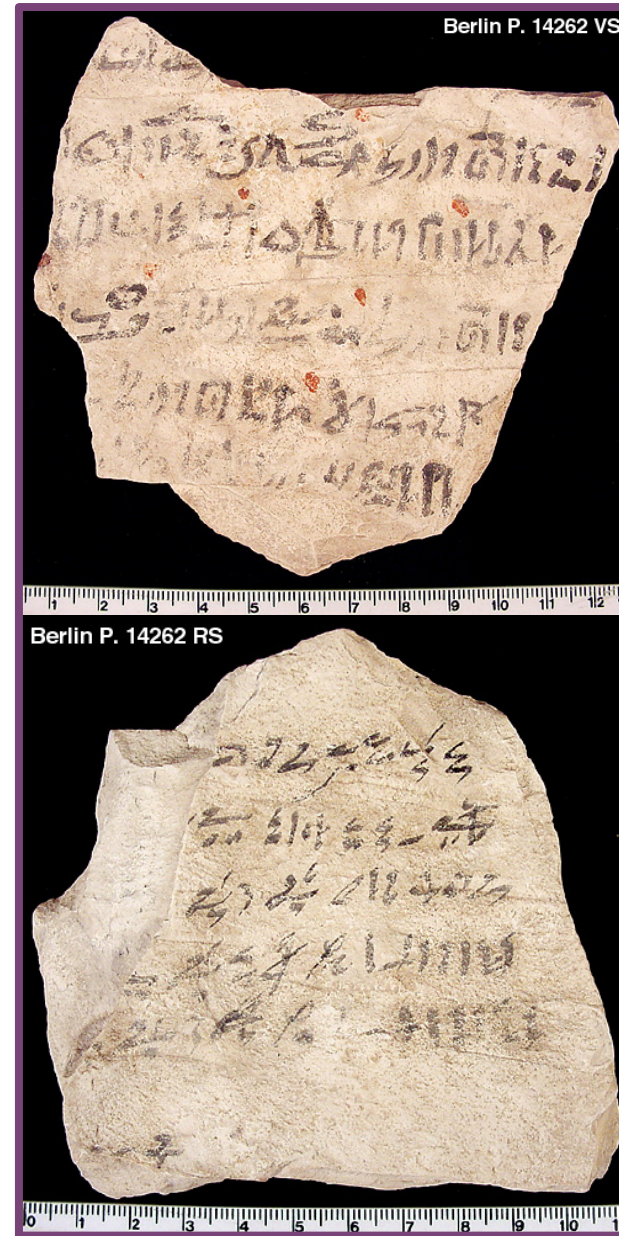
- The original model (2006)

Witness 1	←		R
Witness 2	←		An
Witness 3	←		AOS
Witness 4	←		G
Witness 5	←		C
Witness 6	←		B3

Synoptic edition of *Sinuhe* (Koch 1990: 8,2-7)

+ Evolution of the data model

- The original model (2006)

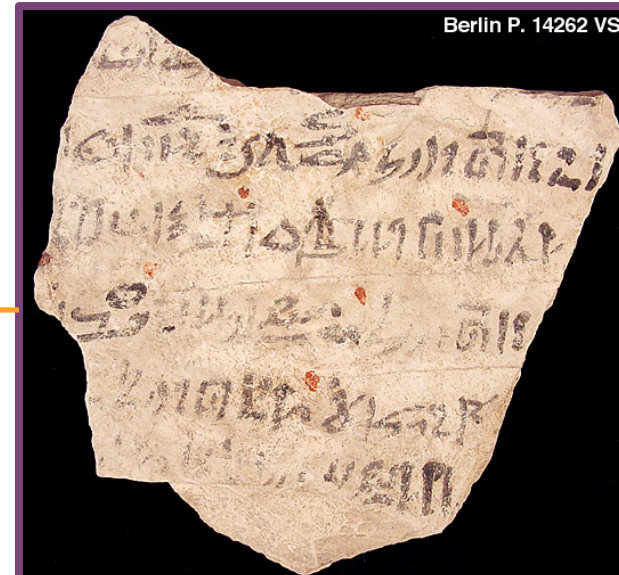


One document

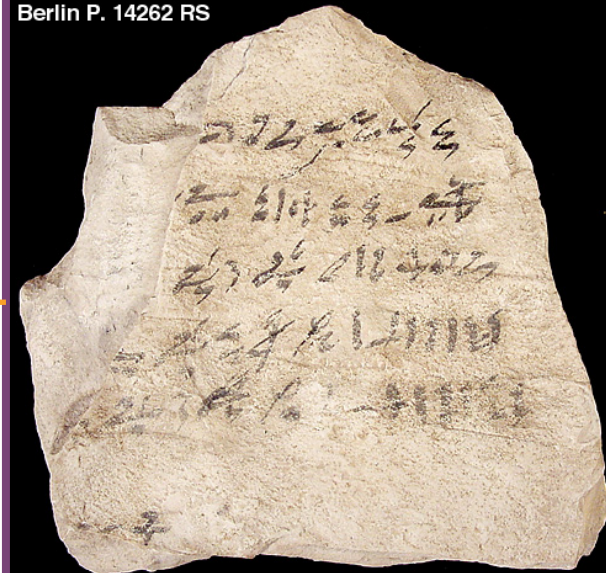
+ Evolution of the data model

- The original model (2006)

Witness 1
Literary text
(hymn)



Berlin P. 14262 RS



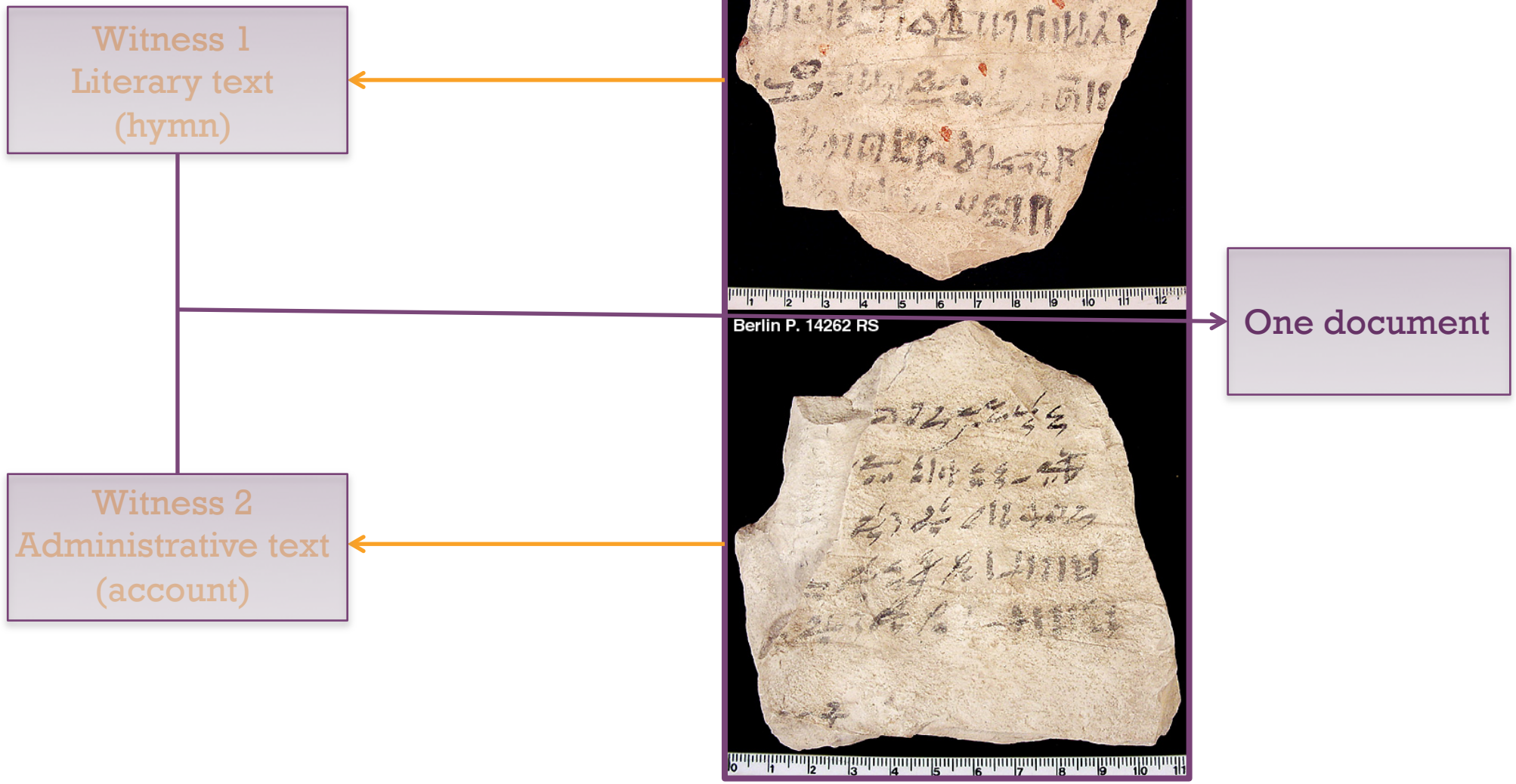
Witness 2
Administrative text
(account)



One document

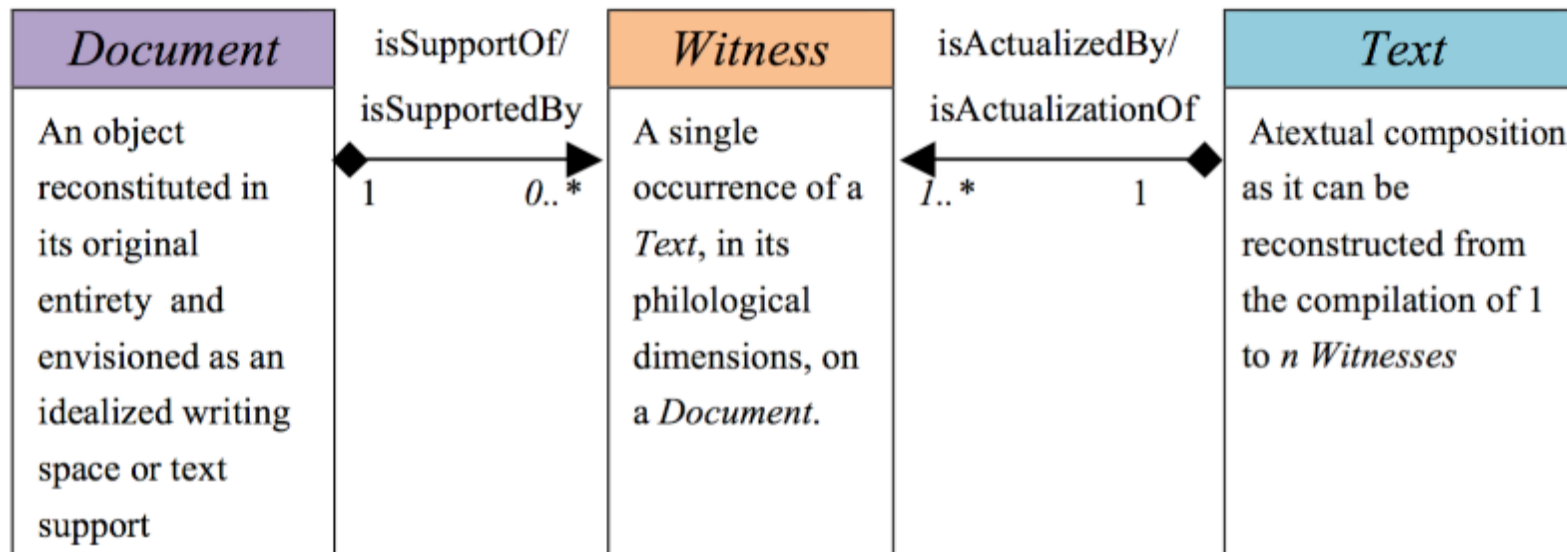
+ Evolution of the data model

- The original model (2006)



+ Evolution of the data model

- The original model (2006)



“The [documentary and textual] dimensions are incapable of disconnection: (...) they negatively constitute one another and (...) this constitution requires human agency at every step, from composition to reception.”

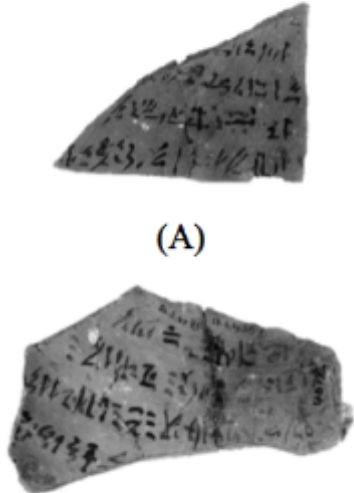
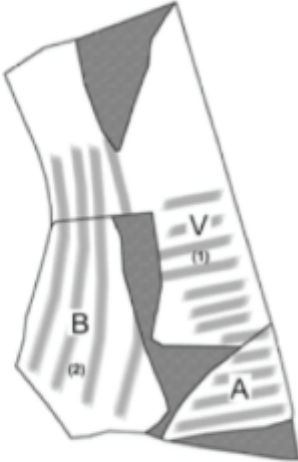
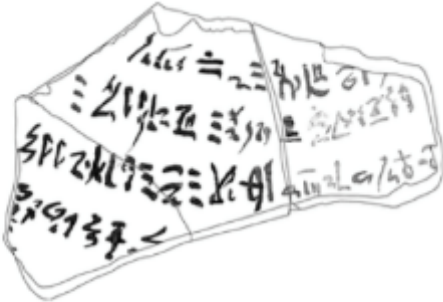

(Eggert 2004: 169; see already Eggert 1998: 57)

+ Evolution of the data model

- The original model (2006)
- The Thot Data Model (TDM – 2016) & material philology

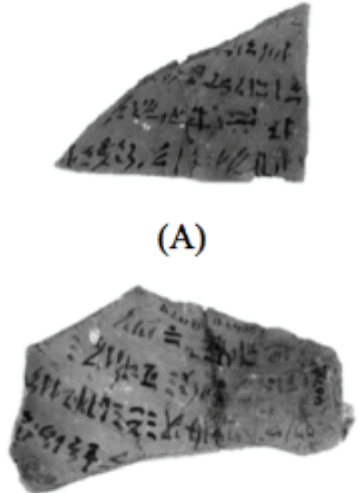
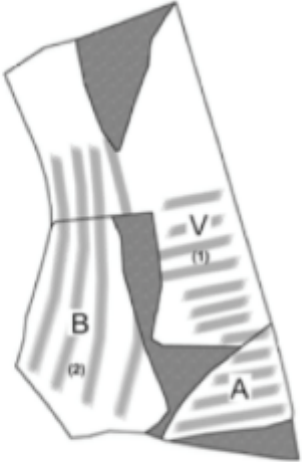
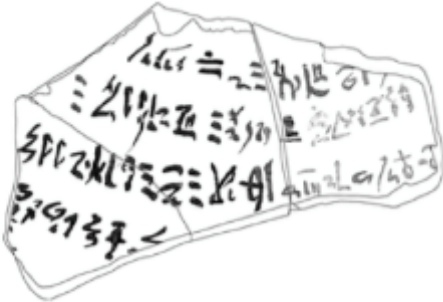
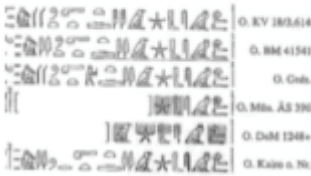
+ Evolution of the data model

- The original model (2006)
- The Thot Data Model (TDM – 2016) & material philology

<i>Object(s)</i>	<i>Document</i>	<i>Witness(es)</i>	<i>Text</i>
 <p>(A)</p> <p>(B)</p>	 <p>V (1)</p> <p>B (2)</p> <p>A</p>		 <p>O. KV 182.614</p> <p>O. BM 41541</p> <p>O. Onb.</p> <p>O. Mn. A3 390</p> <p>O. Dm 1248+</p> <p>O. Kairo n. 30.</p>
O. IFAO 1255 A (v ^o) + O. IFAO 1255 B (v ^o) + (O. Varille 39 [= V]; lost)	Reconstructed original potsherd ostracon based on the three <i>Objects</i>	Occurrence of the <i>Teaching of Amennakhte</i> (§39-48) on O. IFAO 1255 B (v ^o) [+ V <i>lost</i>]	Abstract version of the <i>Teaching of Amennakhte</i> , viewed here as a Synoptic edition

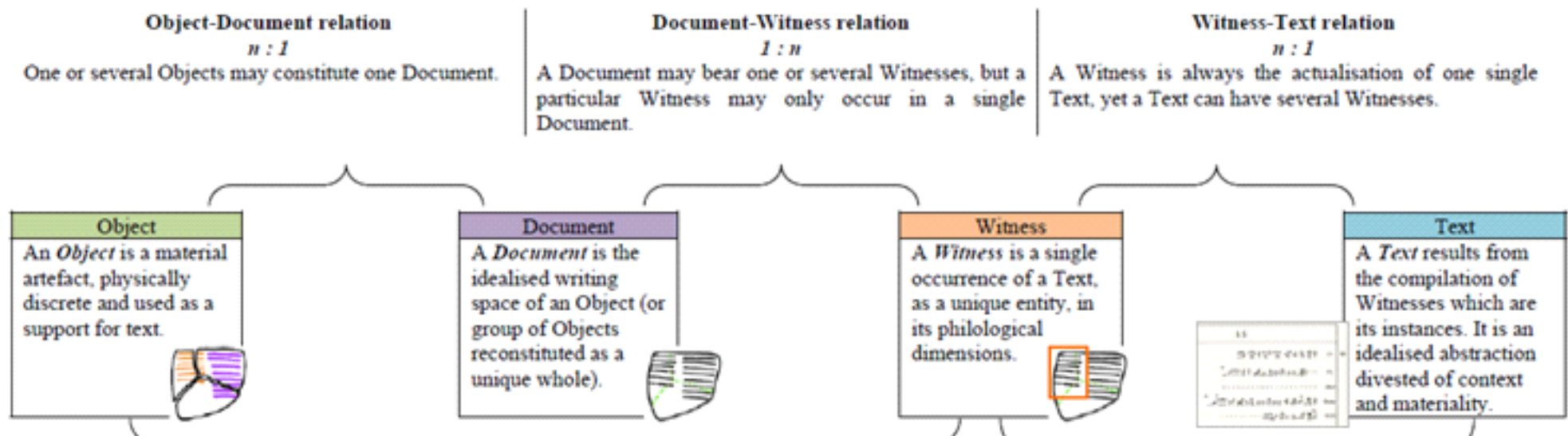
+ Evolution of the data model

- The original model (2006)
- The Thot Data Model (TDM – 2016) & material philology

<i>Object(s)</i>	<i>Document</i>	<i>Witness(es)</i>	<i>Text</i>
 <p>(A)</p> <p>(B)</p>			
<p>O. IFAO 1255 A (v⁰) + O. IFAO 1255 B (v⁰) + (O. Varille 39 [= V]; lost)</p>	<p>Reconstructed original potsherd ostracon based on the three <i>Objects</i></p>	<p>Occurrence of the <i>Teaching of Amennakhte</i> (§39-48) on O. IFAO 1255 B (v⁰) [+ V <i>lost</i>]</p>	<p>Abstract version of the <i>Teaching of Amennakhte</i>, viewed here as a Synoptic edition</p>

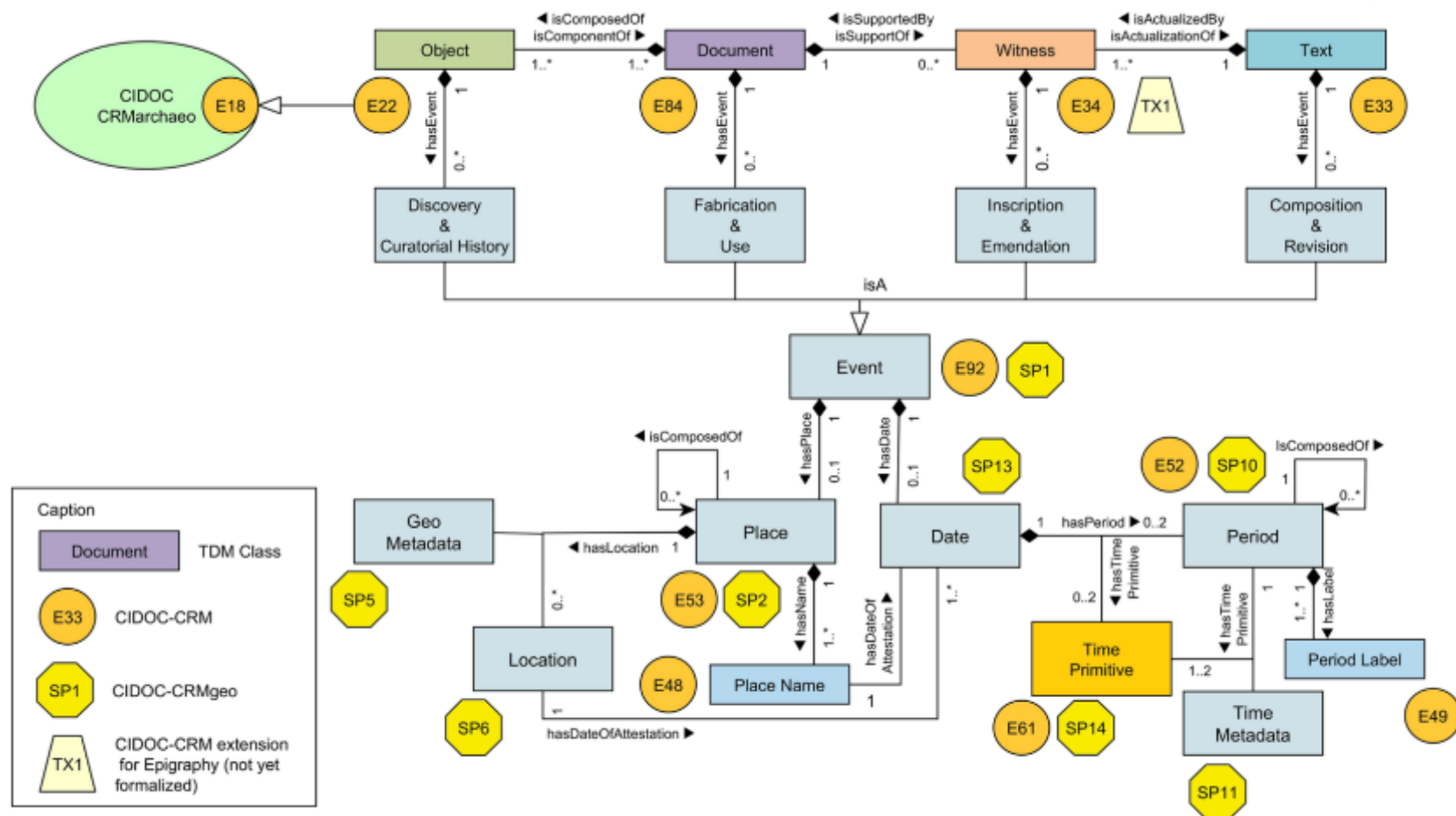
+ Evolution of the data model

- The original model (2006)
- The Thot Data Model (TDM – 2016) & material philology
 - St. Polis & V. Razanajao (2016), Ancient Egyptian texts in contexts. Towards a conceptual data model (the Thot Data Model – TDM), to appear in *Bulletin of the Institute of Classical Studies* (special issue: ‘Digital approaches to the Ancient World’).



+ Evolution of the data model

- The original model (2006)
- The Thot Data Model (TDM – 2016) & material philology



+ Evolution of the data model

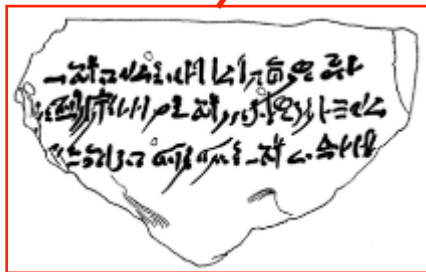
- The original model (2006)
- The Thot Data Model (TDM – 2016) & material philology
- Towards TDM 2.0
 - Why?

+ Evolution of the data model

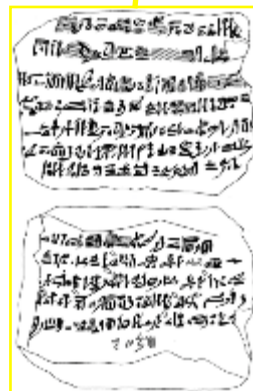
P. Mag. Isis (P. CGT 54051) = the main Witness of the Text (*Isis magical papyrus*)



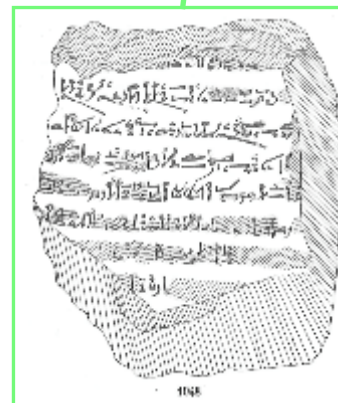
Other Witnesses of the Text, occurring on different Documents (ostraca)



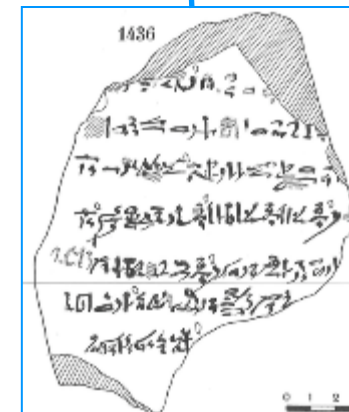
O. Gardiner



O. Nash 14



O. DeM 1048



O. DeM 1046

+ Evolution of the data model

P. Mag. Isis (P. CGT 54051 r^o)



P. Chester Beatty XI (r^o)



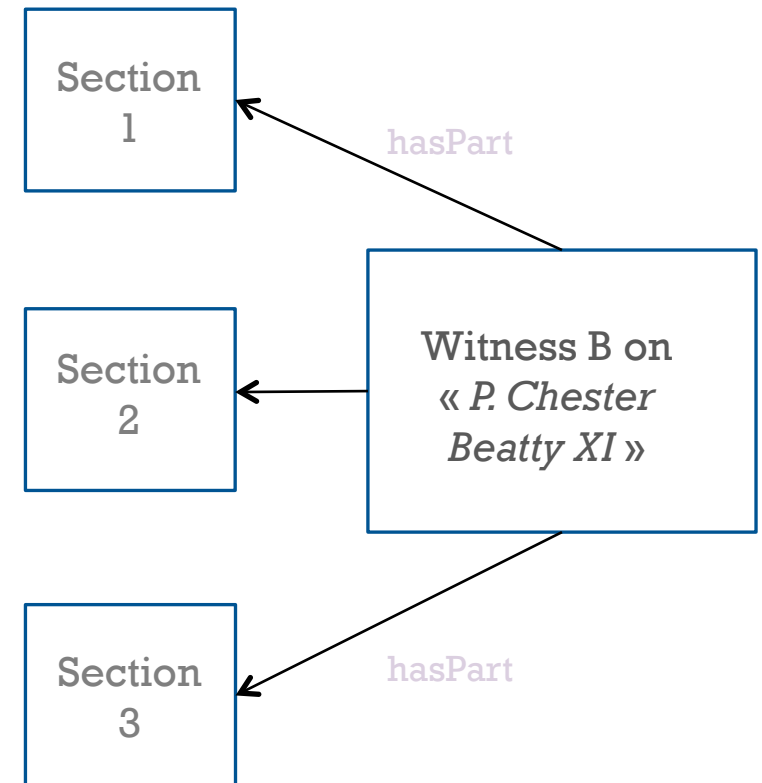
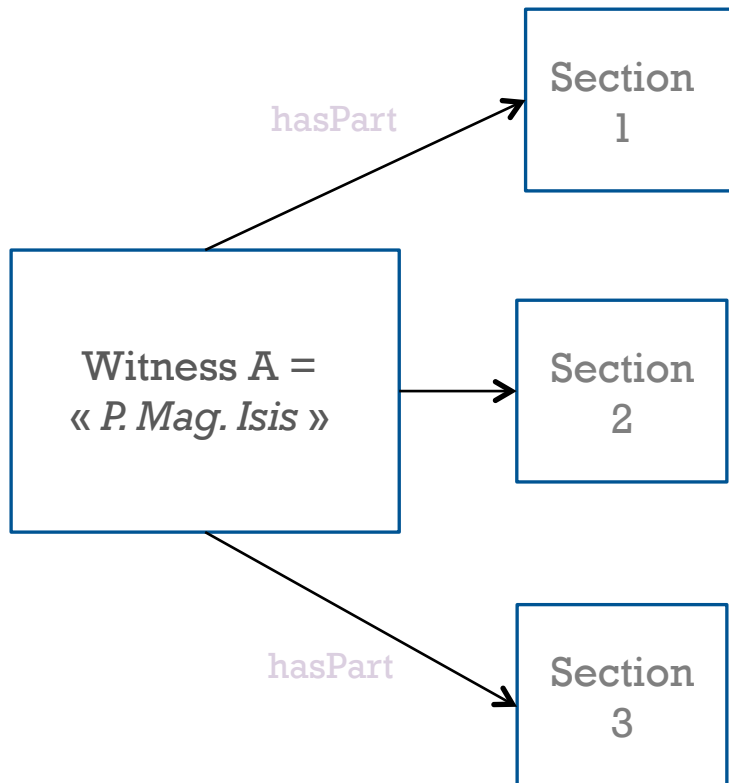
P. CGT 54051 is composed of

- A group of incantations (r^o 2,1–8)
- A formula for bandaging (r^o 2,8–12)
- **The story of Isis & Ra** (r^o 2,12 – 5,5)
- A group of formulae (r^o 5,12 – 6,3)
- A group of formulae (v^o 2,6 – 5,1)

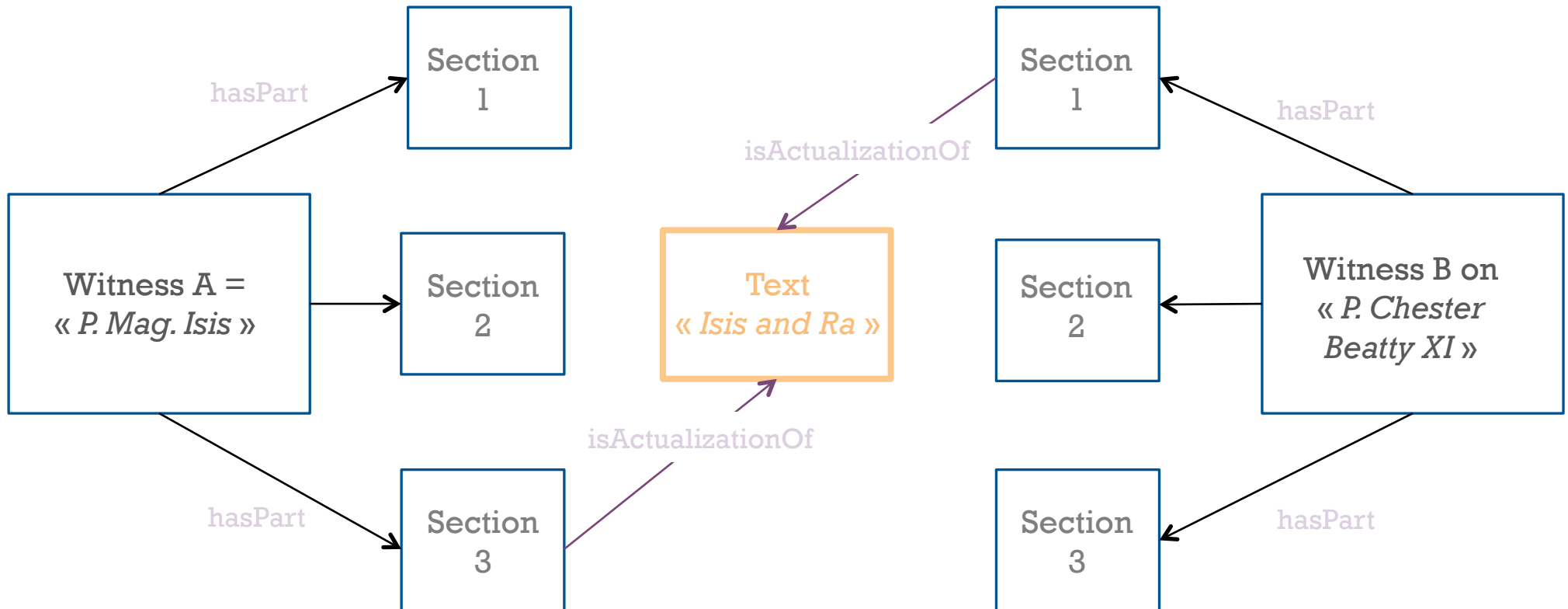
P. Chester Beatty XI is composed of

- **The story of Isis & Ra** (r^o 1,1 – 4,1)
- A group of magical spells (r^o 4,2–10)
- A magical text against scorpions (r^o, fgmts A–L)
- The end of a magical text (v^o 1,1–3)
- An account (v^o 1,4–10)
- An hymn to the god Amun (v^o 2–3)
- A group of formulae for “safety upon the river” (v^o, fgmts A–D)

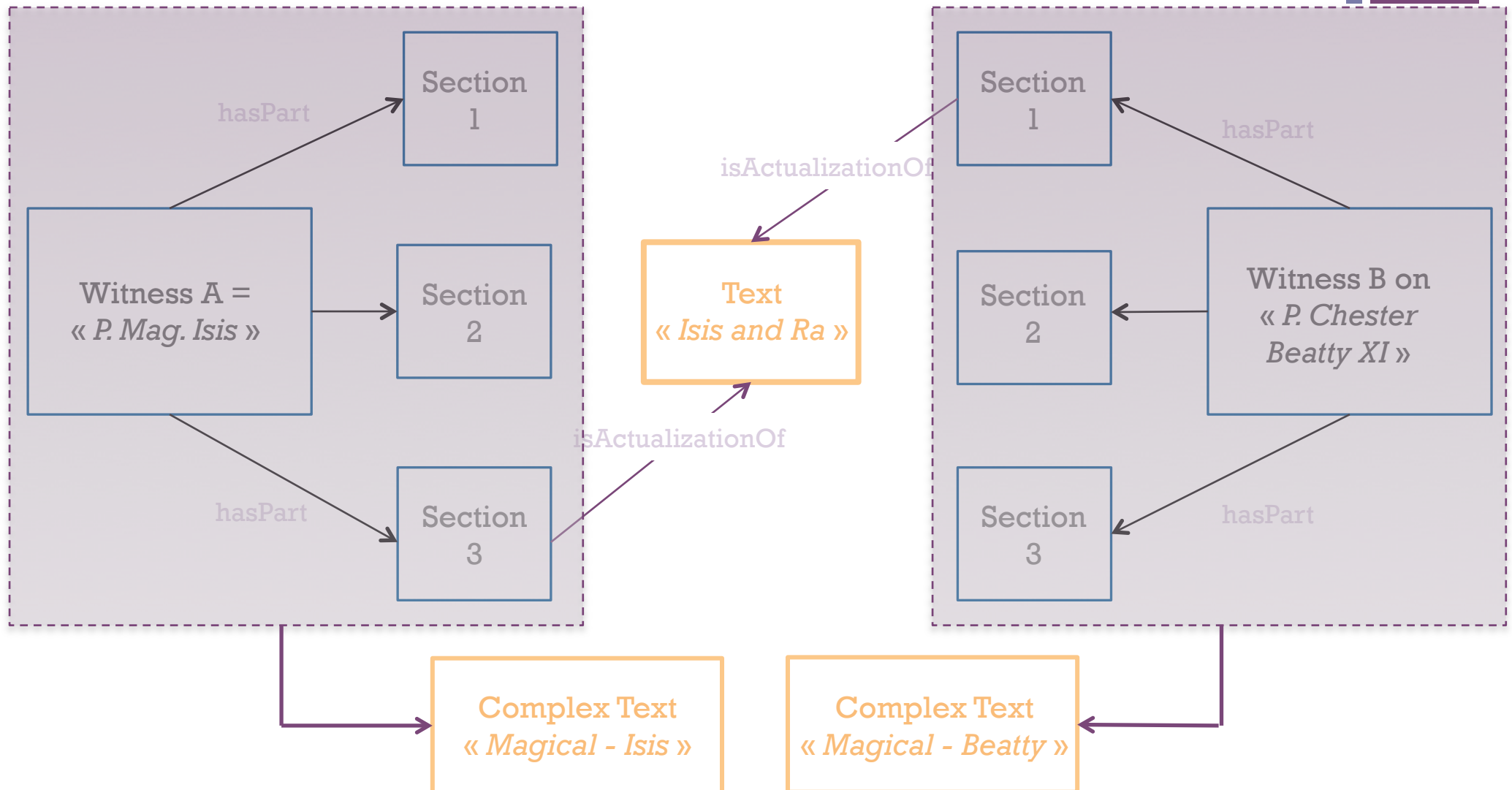
+ Evolution of the data model



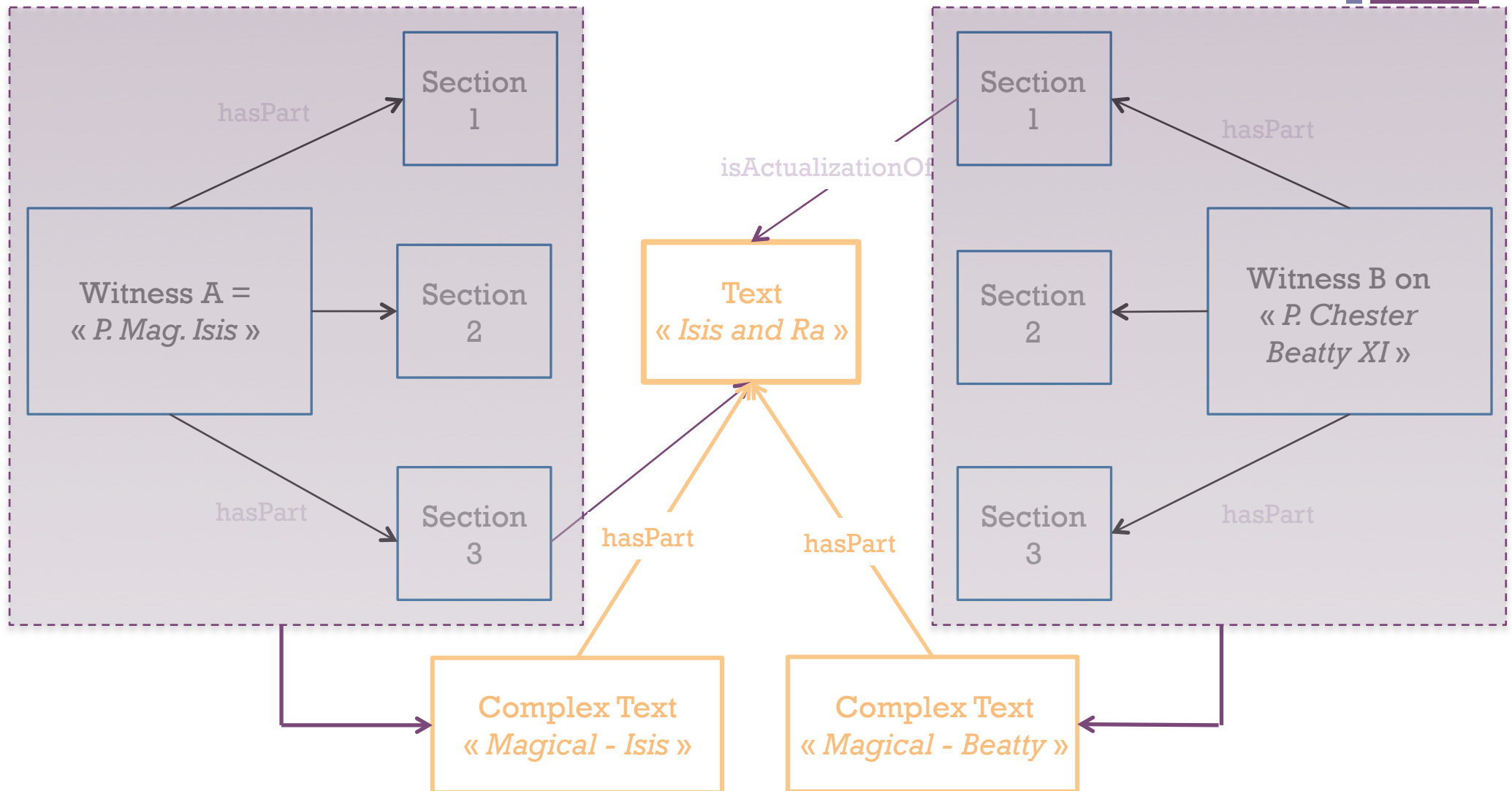
+ Evolution of the data model



+ Evolution of the data model

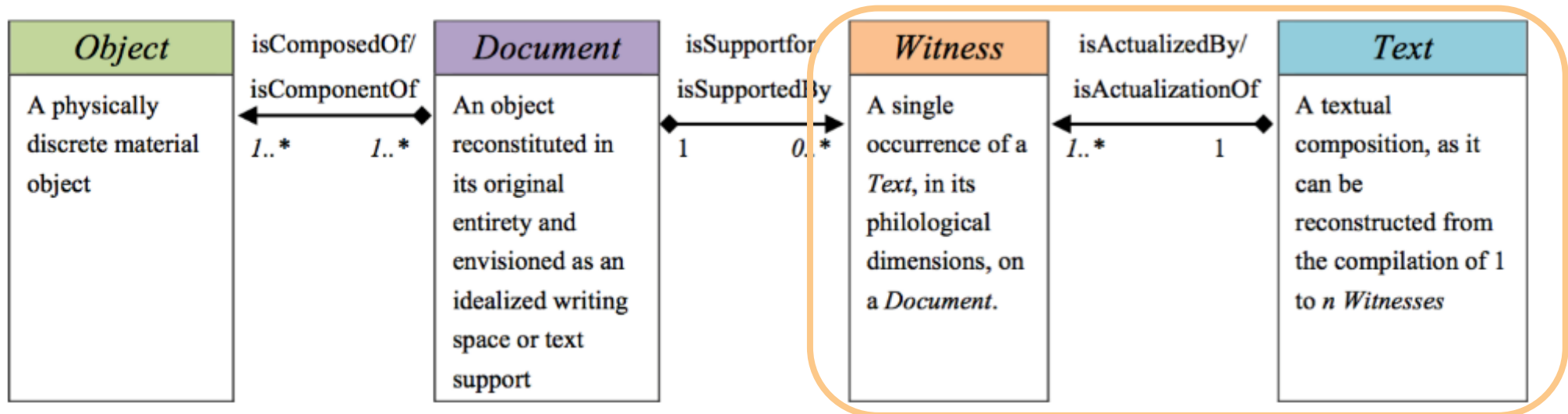


+ Evolution of the data model



+ Evolution of the data model

- The original model (2006)
- The Thot Data Model (TDM – 2016)
- Towards TDM 2.0
 - Why?
 - With which model can we handle such cases?
 - Relationships between witnesses
 - Relationships between texts



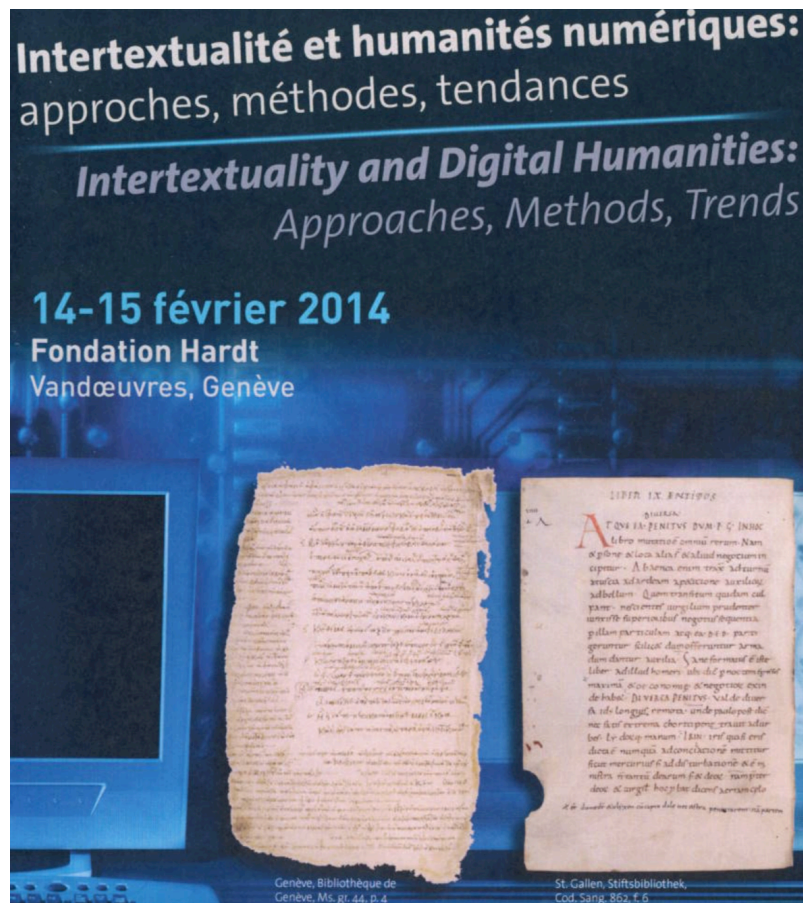


Data models

State-of-the-art in digital editing

+ Data models for digital editing

- 'Digital approaches to intertextuality' is a hot topic



+ Data models for digital editing

- 'Digital approaches to intertextuality' is a hot topic



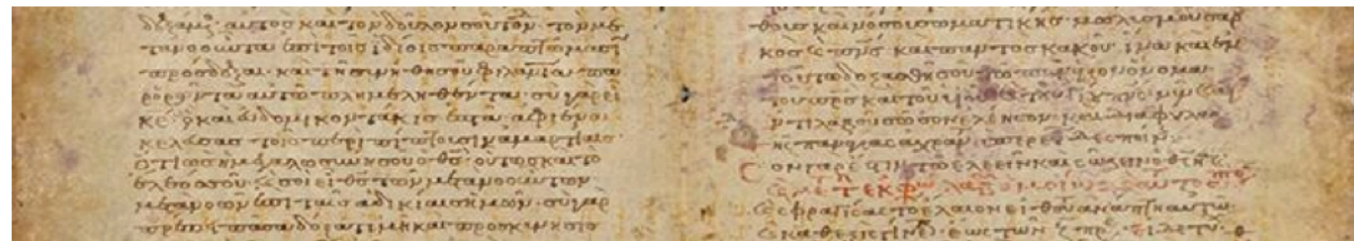
Call for Contribution - Special Issue

2016

Call for Contribution

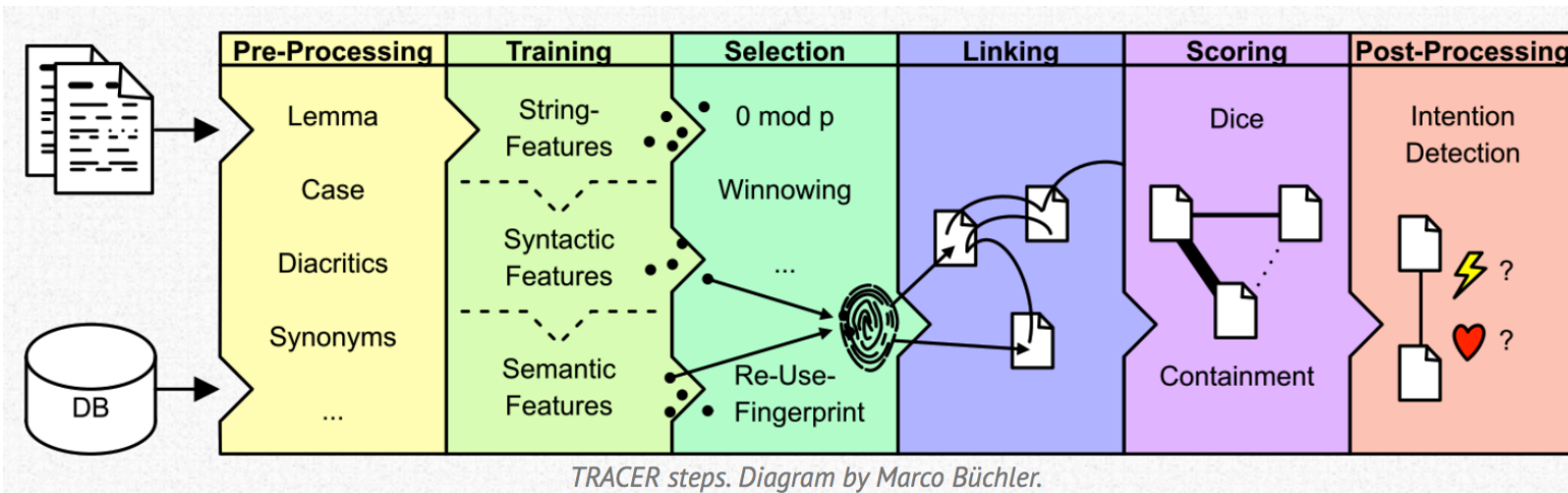
Special Issue on

Computer-Aided Processing of Intertextuality in Ancient Languages



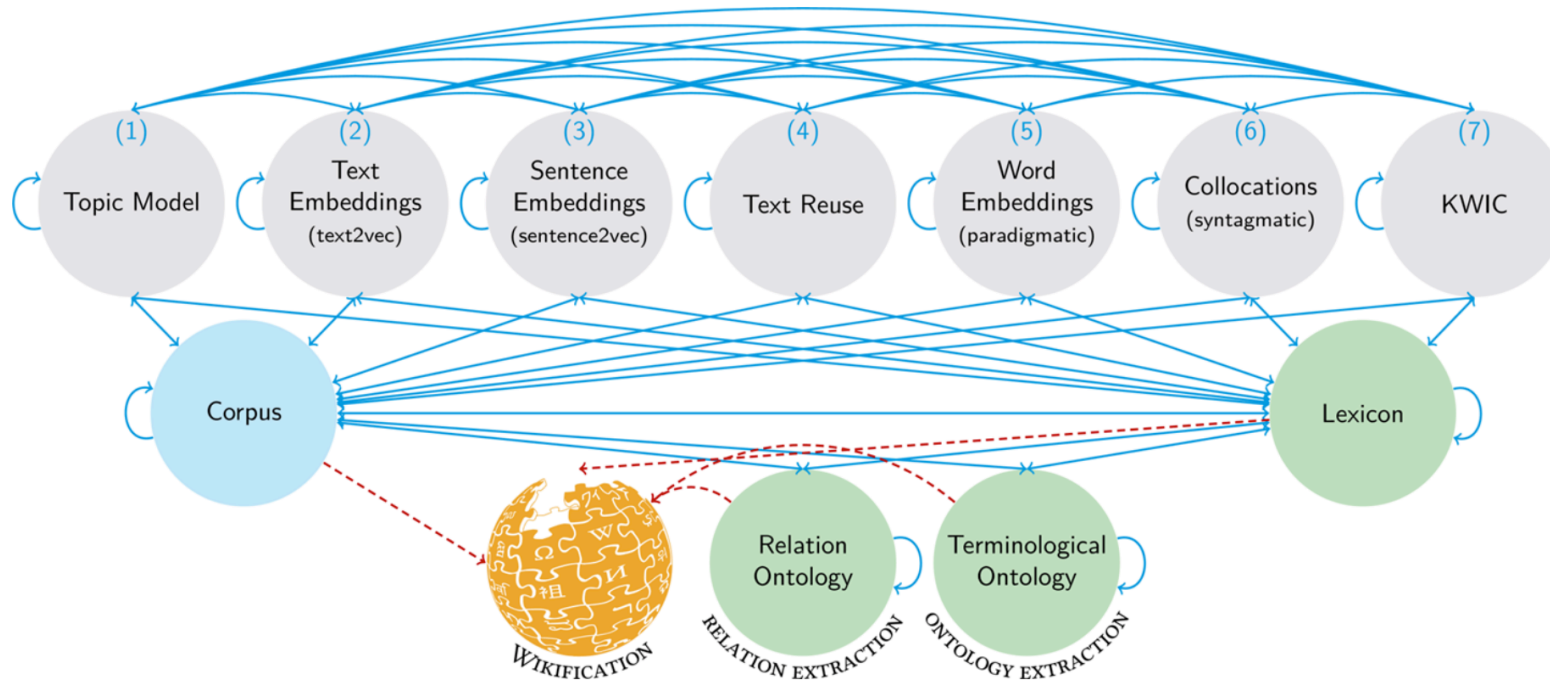
+ Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (NLP)



+ Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)



+ Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)



text-pair

PAIR: Pairwise Alignment for Intertextual Relations

PAIR (Pairwise Alignment for Intertextual Relations) is a simple implementation of a [sequence alignment](#) algorithm for human text analysis designed to identify "similar passages" in large collections of texts. These may include direct quotations, plagiarism and other forms of borrowings, commonplace expressions and the like. We are developing two currently streams:

PhiloLine (for PhiloLogic alignment) is the experimental model, written largely by Mark (in very old fashioned perl), designed to perform all-against-all comparisons in documents loaded into a PhiloLogic database. An entire corpus is indexed and compared against itself, or another database, to find text reuse.

March 2009: We have released version 0e of PhiloLine. To get some sense of what this does and how it currently works, please consult our [ARTFL-Frantext Release Notes](#). We have put this into production for ARTFL and are using it for a number of other projects. We have fairly extensive installation and use instructions in our [Release Notes](#). This is known to work for English, French, [Greek](#), and other language document collections up to 11,500 documents and 520 million words.



text-pair

+ Data models for digital editing

55

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)



Latin Search

The Tesseract project aims to provide a flexible and robust web interface for exploring intertextual parallels. Select two poems below to see a list of lines sharing two or more words (regardless of inflectional changes).

For an overview of all advanced features, see the [instructions](#) page.

Source:

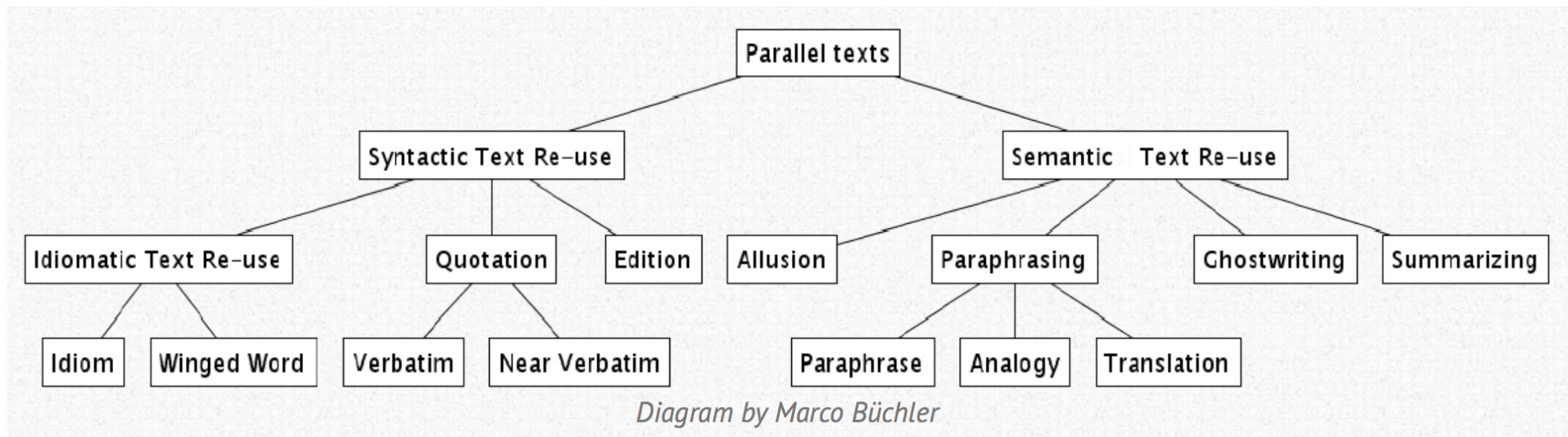
Target:

show advanced



+ Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)



+ Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework

+ Data models for digital editing

- 'Digital approaches to intertextuality' is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework



+ Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework



The text is seen as
“a mosaic of
quotations; any text
is the absorption and
transformation of
another” (1986: 37)

+ Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework



“any text is a new tissue of past citations. Bits of code, formulae, rhythmic models, fragments of social languages, etc., pass into the text and are redistributed within it, for there is always language before and around the text” (1981: 39)

+ Data models for digital editing

- 'Digital approaches to intertextuality' is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework



+ Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework
 - G. Genette (1997 [1982]), *Palimpsests: Literature in the second degree*
 - Intertextuality
 - Paratextuality
 - Metatextuality
 - Hypertextuality
 - Architextuality



+ Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework
- Data models
 - “In the context of digital editing, by modelling we mean at least two types of conceptualization: **the one that tries to organize entities such as texts, documents, works, along with their relationships** and how they have happened to come into being, and the analytical process of establishing the kind and purpose for the production of a new edition, its implied community of users and what features best represent their various needs.” (Pierazzo 2015: 48)

+ Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework
- Data models
 - The FRBR data model (1997 [2009])

**FUNCTIONAL REQUIREMENTS
FOR
BIBLIOGRAPHIC RECORDS**

Final Report

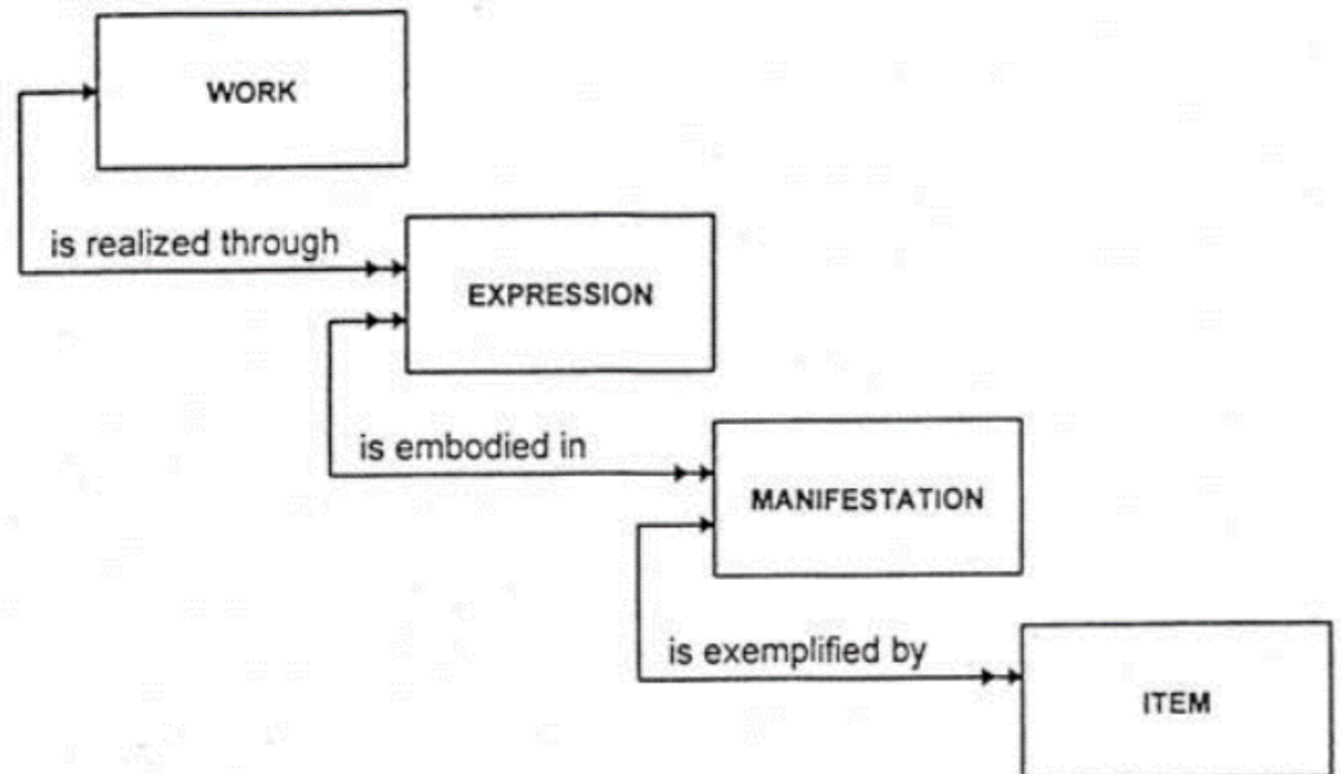
IFLA Study Group on the
Functional Requirements for Bibliographic Records

Approved by the
Standing Committee of the
IFLA Section on Cataloguing

September 1997

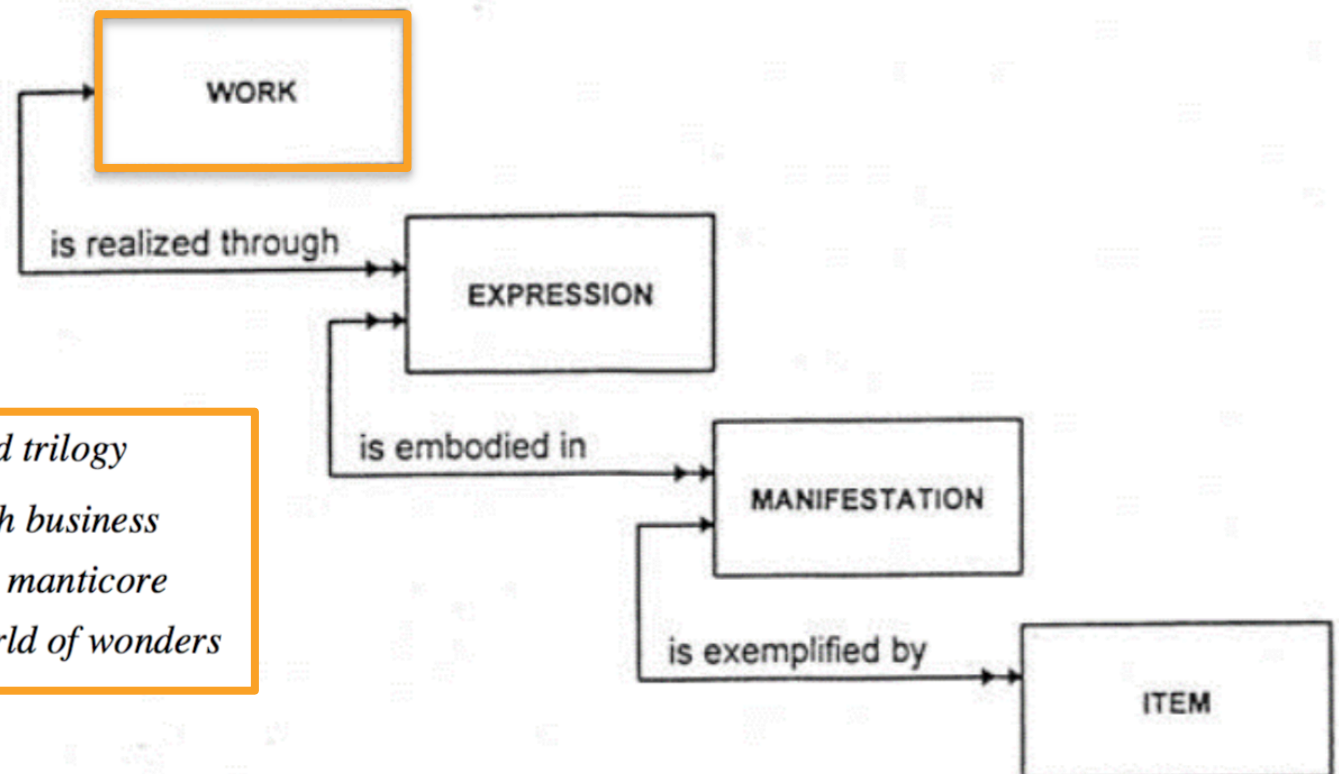
+ Data models for digital editing

- Four basic elements



+ Data models for digital editing

- Four basic elements
- Relationships between these elements

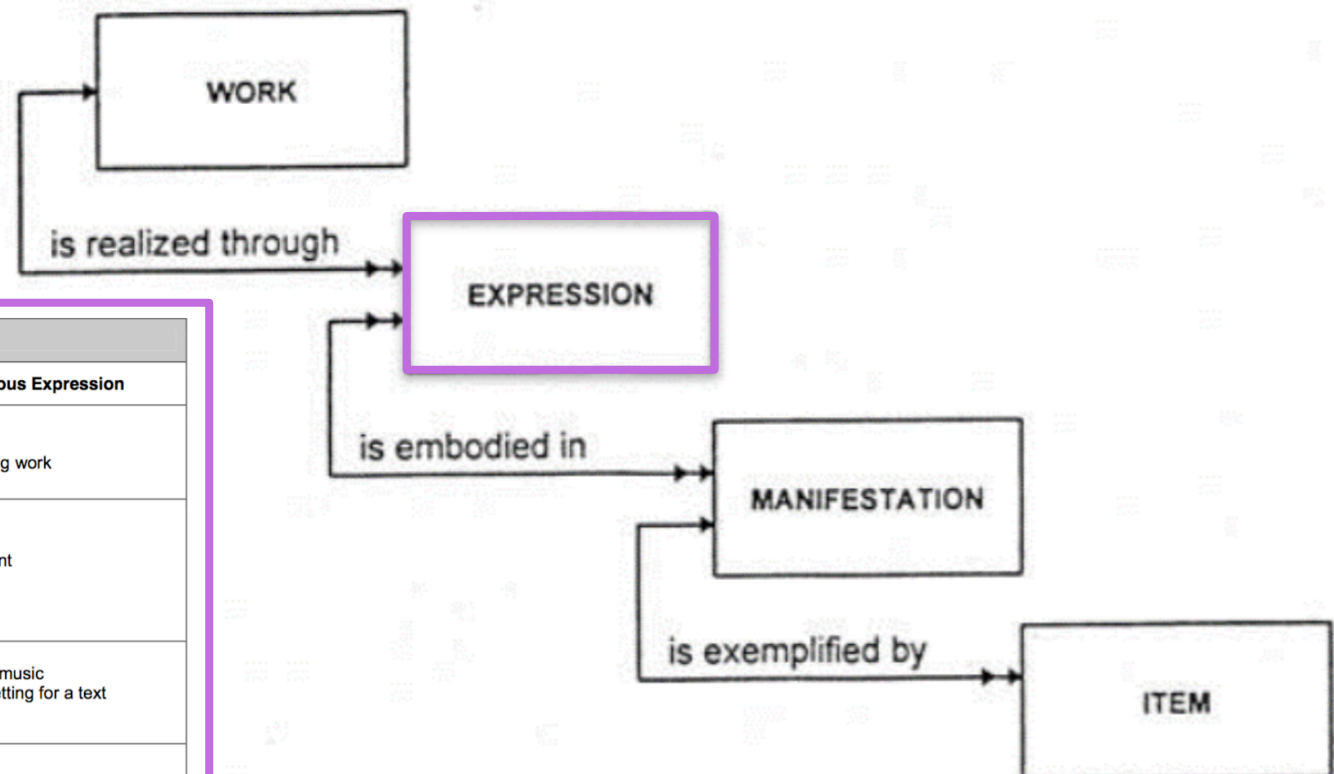


- **w₁** Robertson Davies' *The Deptford trilogy*
 - **w_{1.1}** Robertson Davies' *Fifth business*
 - **w_{1.2}** Robertson Davies' *The manticore*
 - **w_{1.3}** Robertson Davies' *World of wonders*

+ Data models for digital editing

- Four basic elements
- Relationships between these elements

<i>Between expressions of different works</i>		
Relationship Type	Referential Expression	Autonomous Expression
Successor has a successor → ← is a successor to	Sequel	Sequel Succeeding work
Supplement has a supplement → ← supplements	Index Concordance Teacher's guide Gloss Supplement Appendix	Supplement Appendix
Complement has a complement → ← complements	Cadenza Libretto Choreography Ending for unfinished work	Incidental music Musical setting for a text Pendant
Summarization has a summary → ← is a summary of		Digest Abstract



+ Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework
- Data models
 - The FRBR data model (1997 [2009])
‘Neo-platonistic idealism’
Cf. Rafferty (2015) “the notion of *work* points to intertextuality, with all its potential for rich analysis, but at the same time it embeds deep in its system the logocentrism of the ideal signified”

FUNCTIONAL REQUIREMENTS FOR BIBLIOGRAPHIC RECORDS

Final Report

IFLA Study Group on the
Functional Requirements for Bibliographic Records

Approved by the
Standing Committee of the
IFLA Section on Cataloguing

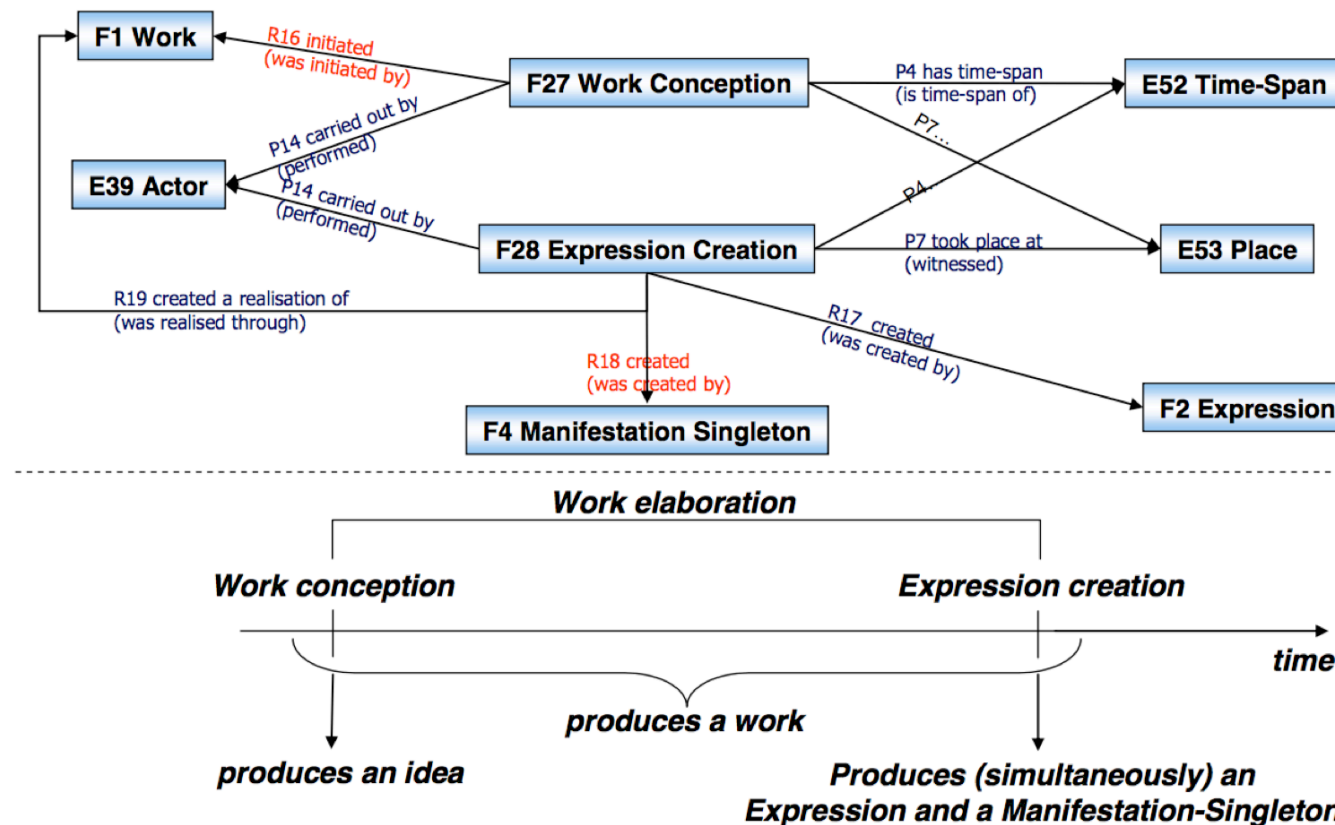
September 1997

+ Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework
- Data models
 - The FRBR data model (1997 [2009])
 - FRBRoo (> CIDOC-CRM)

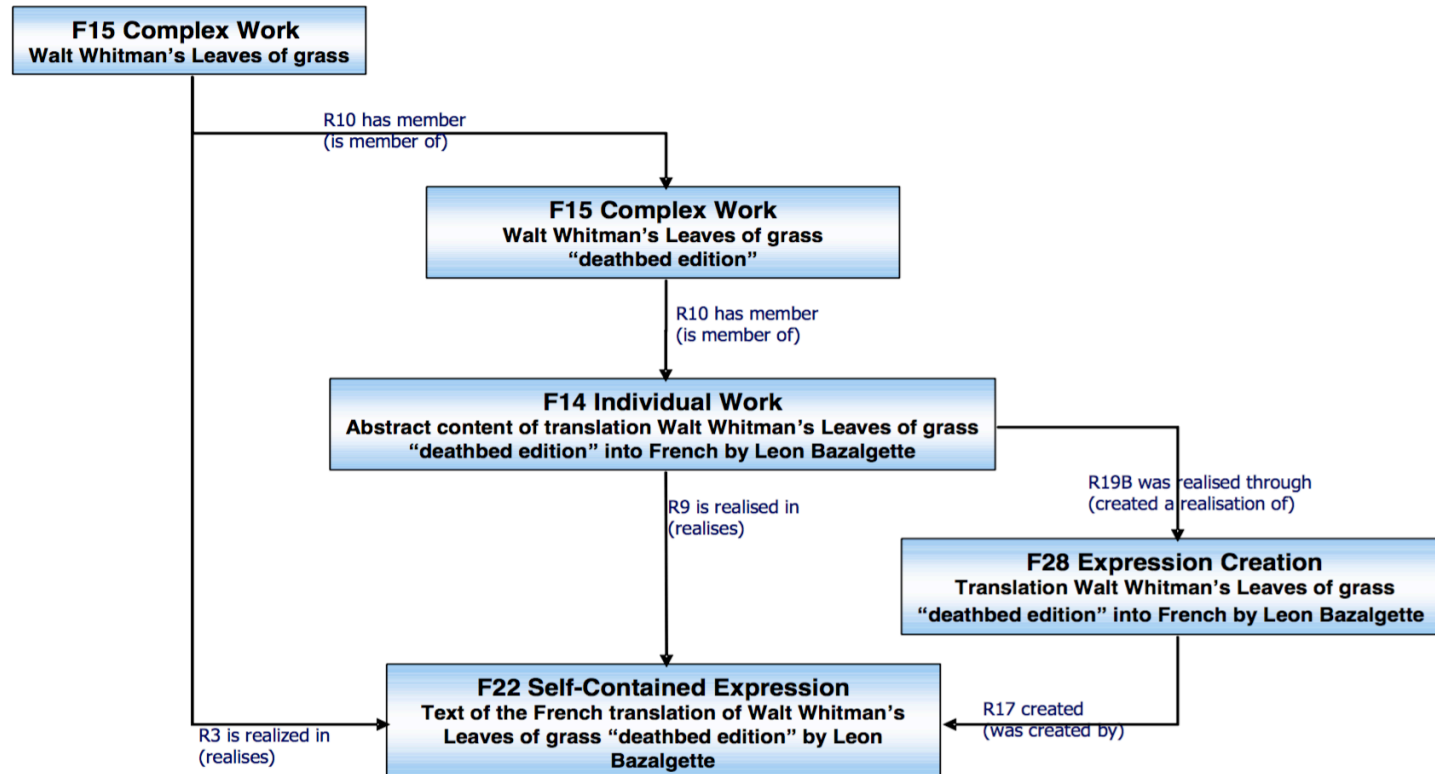
+ Data models for digital editing

- FRBRoo is the FRBR ontology revised and expressed in an object-oriented form compatible with that of the CIDOC-CRM
- Strong orientation towards modelling intellectual processes



+ Data models for digital editing

- FRBRoo is the FRBR ontology revised and expressed in an object-oriented form compatible with that of the CIDOC-CRM
 - Strong orientation towards modelling intellectual processes
 - Addition of some interesting elements



+ Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework
- Data models
 - The FRBR data model (1997 [2009])
 - FRBRoo (> CIDOC-CRM)
 - Pierazzo (2015), *Digital Scholarly Editing*

+ Data models for digital editing

- Pierazzo (2015)
 - Interpretative process
 - Multiple dimensions
 - A single occurrence of ‘intertextuality’ in the book, about the documents’ “literary dimension: style, rhetorical features, genre, intertextuality, citations and allusions.”

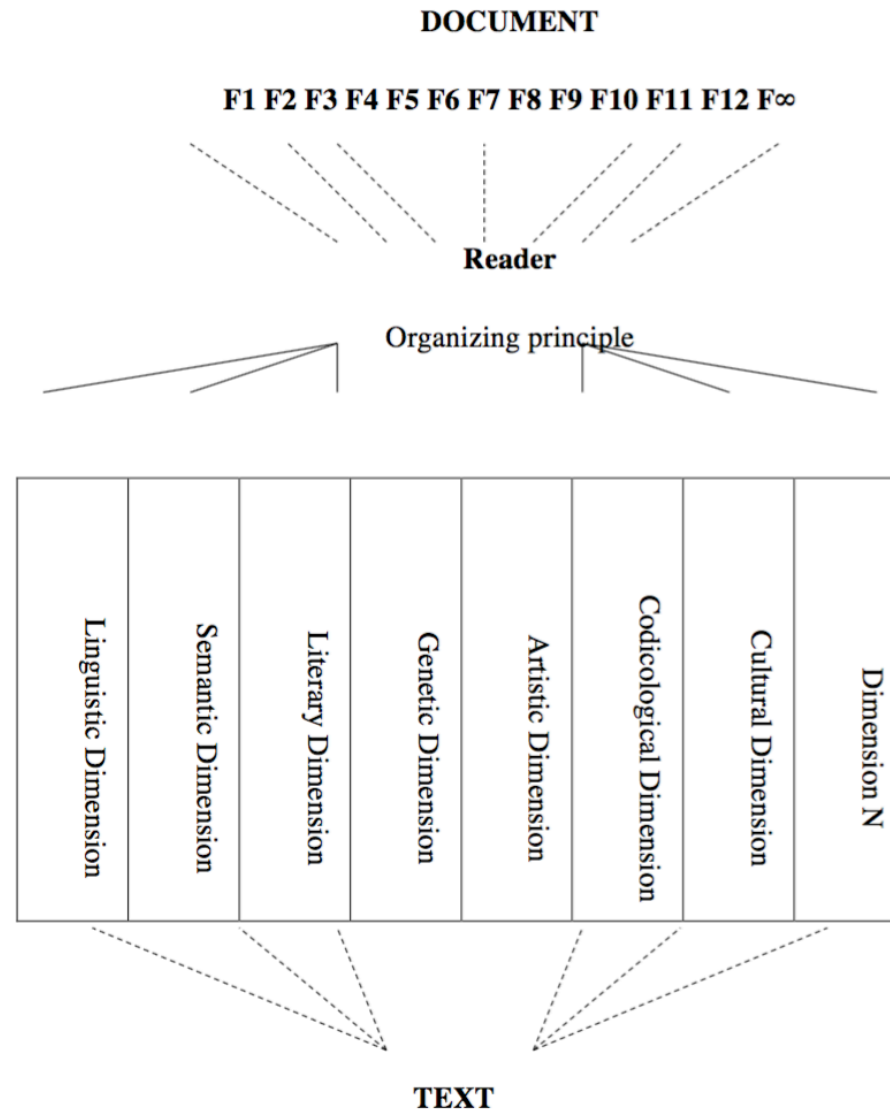
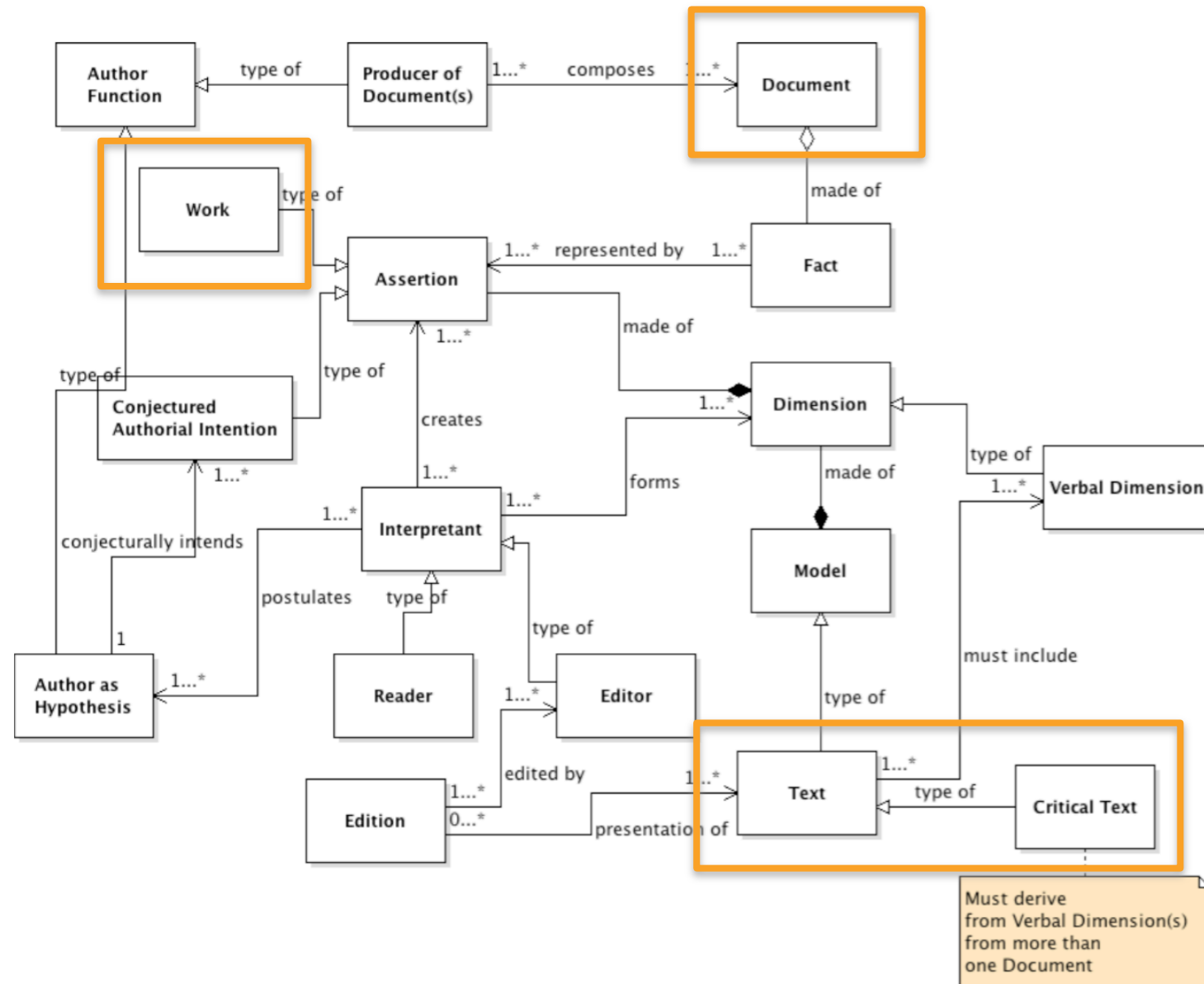
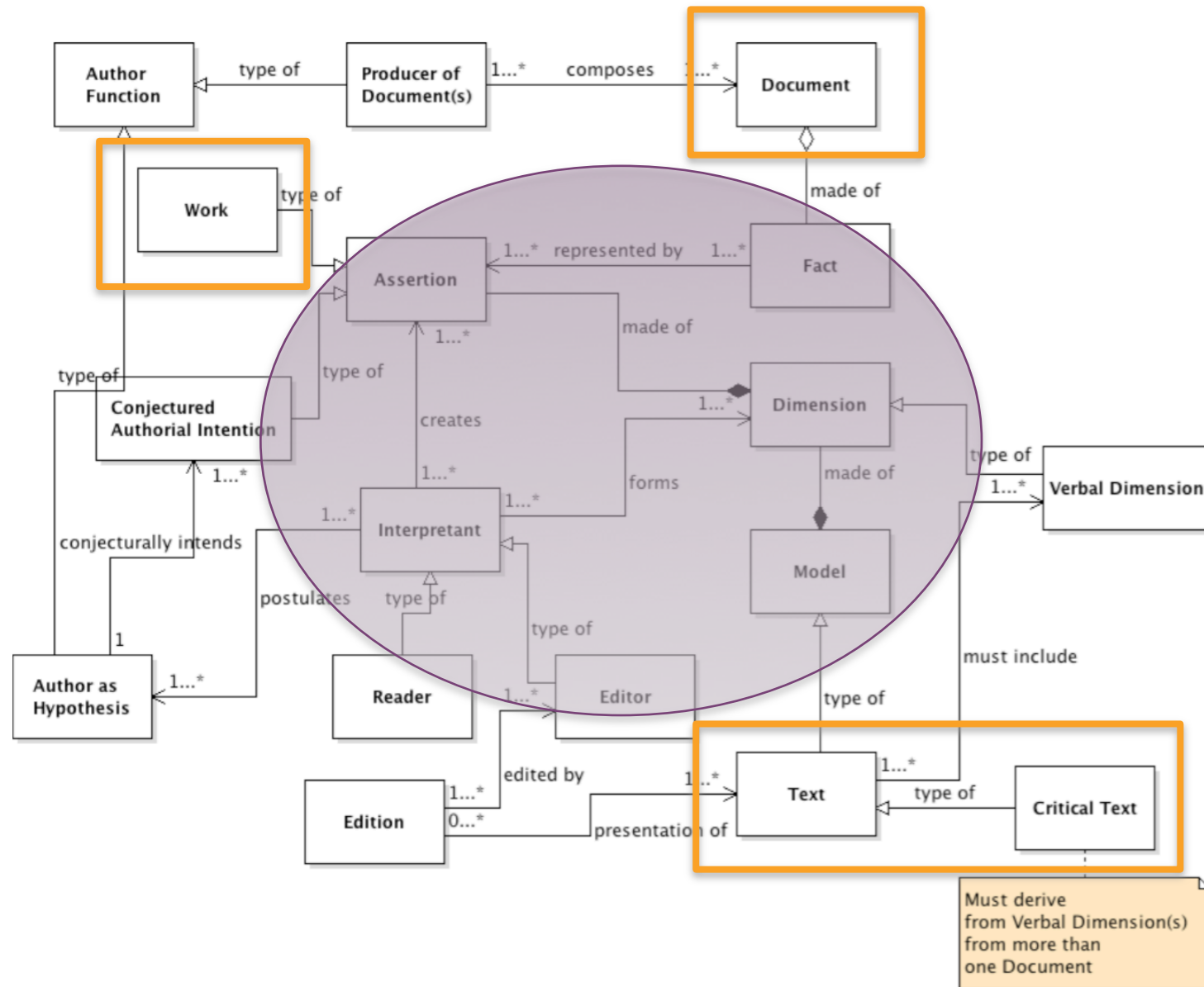


Figure 2.1: Conceptual model of texts and documents

+ Data models for digital editing



+ Data models for digital editing

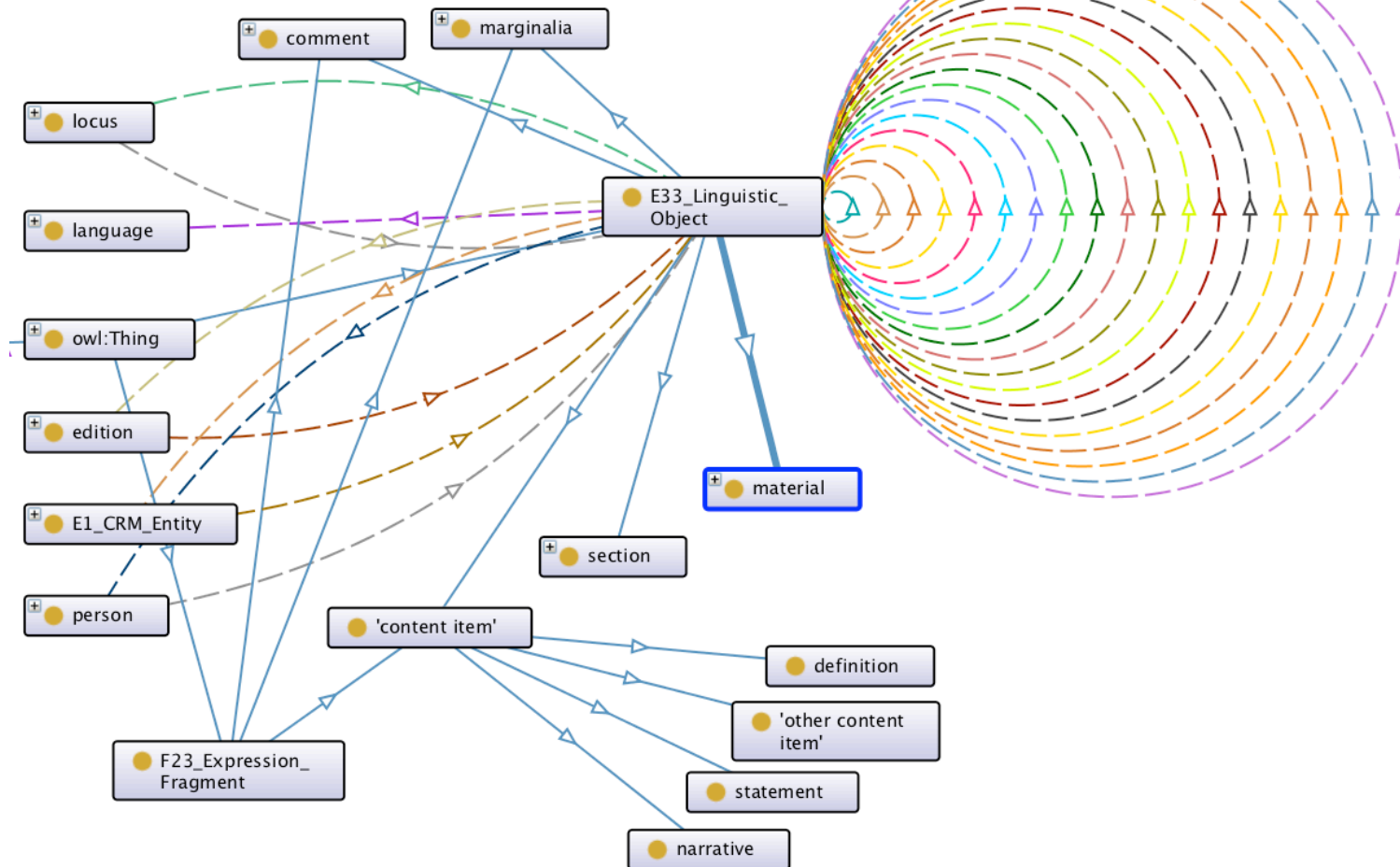


+ Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)
- Theoretical framework
- Data models
 - The FRBR data model (1997 [2009])
 - FRBRoo (> CIDOC-CRM)
 - Pierazzo (2015), *Digital Scholarly Editing*
 - Hedges et al. (2016, in press) – S(haring) A(ncient) W(isdom)S project

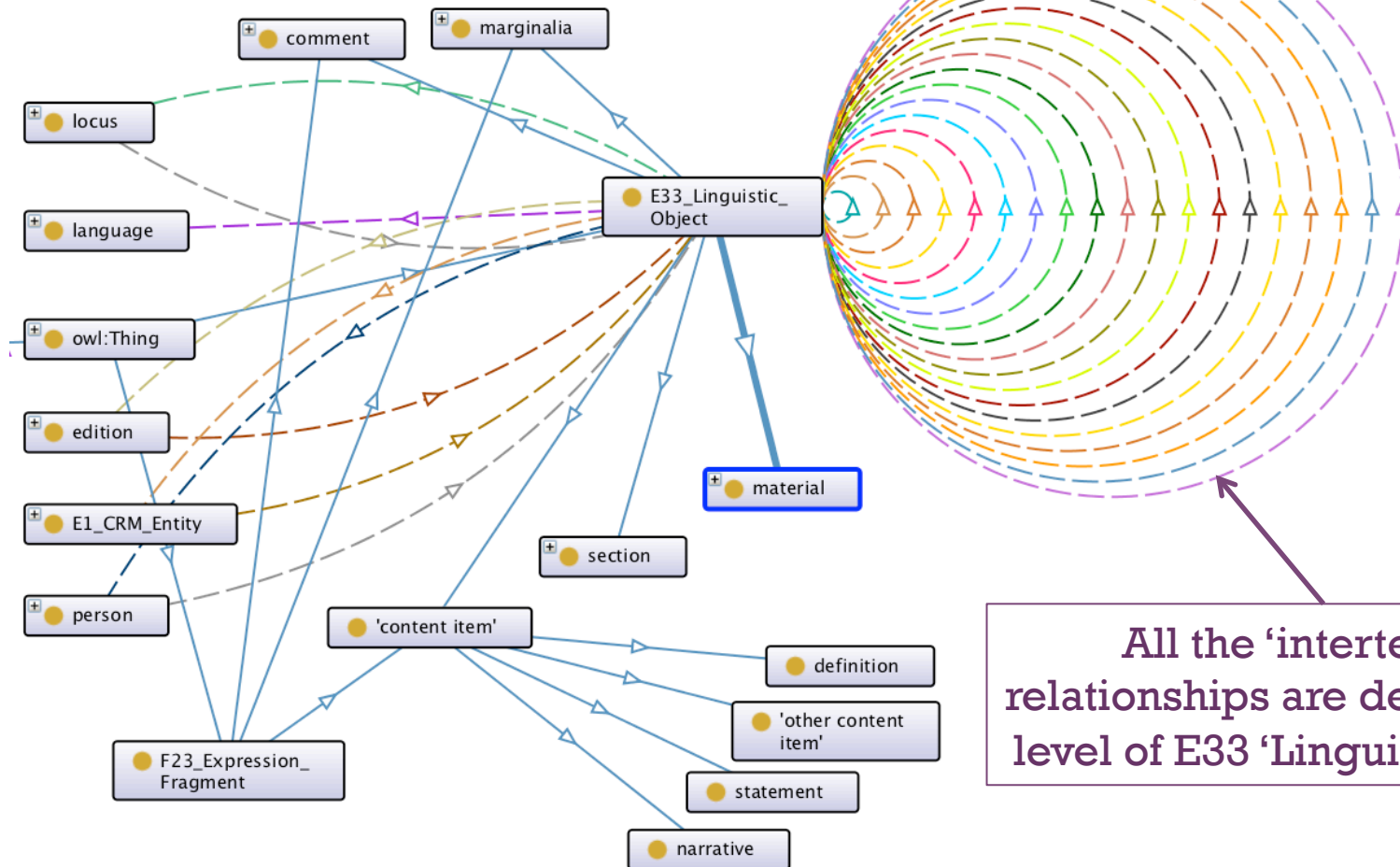
+ Data models for digital editing

- Hedges et al. (2016, in press) – SAWS
 - They worked with an extension of FRBRoo



+ Data models for digital editing

- Hedges et al. (2016, in press) – SAWS
 - They worked with an extension of FRBRoo



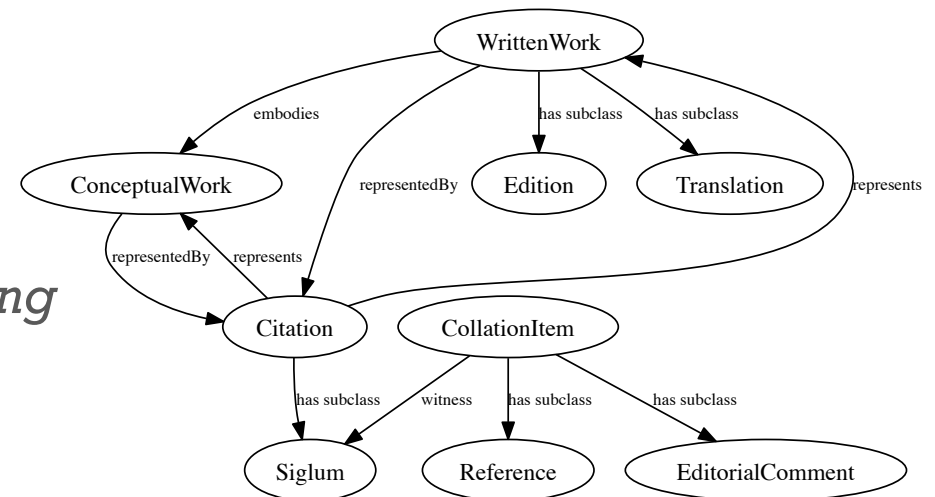
+ Data models for digital editing

- ‘Digital approaches to intertextuality’ is a hot topic
- Many tools (and a huge body of literature) for automatic detection of INTERTEXTS and TEXT REUSE (TAL)

- Theoretical framework

- Data models

- The FRBR data model (1997 [2009])
- FRBRoo (> CIDOC-CRM)
- Pierazzo (2015), *Digital Scholarly Editing*
- Hedges et al. (2016, in press)
- [L(inking) A(ncient) W(orld) D(ata) <http://lawd.info>]



+ Data models for digital editing

- To sum up
 - The concept *work* as an abstract/unifying element represented by witnesses/expressions

+ Data models for digital editing

- To sum up
 - The concept *work* as an abstract/unifying element represented by witnesses/expressions
 - ‘Part-whole’ relationships
 - Work-level: ‘Complex’ works made of ‘simpler’ works
 - Witness-level: ‘Complex’ witnesses made of smaller witnesses

+ Data models for digital editing

- To sum up
 - The concept *work* as an abstract/unifying element represented by witnesses/expressions
 - ‘Part-whole’ relationships
 - Work-level: ‘Complex’ works made of ‘simpler’ works
 - Witness-level: ‘Complex’ witnesses made of smaller witnesses
 - Other relationships
 - Between different works (‘sequel’, ‘imitation’, etc.)
 - Between different witnesses (‘verbatim copy of’, ‘translation’, etc.)



Thot Data Model (TDM 2.0)

+ Thot Data Model (2.0)

- Thomas Tanselle (1989), *A Rationale of Textual Criticism*
 - **Texts of documents**, namely the texts one can derive from physical documents
 - **Texts of works**, “namely the ideal texts that the author had intended to write but which have never been realized in practice” (see Pierazzo 2015)

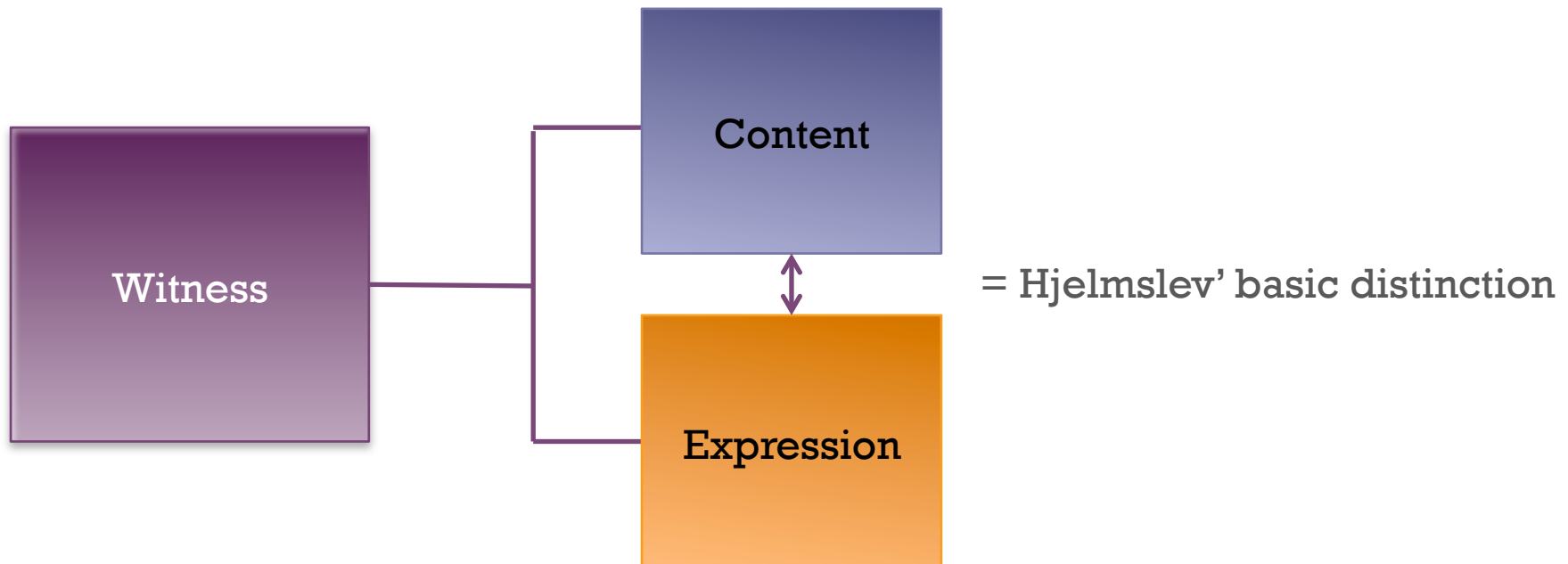
+ Thot Data Model (2.0)

- Thomas Tanselle (1989), *A Rationale of Textual Criticism*
 - **Texts of documents**, namely the texts one can derive from physical documents
 - **Texts of works**, “namely the ideal texts that the author had intended to write but which have never been realized in practice” (see Pierazzo 2015)

Witness

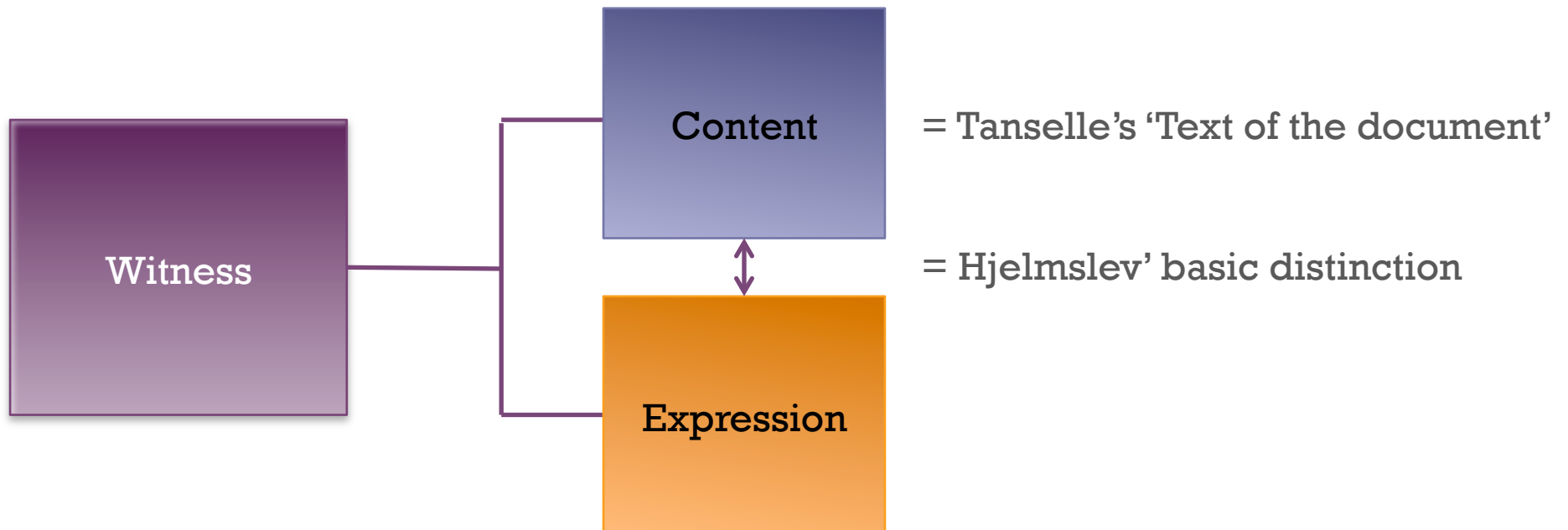
+ Thot Data Model (2.0)

- Thomas Tanselle (1989), *A Rationale of Textual Criticism*
 - **Texts of documents**, namely the texts one can derive from physical documents
 - **Texts of works**, “namely the ideal texts that the author had intended to write but which have never been realized in practice” (see Pierazzo 2015)



+ Thot Data Model (2.0)

- Thomas Tanselle (1989), *A Rationale of Textual Criticism*
 - **Texts of documents**, namely the texts one can derive from physical documents
 - **Texts of works**, “namely the ideal texts that the author had intended to write but which have never been realized in practice” (see Pierazzo 2015)

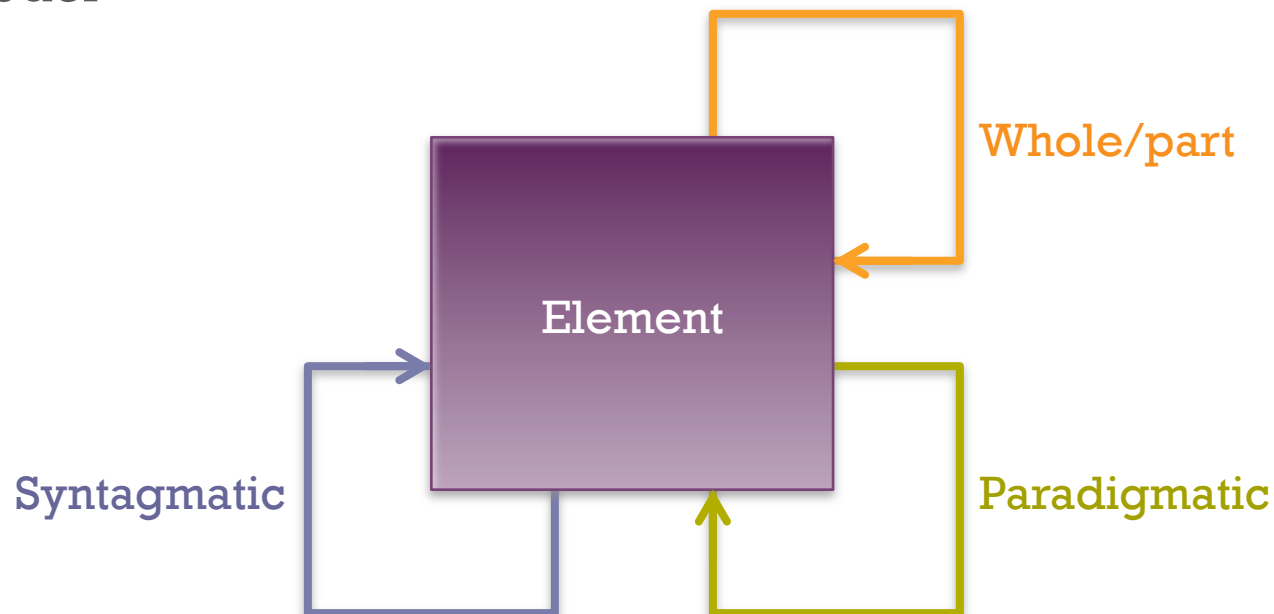


+ That Data Model (2.0)

- The Domain and Range of intertextual relationships become much clearer when this distinction is made (cf. Böhler's opposition between 'syntactic and semantic text reuses')

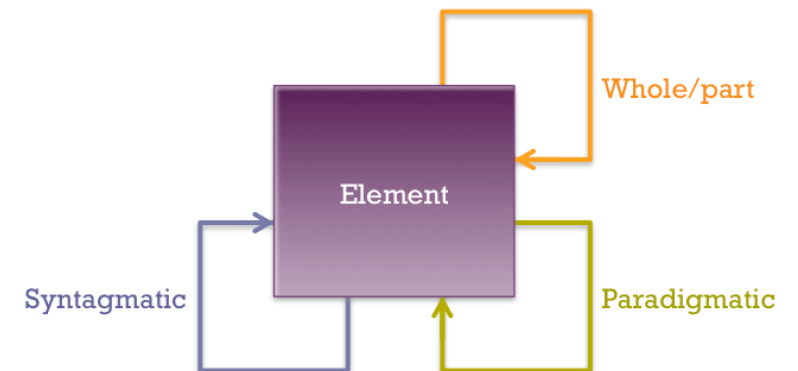
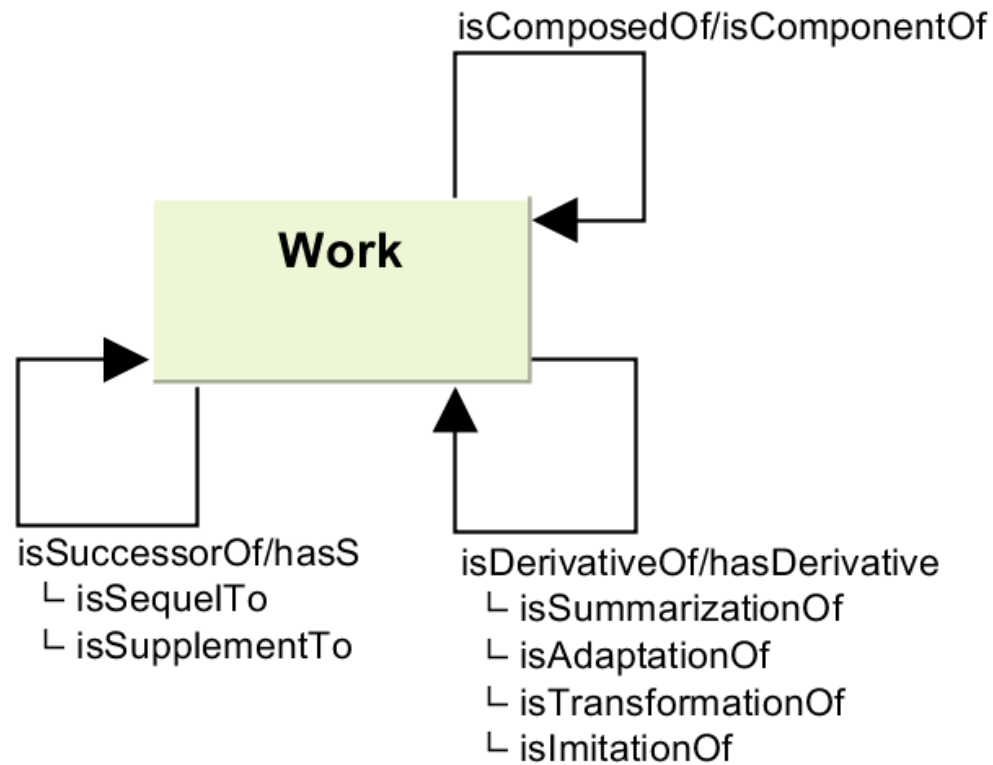
+ Thot Data Model (2.0)

- The Domain and Range of intertextual relationships become much clearer when this distinction is made (cf. Büchler's opposition between 'syntactic and semantic text reuses')
- An analysis of the relationships suggested in the literature suggests that they can be organized according to three types for any element of the model



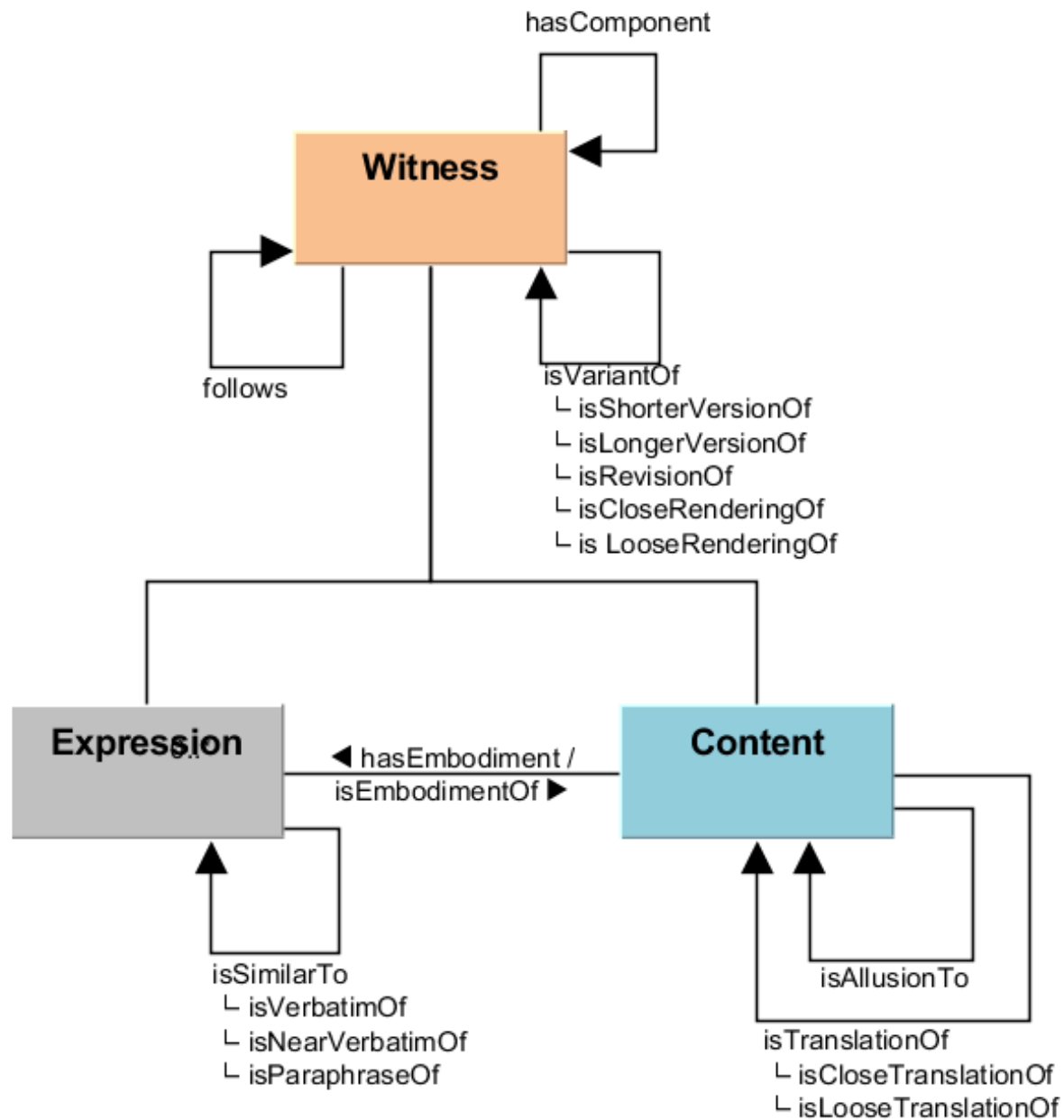
+ Thot Data Model (2.0)

■ The *Work*



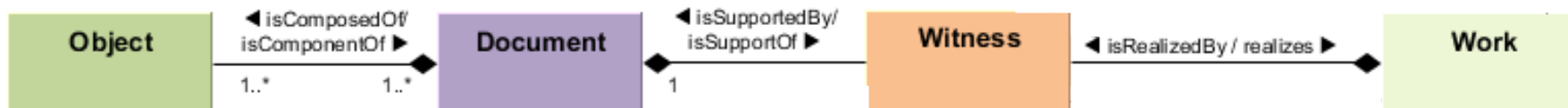
+ Thot Data Model (2.0)

- The *Work*
- The *Witness*
 - Expression
 - Content



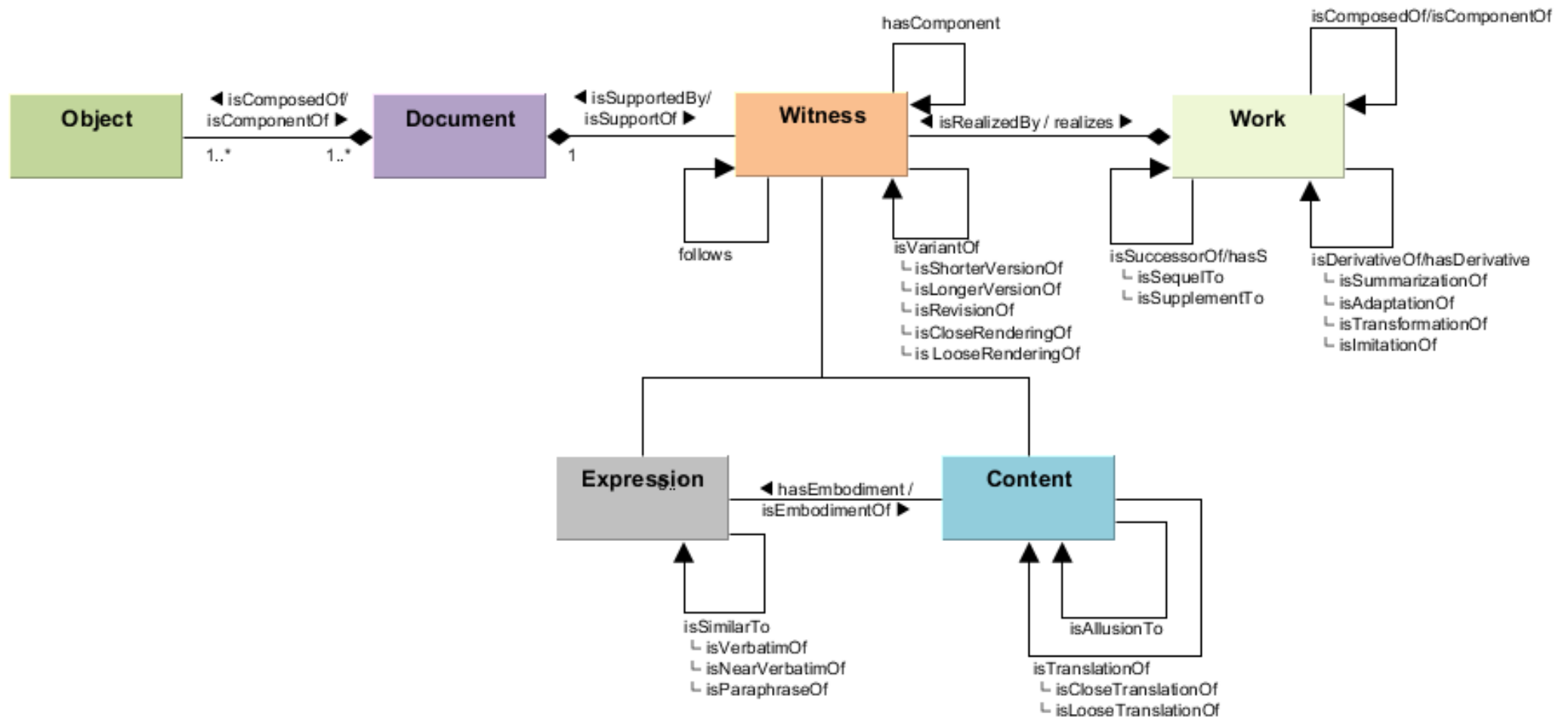
+ Thot Data Model (2.0)

- General model



+ Thot Data Model (2.0)

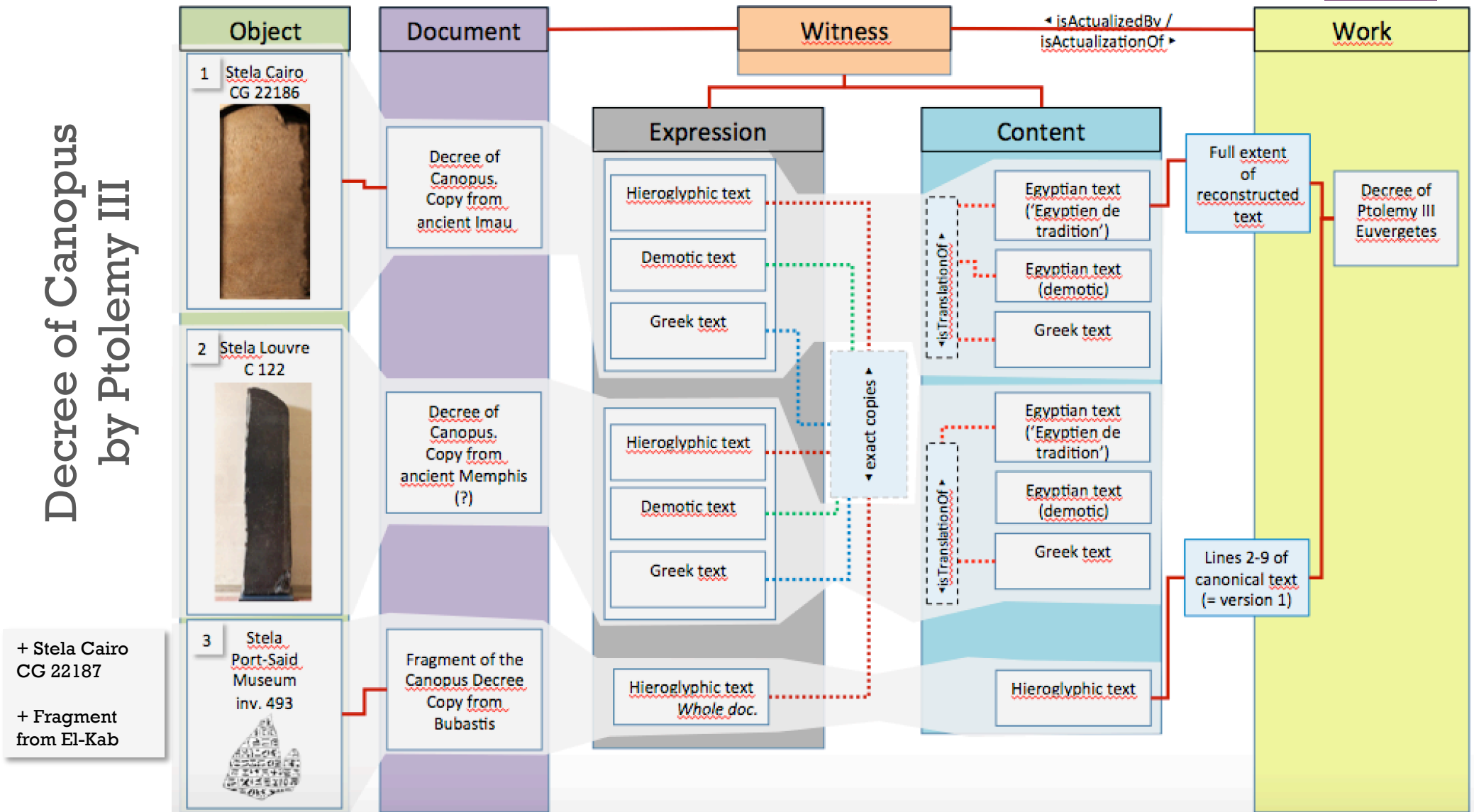
■ General model



+ Thot Data Model (2.0)

Illustration 1. Multilingual decree

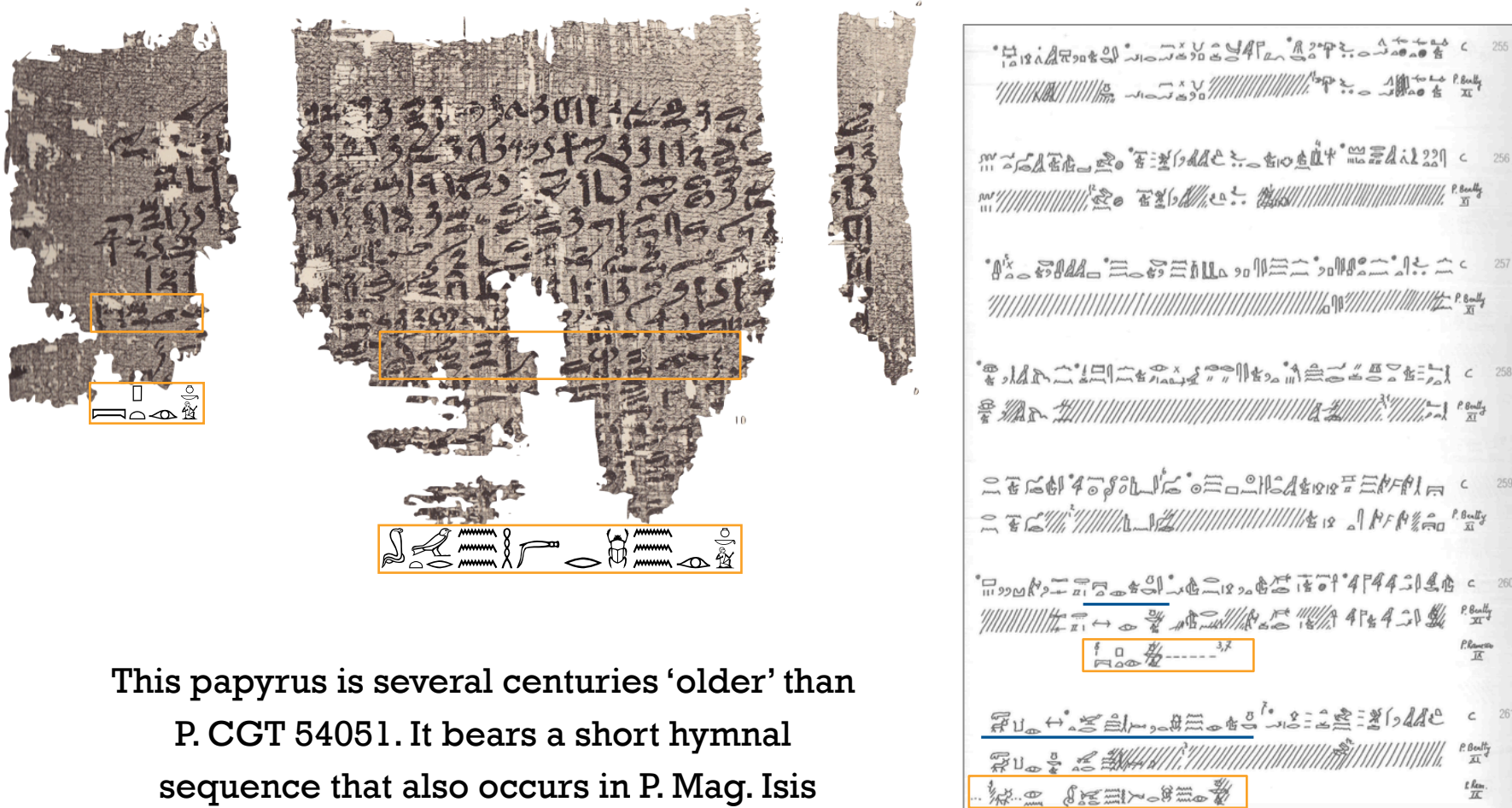
Decree of Canopus by Ptolemy III



+ Thot Data Model (2.0)

Illustration 2. Distant text reuse

P. Ramesseum IX



This papyrus is several centuries 'older' than P. CGT 54051. It bears a short hymnal sequence that also occurs in P. Mag. Isis

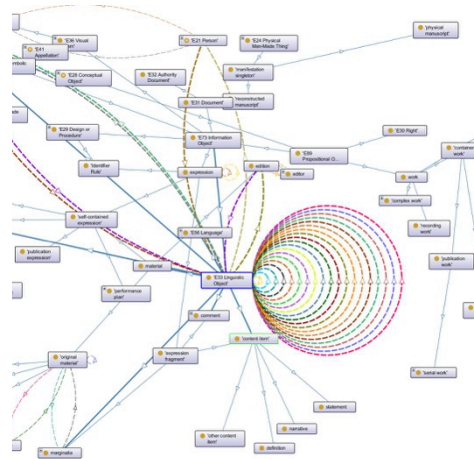
+ Thot Data Model (2.0)

Implementation





Thanks!



TEI <Text Encod

Home Guidelines Activities Tools Member

[Home](#)

TEI-C News

- TEI 2014 Members Meeting website online
Posted on: 2014-05-01
- Final Call for Papers - 2014 TEI Conference
Posted on: 2014-04-18
- Social Media Coordinator for the TEI: position available
Posted on: 2014-03-20
- TEI Conference 2014: First Call for Papers
Posted on: 2014-02-04
- Issue 6 of the Journal of the TEI published: Selected Papers from the 2012 TEI Conference
Posted on: 2014-01-27
- TEI P5 Guidelines version 2.6.0 is released!
Posted on: 2014-01-23

Other News

- Digital Scholarly Editions Initial Training Network (DixIT)
Posted on: 2013-07-19
- EpiDoc Workshop, London, April 22-25, 2013
Posted on: 2013-01-14
- CFP: Issue 6 of the Journal of the Text Encoding Initiative
Posted on: 2012-10-31
- Call for TAPAS test data
Posted on: 2012-09-26
- Seminar, Taking TEI Further: Transforming and Publishing TEI Data
Posted on: 2012-08-15

[Older items](#)

TEI:
The Te digital f social s to pres of reso software
The TE around
Want to come to