

ANCIENT EGYPTIAN TEXTS IN CONTEXT. TOWARDS A CONCEPTUAL DATA MODEL (THE THOT DATA MODEL — TDM)*

STÉPHANE POLIS AND VINCENT RAZANAJAO

In Ancient Egypt, writing is everywhere: from religious and funerary buildings to everyday objects, most architectural and many artefactual productions have been inscribed. This pervasiveness of writing lasted for more than three millennia, starting with the very first manifestations (end of the fourth millennium BC), and continuing until the last manifestations during the Ptolemaic and Roman periods (fourth century BC–fourth century AD).

Despite the inextricable link between the texts and their supporting contexts¹ (whether monuments, smaller artefacts, or even the physical landscape), current Egyptological text-oriented projects² often fail to handle both dimensions together satisfactorily, and generate a significant amount of data about texts that are (at least partly) divorced from their contexts.³ For this reason, ancient Egyptian textuality is difficult to apprehend within

* We are in debt to many members of the Egyptological community for fruitful discussions on the topic: Francisco Bosch-Puche (Oxford), Laurent Coulon (Paris), Peter Dils (Leipzig), Todd Gillen (Liège), Silke Grallert (Berlin), Eitan Grossman (Jerusalem), Ingelore Hafemann (Berlin), Lutz Popko (Leipzig), Sebastian Tonio Richter (Berlin), Serge Rosmorduc (Paris), Simon Schweitzer (Berlin), Nathalie Sojic (Liège), Daniel Werning (Berlin), and Jean Winand (Liège). We are grateful to the editors and reviewers for their comments and suggestions on earlier drafts of this paper, and we duly take ownership of all remaining mistakes and shortcomings.

¹ See e.g. C. Haas, *Writing technology: Studies on the materiality of literacy* (Mahwah 1996) for a discussion of the inextricable link between writing and the material world, as well as the conceptualization of writing as an embodied practice.

² See especially the *Thesaurus Linguae Aegyptiae* (<http://aaew.bbaw.de/tla/>) and *Ramses Online* (<http://ramses.ulg.ac.be>). For a detailed presentation of these corpora, see respectively P. Dils and Fr. Feder, 'The *Thesaurus Linguae Aegyptiae*. Review and perspectives' and St. Polis, A.-Cl. Honnay, and J. Winand, 'Building an annotated corpus of Late Egyptian. The Ramses Project: Review and perspectives', in *Texts, Languages & Information Technology in Egyptology*, ed. St. Polis & J. Winand (Liège 2013) 11–23 and 25–44.

³ Conversely, it is quite obviously in the interests of archaeological projects to integrate textual data (see e.g. B. J. Little, *Text-aided archaeology* (Boca Raton 1991); for written documents as constitutive of the past rather than evidence about the past, see J. Moreland, *Archaeology and text* (London 2001); J. Moreland, 'Archaeology and Texts: Subsistence or enlightenment', *Annual Review of Anthropology* 35 (2006) 135–151), since it helps to understand the space being excavated, for instance through descriptions, or mentions of names, of buildings or of institutions. An example of poor incorporation of the textual material is the 1980s project on Naukratis, which denied the existence of a great Egyptian *temenos* or sacred area, despite the discovery by Petrie (a century earlier) of foundation deposits bearing the name of Ptolemy I in hieroglyphs (A. J. Spencer, 'The Egyptian temple and settlement at Naukratis', *British Museum Studies in Ancient Egypt and Sudan* 17 (2011) 35).

its full (present) archaeological and (past) cultural environment,⁴ especially when using digital resources.⁵

The aim of this contribution is to present a conceptual data model — the Thot Data Model (TDM) — that allows an integrated digital approach to ancient Egyptian writings, by capturing within a single framework the whole range of (meta)data relating to written performance.⁶ Furthermore, since explicit definitions of the elements and of their relation types are provided, this model could be used as a *generic* data model for other philological disciplines dealing with textual material of the ancient world and beyond. As such, we situate this paper as a further step within the creation ‘of an encoding model that could combine the textual as well as the material dimension of an archeological object-bearing text’.⁷

Our contribution is based on research that was first developed within two independent projects,⁸ hosted respectively at the Griffith Institute (University of Oxford) and the

⁴ This state of affairs is intimately linked to the strong distinction between text-oriented, art-oriented, and archaeology-oriented approaches in Egyptology (besides the usual disciplinary oppositions, see St. Quirke *Exploring religion in Ancient Egypt* (Chichester, 2015) 9 for underlying epistemological reasons), and to the history of the field, strongly emphasizing the textual aspect over other dimensions. In this sense, two illuminating examples can be mentioned: (1) The *Pyramid texts* (the oldest religious corpus in the history of Egypt) were, until recently, mostly studied based on a synoptic edition (K. Sethe, *Die altägyptischen Pyramidentexte* (Leipzig, 1908–1922)). Reconsidered in their physical context, many formulae can now be better understood using the clues given by the actual position in a room, on a wall, or at the bottom of a corridor entrance (e.g. J. Allen, ‘Reading a pyramid’, in *Hommages à Jean Leclant*, Bibliothèque d’étude 106, ed. C. Berger, G. Clerc and N. Grimal, vol. I (Cairo 1994) 5–28; H. M. Hays, *The organization of the Pyramid Texts: Typology and disposition*, *Probleme der Ägyptologie* 31 (Leiden/Boston 2012) 79, 83; B. Mathieu, ‘Linguistique et archéologie: l’usage du déictique de proximité (pn/tm/nn) dans les Textes des Pyramides’, in *Aere perennius. Mélanges égyptologiques en l’honneur de Pascal Vernus*, *Orientalia Lovaniensia Periodica* 242, ed. Ph. Collombert, D. Lefèvre, St. Polis, and J. Winand (Leuven 2016) 407–427). (2) It is significant that the architectural study of the temple of Hathor in Dendara (P. Zignani, *Le temple d’Hathor à Dendara: Relevés et étude architecturale*, Bibliothèque d’étude 146, 2 vols (Cairo 2010)) was published nearly eighty years after the first volume of inscriptions was released (Ém. Chassinat, *Le temple de Dendara* (Cairo, 1934–1935)), all the more given that the significance of spatial arrangement of texts in temples from the Greco-Roman period has been known since the 1960s (Ph. Derchain, ‘Un manuel de géographie liturgique à Edfou’, *Chronique d’Égypte* 37/73 (1962) 31–65).

⁵ In Egyptology, a field of study that explicitly envisions writing as a *material practice* has recently emerged. It provides empirical and methodological observations that are directly relevant for our purposes. For the *material* aspect, see e.g. K. E. Piquette, ‘Re-materializing script and image’, in *Current Research in Egyptology 2008*, ed. V. Gashe and J. Finch (Bolton 2008) 89–107, as well as several contributions in *Writing as Material Practice. Substance, surface and medium*, ed. K. E. Piquette and R. D. Whitehouse (London, 2013). Regarding writing as a *practice*, see the ongoing work of Chl. Ragazzoli on the New Kingdom scribal elite (e.g. ‘The social creation of a scribal place: The visitors’ inscriptions in the tomb attributed to Antefiqer (TT60) (with newly recorded graffiti), *Studien zur Altägyptischen Kultur* 42 (2013) 269–323).

⁶ In the long run, this model could also benefit large-scale (non-textual) databases such as Trismegistos (<http://www.trismegistos.org>), which does not distinguish systematically between the document (identified by a TM number), the physical object (e.g. a piece of papyrus), and the text (e.g. a given Demotic letter); this represents a major difference with the TDM argued for in this paper. For the guidelines followed in problematic cases, see http://www.trismegistos.org/about_identifiers.php.

⁷ Emmanuelle Morlock and Eleonora Santin, ‘The inscription between text and object. The deconstruction of a multifaceted notion with a view of a flexible digital representation’, in *First EAGLE International Conference on Information Technologies for Epigraphy and Cultural Heritage, Sep 2014*, ed. S. Orlandi et al. (Paris 2014) 325–350. To the best of our knowledge, this study is the only one that addresses directly the questions discussed here. D. Buzzetti, ‘Digital representation and the text model’, *New Literary History* 33 (2002) 61–88 is chiefly concerned with the merits of strongly vs weakly embedded mark-ups for representing the structure of the form and content of a text. The novelty of this study lies in the fact that it sees the *Text* as a ‘coherent and non-linear representation’ both in terms of *expression* (the different witnesses) and *content* (the related interpretations) (76–81). This point shall be discussed in section 4.1.

⁸ Namely the *TopBib* (<http://topbib.griffith.ox.ac.uk>) and *Ramses* (<http://ramses.ulg.ac.be>) projects.

Department of Egyptology at the University of Liège. While working together on a TEI-compliant interchange format for Ancient Egyptian-Coptic textual resources (see section 4.2), a group of shared concerns on the topic arose from the topographical orientation of the former and the textual dimension of the latter.⁹

The paper is structured as follows. In section 1, we provide a definition of the four main conceptual elements of the data model, namely *Object*, *Document*, *Witness*, and *Text*, and we describe how they relate to one another. Metadata that apply to these elements and relationships are then discussed in section 2. Section 3 is devoted to a brief discussion of the advantages of this data model in terms of localization of written production. The links between the TDM and other conceptual models (as well as TEI elements) are presented in section 4. The conclusions provide perspectives concerning the online resources that would allow a linked open data implementation of the model in Egyptology.

1. The Conceptual Data Model

1.1. The core classes: Object, Document, Witness, and Text¹⁰

In this section, we take as a point of departure the definitions of the four key elements of the TDM and we proceed with a discussion of their relationships.¹¹ These elements are:

- *Object*, which refers to a physically discrete material object (in its present state, based on current available evidence);¹²
- *Document*, which refers to an artefact reconstituted in its original entirety (*i.e.* materially made up of 1 to *n* *Objects*) and envisioned as an idealized writing space or text support;¹³
- *Witness*, which refers to a single occurrence of a *Text*, in its material (and more broadly philological) dimensions, on 1 to *n* *Document*;¹⁴
- *Text*, which refers to a textual composition as it can be reconstructed from (the compilation of 1 to *n*) *Witness(es)*.¹⁵

⁹ A first paper on the topic was presented at the *Text encoding initiative conference and members' meeting 2015* (October 28–31, Lyon – France) by L. Coulon, F. Elwert, E. Morlock, S. Polis, V. Razanajao, S. Rosmorduc, S. Schweitzer, and D. Werning.

¹⁰ The basic elements introduced in this section (as well as their relationships) were first presented in an internal report of the Ramses Project by T. Gillen, St. Polis, and N. Sojic, *Metadata: A progress report for the Ramses Project* (2014/05/14).

¹¹ These four elements are identified (although with other labels) by Morlock and Santin, 'The inscription between text and object' (n. 7, above) 141 (see n. 12–15, below). The main originality of the TDM lies, therefore, in the explicit definition of all the possible types of relationships between these elements.

¹² Compare with the 'physical object part' defined as 'a detachable physical part of a material object that can be physically isolated' in Morlock and Santin, 'The inscription between text and object' (n. 7, above) 341.

¹³ Compare with 'text-bearing object' defined as 'a material object that bears one or several inscribed texts' in Morlock and Santin, 'The inscription between text and object' (n. 7, above) 341. Note that *Document* is explicitly defined in the TDM in relation to a writing space. No matter how interesting the history of a *Document* before it was envisioned as a text-bearing object, the TDM does not record it. However, the various features that can account for its selection, fabrication, reshaping, *etc.* as a text-bearing object are of course recorded in the metadata pertaining to the *Document* and its *Writing surface*.

¹⁴ Compare with the 'inscribed entity' defined as 'the set of marks that were inscribed on a material support' in Morlock and Santin, 'The inscription between text and object' (n. 7, above) 341.

¹⁵ Compare with the 'abstract text' defined as 'an abstract entity corresponding to the 'object of thought' that is the

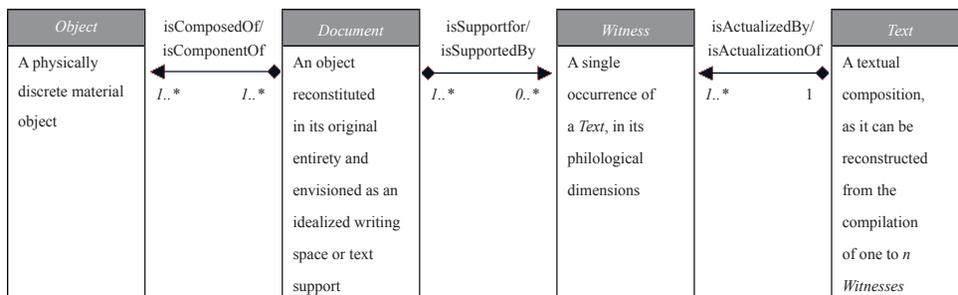


Figure 1: The four basic elements and their relationships.

Among these four elements, two refer to dimensions which can be considered concrete in the present day, that is, *Object* (as a discrete material object) and *Witness* (as the actual occurrence of a text), while two refer to abstract dimensions, namely *Document* (as an object reconstituted in its original integrity)¹⁶ and *Text* (as the reconstruction of a textual composition). It should be stressed that we adopt a modern perspective here, but that the elements *Document* and *Text* were obviously ‘concrete’ entities at some point in the cultural environment of ancient Egypt: the metadata that describe these elements (section 2) should suffice to demonstrate that they fully belong to the scribal practices of the past (*i.e.* were produced and inscribed by individuals in a specific temporal and geographical context with an intended audience and specific communicative goals).

The relationships between the core elements of the TDM allow us to handle the empirical cases that have been encountered so far. Beside the self-explanatory one-to-many relationships between the elements *Text* and *Witness* (a *Text* can be actualized by several *Witnesses*), the many-to-many relationships between the elements *Object* and *Document*, on the one hand, and the elements *Document* and *Witness*, on the other, require further discussion.

The many-to-many relationship between *Object* and *Document* allows us to cope with two situations, namely: (1) when one *Document* is nowadays broken and reconstructed based on several *Objects* (one *Document* – many *Objects*; see the example discussed below in section 1.3); (2) when one *Object* has been used for several *Documents* (one *Object* – many *Documents*). The latter occurs in cases of reuse of an *Object*, meant to function as a new text-bearing *Document*. Consider for instance the so-called Second Stela of Kamose¹⁷ (*c.* 1550 BC), which was originally a doorjamb within a monument of Senusret I (*c.* 1950 BC).¹⁸ The *Object* as we know it today was first used as part of a larger *Document* (*i.e.* a building not identified with any certainty) and inscribed with the royal titulary (as well as representations of the king and deities). In a second step, the building was dismantled and the same *Object* was reshaped as a new text-bearing *Document* (a *stela*) recording Kamose’s victory against the Hyksos. A single artefact (made of limestone) was therefore used for two

denotata of the inscribed entity or its intellectual content’ in Morlock and Santin, ‘The Ininscription between text and object’ (n. 7, above) 341.

¹⁶ See D. Buzzetti, ‘Digital representation and the text model’, 76–81 and the discussion above in n. 7.

¹⁷ L. Habachi, *The second stela of Kamose and his struggle against the Hyksos ruler and his capital*, *Abhandlungen des Deutschen Archäologischen Instituts Kairo, Ägyptologische Reihe 8* (Glückstadt 1972).

¹⁸ See S. Biston-Moulin, ‘De Sésostris Ier à Kamosis: note sur un remploi de Karnak’, *Égypte Nilotique et Méditerranéenne 4* (2011) 81–90 [<http://www.enim-egyptologie.fr/index.php?page=enim-4&n=4>].

Documents (a doorjamb and a *stela*), each receiving specific *Witnesses*.¹⁹

The many-to-many relationship between *Document* and *Witness* applies in two cases: (1) when one *Document* bears several *Witnesses* (see the example discussed below in section 1.3); (2) when one *Witness* spans over several *Documents*. The latter can be illustrated by the following example: the *Witness* of a single magical text spans over two scrolls (P. Louvre E 3237 & E 3239), respectively numbered by the scribe ‘first’ and ‘second’.²⁰

1.2. Complementary classes and relationship attributes

In order to document how texts and text bearers interact with each other, our data model implements a class *Writing Surface*, an interface between the document and the text written on it. *Writing Surface* is a concept attached to *Document* and not to *Object*, since a writing surface has to be defined regardless of the current materiality of a text bearer (be it made of several objects or in a poor state of preservation).

The concept of the *Writing Surface* is not sufficient to link a document to a text, and an extra layer of information has to be provided: the organization or layout of the text on the *Writing Surface*. *Layouts* are sections of text that form discrete entities, which can be defined by means of features such as ‘columns’ or ‘lines’, and characterized in relation to the immediate, more iconographical context.

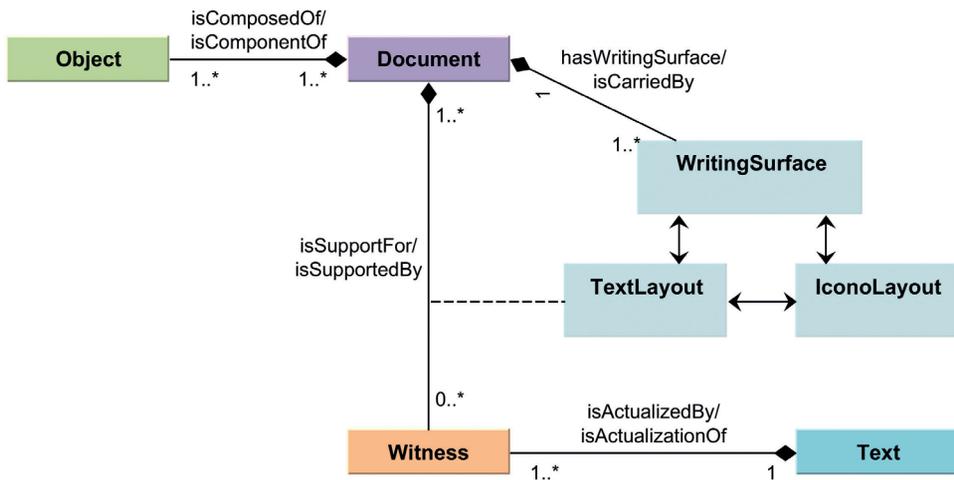


Figure 2: Enriched data model.

¹⁹ At a conceptual level, such cases should be distinguished from the reuse of a single *Document* for different *Witnesses* (with erasures, re-inscriptions, etc.): the relationship between one text-bearing *Document* and many *Witnesses* is perfectly able to handle such cases.

²⁰ See G. Posener, ‘Les ostraca numérotés et le Conte du Revenant’, in *Древний Восток. Сборник 1: к семидесятилетию академика М. А. Коростовцева*, ed. I.S. Katsnel’son (Moscow 1975) 107.

The conceptual data model under examination would thereby allow analyses of the place that a text occupies within its direct writing environment.²¹ This opens up new perspectives for studying the way texts are organized within a monument as well as the manner of their organization.²²

Attributes could also apply to the relationship between *Objects* and *Documents* as well as to the relationship between *Witnesses* and *Texts*. Indeed, much as the notions ‘written surface’ and ‘layout’ enable a precise positioning of *Witnesses* on *Documents*, the position of *Objects* within a *Document* could also be documented by an attribute on relationship ‘isComponentOf’, and the position of *Witnesses* in a *Text* by an attribute on relationship ‘isRealizationOf’. In the latter case, the value of the positioning attribute would have to be defined according to sections or paragraphs as outlined in major synoptic editions of (literary) texts.²³

1.3. Example of the basic data model

In order to illustrate this data model, we will use a literary text from the Ramesside period known as the *Teaching of Amennakhte*.²⁴ This *Text* is a rare *œuvre* for which we can quite safely identify the historical author, Amennakhte, who composed it around 1150 BC.²⁵ One reconstructs — usually through a synoptic edition following the Egyptological scholarly tradition — this ideal *Text* on the basis of approximately twenty *Witnesses*, all of which come from the village of Deir el-Medina and its surroundings.

One of these *Witnesses* occurs on the verso of a *Document* that was reconstructed²⁶ on the basis of two fragments (*Objects*) from the French Institute in Cairo (O. IFAO 1255 A and O. IFAO 1255 B²⁷), and by using a previous transcription made by J. Černý (*Mss.* 17.108, f^o 73–74 = KRI VII, 300,6-302,4) of the O. IFAO 1255 together with O. Varille 39,

²¹ In order to position *Witnesses* on *Documents*, a controlled vocabulary should be specified (see *TopBib* and *Thot* projects in section 5). For two-dimensional objects, labels such as ‘recto’ and ‘verso’, as well as ‘columns’, ‘pages’, etc. should suffice. For objects with a more complex materiality (e.g. temples, tombs, statues, etc.), we are developing thesauri with parts of those objects (e.g. court, wall, register, etc.). This positioning through a controlled vocabulary should be linked to the ‘Typological object part (or “physical feature”)’ in Morlock and Santin, ‘The Inscription between text and object’ (n. 7, above) 341, i.e. ‘a non-detachable part of an object identified with reference to a given epigraphic or archaeological typology.’

²² For instance, this would allow encoding of information such as the presence of a dedication formula on the shoulders of a statue, or of texts on the back-pillar. This would also help in analysing the network of intertextual references at temple scale, as done especially in the so-called Ptolemaic studies.

²³ This will, in a second step, significantly facilitate the alignment of the *Witnesses* of a single *Text*.

²⁴ S. Bickel and B. Mathieu, ‘L’écrivain Amennakht et son *Enseignement*’, *Bulletin de l’Institut Français d’Archéologie Orientale* 93 (1993) 31–51; A. Dorn, ‘Die Lehre Amunnachts’, *Zeitschrift für ägyptische Sprache und Altertumskunde* 131 (2004) 38–55; A. Dorn, ‘Zur Lehre Amunnachts: ein Join und Missing Links’, *Zeitschrift für ägyptische Sprache und Altertumskunde* 140 (2013) 112–125.

²⁵ St. Polis, ‘The scribal repertoire of Amennakhte son of Ipu: Describing variation across Late Egyptian registers’, in *Beyond free variation: Scribal repertoires from Old Kingdom to Early Islamic Egypt*, ed. J. Cromwell and E. Grossman (Oxford 2016).

²⁶ P. Grandet, ‘Un document relatif aux grèves de Deir el-Médinéh en l’an 29 de Ramsès III et un fragment de l’*Enseignement d’Amennakhté*, sections 39–48: O. IFAO 1255 A-B (ONL 514 A-B)’, in *Aere perennius. Mélanges égyptologiques en l’honneur de Pascal Vernus*, Orientalia Lovaniensia Analecta 242, ed. Ph. Collombert, D. Lefèvre, St. Polis and J. Winand (Leuven, 2016) 327–359.

²⁷ When discovered, O. IFAO 1255 B was broken in three pieces. It is however considered here as a single *Object*, since it is nowadays physically ‘non-detachable’ and accordingly received a single inventory number at the IFAO. If the modern reconstruction is proved to be wrong, the TDM allows of course for an *Object* to be split up.

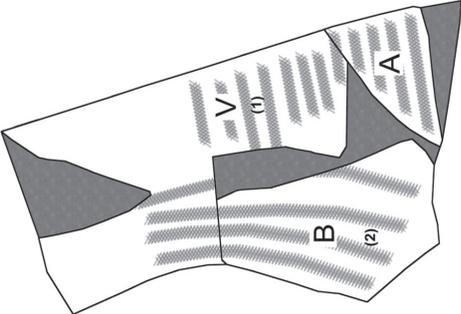
| Object(s) | Document | Witness(es) | Text |
|--|--|---|---|
|  <p>(A)</p> <p>(B)</p> |  | <p>Occurrence of the <i>Teaching of Amennakhte</i> (§39-48) on O. IFAO 1255 B (v^o) [+ V lost]</p> | <p>Abstract version of the <i>Teaching of Amennakhte</i>, viewed here as a synoptic edition</p> |
| <p>O. IFAO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

Figure 3: Illustration of the basic data model with the *Teaching of Amennakhte*.

an *Object* which has since disappeared but for which we have contemporary evidence. The ‘original’ *Document* is thus made up of three *Objects* (one of which is now lost) belonging to two different collections. This illustrates nicely the fact that the element *Document* refers to a reconstructed abstraction, which allows us to define the original *Writing Surfaces* on which the scribe wrote the *Witness(es)* at some point in the past.

This example further allows us to discuss the *n:n* relationship between *Objects* and *Witnesses*. In conceptualizing the writing space of a *Document* as a single complete entity, it is not only possible to link related *Objects* together into a meaningful whole, but it also allows the linking of related *Witnesses* into the same meaningful whole. The *n:n* relationship between *Objects* and *Witnesses* that can be documented with this model overcomes the difficulty of positioning a continuous text *Witness vis-à-vis* its fragmented *Objects*. In the aforementioned example, the *Witness* of the *Teaching of Amennakhte* spans two *Objects*, O. IFAO 1255 B and O. Varille 39, and comes along with the *Witness* of another *Text*, namely an administrative text dealing with the ‘strikes’ that occurred at the end of the reign of Ramesses III (twentieth dynasty). Made of three *Objects*, the *Document* thus preserves two *Witnesses* of two different *Texts*.

Finally, it should be stressed again that the element *Text* is conceptual and therefore plays a similar role as *Document*, since it allows the linking together of *Witnesses* with commensurable textual content (for instance, the twenty witnesses of the *Teaching of Amennakhte*).²⁸

This first illustration of the data model is admittedly simple, since it simply consists of an ostrakon inscribed with a few sentences. However, it would similarly apply to a temple inscribed with dozens of scenes comprising thousands of lines and columns of text: the interplay between concrete (*Object, Witness*) and abstract (*Document, Text*) elements indeed allows a precise description of both the material and textual aspects in all their complexity.

2. Refining the model: metadata enrichment

The data model of section 1 can be enriched with metadata to characterize the four basic elements. Our aim is not to provide a full list of possible descriptors — each project will have its own *desiderata* in this respect — but to discuss some of them to refine the definitions of the elements, and to illustrate the potential of the TDM. The metadata specified here are thus merely suggestive, and should be supplemented as necessary. We begin with metadata pertaining to chronological and spatial dimensions, which apply to the four elements of the TDM (section 2.1). In a second step we examine metadata that are specific to a single element (section 2.2).

2.1. Metadata pertaining to space and time

Metadata related to ‘space’ and ‘time’ have an important role in documenting ancient Egyptian texts: they are needed at all levels of the model in order to situate a piece of writing within its geographical and historical contexts, both ancient and modern.

The distinction between *Object* (defined as a discrete material entity), *Document* (envisioned in relation to an ideal writing surface), and *Witness* (understood as the actualization of a *Text* on 1 to *n Document*), allows us to divide the metadata pertaining

²⁸ Note that our data model does not make any assumptions about the existence of a single *Urtext*, but simply states that one (or several) *Witness(es)* relate(s) to a single element (called *Text*).

to the spatiotemporal dimensions into two broad categories: *place of creation* and *place of provenance* on the one hand, and *date of creation* and *date of discovery* on the other.

| | | <i>Object</i> | <i>Document</i> | <i>Witness</i> | <i>Text</i> |
|--------------|--------------|--|-------------------------------|--------------------------------------|-------------------------------------|
| EVENT | PLACE | Place(s) of provenance | Place of creation | | |
| | | | Place(s) of fabrication & use | Place(s) of inscription & emendation | Place(s) of composition & revision* |
| | DATE | Date(s) of discovery & change(s) of custody | Date of creation | | |
| | | | Date(s) of fabrication & use | Date(s) of inscription & emendation | Date of composition & revision |

Figure 4: Metadata about space and time.

* It is not clear to us whether metadata about the place of composition and emendation would be relevant in the field of studies on the ancient world, but this of course logically possible.

While *place of provenance* and *date of discovery* refer to modern places and times, the broad categories *place of creation* and *date of creation* apply to ancient realities,²⁹ which can be further subdivided depending on the element to which they are connected:

- Place of creation
 - *of a Document = place of fabrication or origin;*
 - *of a Witness = place of inscription or writing;*
 - *of a Text = place of composition.*
- Date of creation
 - *of a Document = date of fabrication or origin;*
 - *of a Witness = date of inscription or writing;*
 - *of a Text = date of composition.*

Thus, the ‘discovery’ of an *Object*, the ‘fabrication’ of a *Document*, the ‘inscription’ of a *Witness*, and the ‘composition’ of a *Text* are all subclasses of *events located in space and time*. Figure 5 captures this observation: the metadata pertaining to space and time are always connected to some *event* that affects an element of the TDM.

We shall not discuss at length all the types of events that can apply to the four elements of the TDM. Besides the ‘discovery’, the ‘fabrication’, the ‘inscription’, and the ‘composition’, any *Object*, *Document*, *Witness*, or *Text* is indeed likely to undergo several changes. As an illustration, a *Witness* can be inscribed and later emended (by the same scribe or another hand), a *Document* can be reused in another context, and an *Object* can have a complicated curatorial history. As illustrated in Figure 5, all these events are accommodated within the framework of the TDM by the one-to-many relationships between elements and events.

²⁹ It is crucial to distinguish the two types of place, as it regularly happens that the provenance of an *Object* differs from the original place of creation of the *Document* or of inscription of the *Witness*. P. Louvre E. 27151, for instance was discovered on the island of Elephantine (near Aswan), but is ‘most probably of Theban origin’, which means that it was probably written and sent by a scribe from the Theban region (see P. Posener-Kriéger, ‘A letter to the governor of Elephantine’, *Journal of Egyptian Archaeology* 64 (1978) 84–87).

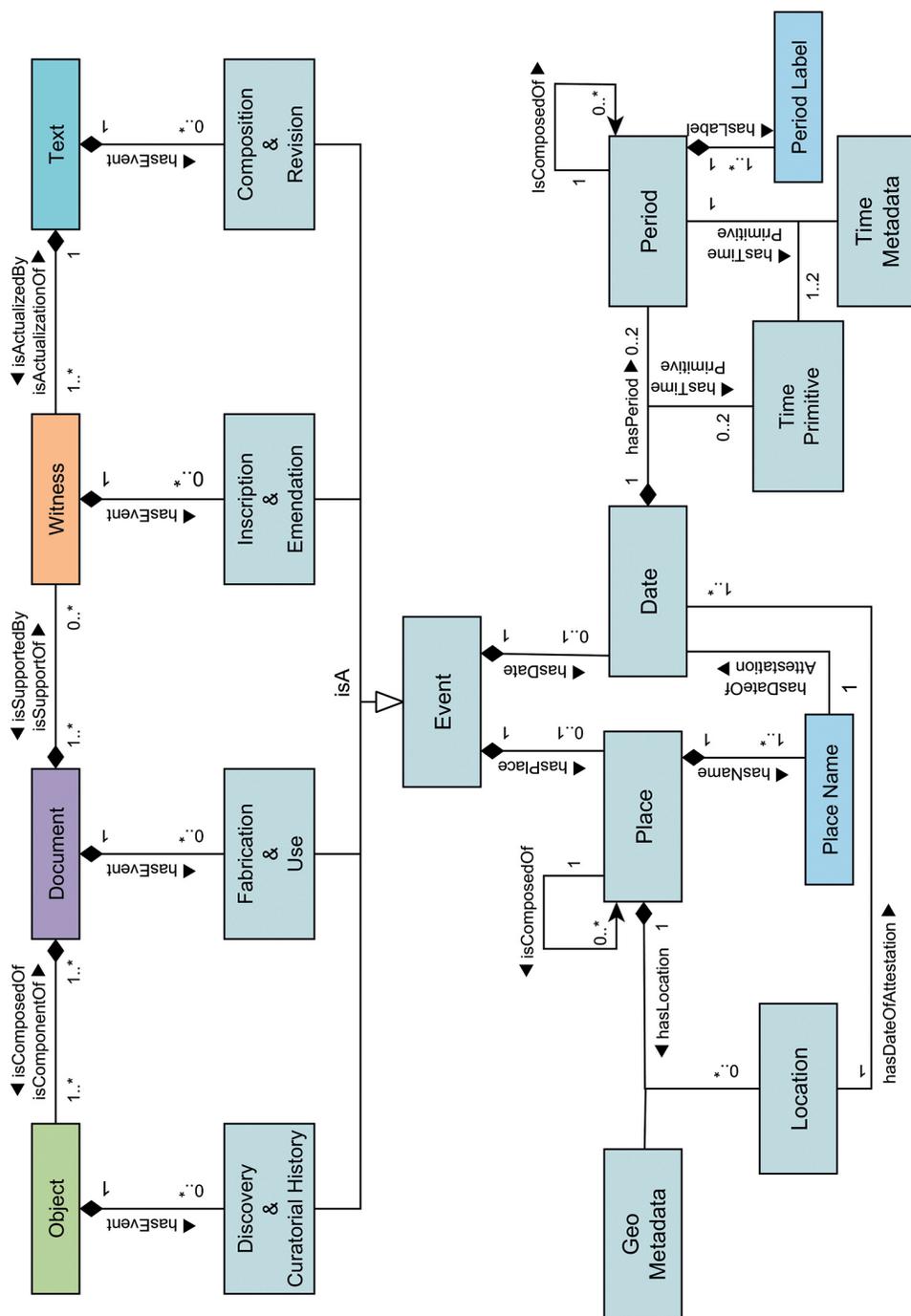


Figure 5: The Thot Data Model: core elements with events located in space and time.

2.1.1. Space related metadata

We understand a place as a geographical ‘thing’ in a very similar way to the Pleiades Data Model;³⁰ that is to say that it can be defined by its association with up to two other classes, namely *Place Name* and *Location* (following Pleiades’ phrasing). However, unlike Pleiades, we argue that a place cannot exist without a name.³¹ A place is a discrete thing on earth, and words can be used to designate it, through proper place names used by locals, the administration, or labels created by archaeologists or epigraphers; either this, or it is something alluded to in texts, and it has, *de facto*, at least one proper name. In the Egyptological context, the name of an archaeological site will be in many cases an Arabic name, often adopted as a transcription in a European language. When located in a site or a monument, places will very often be rendered in the language of those excavating the monument (*e.g.* ‘temple d’Hatchepsout’, ‘Theban Tomb 47’, or ‘Grand puits’ in Deir el-Medina). Whether mentioned in ancient texts, a modern toponym, or a name created by Egyptologists, it is important to keep in mind that a place name only acts as a label for a place and not as an identifier.

The second set of properties appertains to the ‘physical’ reality of a place, which is modelled through a class *Location*. By *Location* we understand a discrete, physically definable space or entity, which can be the extent of a town or that of an archaeological site, a monument, or part of one. Such entities may or may not have a precise, georeferenced location, which is documented in *GeoMetadata*. This flexibility allows us to model places that no longer have a physical reality, such as an ancient tomb that was once excavated but is now lost.³²

The granularity of the above examples shows how any instance of a place may relate to or enclose another place (and *vice versa*): ‘Egypt’ encloses ‘Cairo’, which encompasses ‘Midan Tahrir’, *etc.* What is true for places within the same timeframe (Egypt, Cairo, and Tahrir all belong to modern times), is also valid for places from different timeframes, making the TDM an adequate framework for modelling relationships between ancient and modern places. The TDM thus enables a more accurate modelling of the ancient Egyptian space than other projects,³³ as it clearly decouples ancient places and their names from modern ones. For instance, Husseniya (in the Eastern Delta), is the *Modern Place* in which lies the archaeological site and *Modern Place* Tell el-Far’un, aka Nebesheh. The remains of the *Ancient Place* known as Imet in ancient Egyptian texts are situated here; it is also mentioned as Βουτώ in Herodotus (Hdt. II 75).³⁴ There are, therefore, two distinct types of places, ancient and modern, and the pieces of information about them, *i.e.* location and place names, have to be documented in separate records. In semantic web terms, this would imply that the *Modern Place* Tell el-Far’un would bear a specific URI (and the same for its place names), whereas the *Ancient Place* named Imet would be identified by another URI.³⁵

³⁰ <http://pleiades.stoa.org/help/pleiades-data-model>.

³¹ See E. Relf, *Place and Placelessness* (London 1976), quoted in P. L. Price, ‘Place’, in *The Wiley-Blackwell companion to cultural geography*, ed. N. C. Johnson, R. H. Schein, J. Winders (2013) 122.

³² L. Manniche, *Lost tombs: A study of certain Eighteenth Dynasty monuments in the Theban necropolis*, Studies in Egyptology (London 1988).

³³ Despite the clear statement in the introductory page of Trismegistos Places (http://www.trismegistos.org/geo/about_datastandards.php), this project does not seem to have actually implemented a distinction between ancient and modern places; see example in n. 35.

³⁴ <http://www.perseus.tufts.edu/hopper/text?doc=Perseus:abo:tlg,0016,001:2:75>.

³⁵ This model is being implemented in the *Digital Topographical Bibliography* developed by the Griffith Institute (University of Oxford); see topbib.griffith.ox.ac.uk/dtb.html?topbib=401-150 (modern place Tell el-Far’un) and

2.1.2. Time related metadata

As we envision *Object* as an artefact in its present state, the chronological metadata pertaining to it should relate to events of its modern history (e.g. the date of discovery and other events linked to its museographic ‘life’). The three remaining elements of the TDM can be enriched with metadata that pertain broadly to the concept of ‘date of creation and modification’ (i.e. in ancient times):

- When referring to *Text*, time-related metadata correspond to the (inferred) *date of composition* of a text or to a possible *date of revision* of the same *Text*;
- This can be different from the *date of inscription* (for instance when a scribe copies an older literary text), or from a *date of emendation* of the *Witness*;
- When linked to events concerning the element *Document*, time-related metadata refer to the *date of fabrication* and (possibly) to the *date(s) of reuse* of a *Document*. This is most relevant for monuments, statues, and other objects for which the *date of fabrication* can differ substantially from the *date of inscription*.

Even if information about the events discussed in section 2.1 above is elusive and can often only be inferred in our historical disciplines,³⁶ we think that a careful distinction between the types of events is essential when documenting digital resources, given the various processes they refer to in terms of scribal practices.

2.2. Element specific metadata

In order to document as precisely as possible all aspects of ancient Egyptian texts, various metadata can be used for describing further the elements of the model. In this section, we list and briefly discuss some of them, with the idea that the use and implementation of these metadata might differ according to project priorities:

- In our data model, *Object* embodies the materiality of an inscribed artefact and can be enriched with metadata related to e.g. *Dimensions*, *Material*,³⁷ and *State of preservation*;
- *Document* being the idealized reconstruction of a text-bearing object, it is the element on which typological characterization (*Object type*, such as coffin, ostrakon, papyrus, statue, stela, temple, etc.) as well as *Reconstructed dimensions* will be applied;
- As physical representation of a *Text*, *Witness* is an element that can be characterized in terms of *Scripts*, *Scribes*, *Paleographical features*, *Modes of inscription*, etc.,³⁸

<http://topbib.griffith.ox.ac.uk/dtb.html?topbib=901-119-001> (ancient place Imet). The full range of geographical data (namely places and place names) relating to these places is recorded under a single TM entry (<http://www.trismegistos.org/place/2824>).

³⁶ In this respect, the data model allows for time spans (Terminus Post Quem–Terminus Ante Quem).

³⁷ K. E. Piquette, ‘Re-materializing script and image’ (n. 5, above) 96 (and elsewhere later on) uses the distinction of J. Gibson, *The Ecological Approach to Visual Perception* (1979) 16, who distinguishes between Medium (sensory perception, etc.), Substance (the resistant thing in the world), and Surface. *Material* metadata are linked to the Substance of objects.

³⁸ Whether by addition (applied colour, inlay, etc.), subtraction (carving, incision, etc.), or transformation (impression in clay, fired and unfired, etc.).

- To *Text* are attached metadata referring *e.g.* to *Author*,³⁹ *Language*, and *Text genre*.

In order to fully grasp the materiality of texts, visual representations can be appended to the elements of the TDM. In the near future, three-dimensional images are likely to become more and more common for several types of objects.⁴⁰

3. *An illustration of the benefits expected from the data model as regards spatial domain*

In this section, we discuss some new avenues for research that the conceptual model opens up, focusing on the inferences that can be made in the *spatial* domain of text analysis. The chain of knowledge (from *Object* to *Text*) resulting from the model is particularly promising and major outcomes can be expected regarding toponymy and the study of ancient Egyptian place names.

A first prospect is the study of the geographical distribution of ancient place names, which is made possible by the link (through the concept of a *Document*) between the georeferenced provenance of an *Object* and the place name(s) mentioned in the *Witness*. This would open up possibilities in terms of navigation through a space ‘populated’ with place names linked to the locations of their occurrences. Of course, the occurrence of a place name on an *Object* with a known provenance does not imply that the object’s place of provenance was named as such (place names can occur in a text for very many reasons). Nonetheless we believe this feature will open up new perspectives in the study of ancient Egyptian toponyms, since this is an ideal starting point for locating a place based on inferences from distribution patterns. Additionally, the TDM would allow mapping of relationships between ancient places by drawing links between the places and the incidences of their names.

A second perspective is toponymy as a linguistic field, and the opportunity offered by the data model of correlating ancient place names with ‘surrounding’ modern place names. It is again the *Document* that acts as glue, this time between the occurrences of ancient place names in a *Witness* and the modern place names attached to the *Object*’s provenance. The goal is not to establish a connection with the name of the provenance alone, but rather to allow comparison between ancient place names and the full set of modern toponyms that are located around the provenance of a *Document*, extracted for instance from GeoNames.⁴¹ This would help when compiling data samples for identifying ancient Egyptian place names with modern Arabic ones.

A third possibility could be the study of cases for which the ‘place of origin’ of the *Document* and the ‘place of provenance’ of the *Object* do significantly differ. This would indeed give us some insight into the movements of (text-bearing) objects in ancient Egypt, which, although a fundamental aspect of archaeological research, are hardly quantifiable at the moment.

These research questions are merely illustrative of the types of inquiry that would be possible if the TDM were fully implemented. In this respect, the ability to link this model to online resources is crucial. This point is developed in the conclusions.

³⁹ On the problematic notion of *Author* (especially in the ancient Egyptian cultural environment), see St. Polis, ‘The scribal repertoire of Amennakhte son of Ipu’, with references to previous literature on the topic.

⁴⁰ See *e.g.* the toolbox developed by the ‘Digital Epigraphy and Archaeology’ project for the squeezes of Greek and Latin inscriptions: <http://www.digitalepigraphy.org/toolbox/info.html>.

⁴¹ Using for instance the API ‘FindNearBy’, see <http://www.geonames.org/export/web-services.html#findNearby>.

4. Mapping with other data models

Data models developed by other projects dealing with cultural heritage and ancient texts have been an inspiration while building the TDM.⁴² Although some elements can be easily matched between different models, a general mapping is difficult as significant dissimilarities appear when considering the relationships and hierarchies between classes across conceptual models.

4.1. CIDOC-CRM

This is particularly true for the obvious candidate in the domain of cultural heritage, *i.e.* the CIDOC-CRM, the conceptual reference model developed by the International Council of Museums.⁴³

Three of the four core elements of the TDM match CIDOC-CRM elements rather unambiguously. The elements *Document*, *Object*, and *Text* find an equivalent respectively in CIDOC-CRM *E84 Information Carrier*, *E22 Man-Made Object*, and *E33 Linguistic Objects*. However, this matching raises difficulties since in CIDOC-CRM, *E84* is a subclass of *E22*, which implies that *E84* forms part of (CIDOC-CRM P148i) *E22*, whereas in the TDM *Document* and *Object* are two disjoint classes, the former being linked to the latter by the inverse property (see Figure 6 and section 1.1).

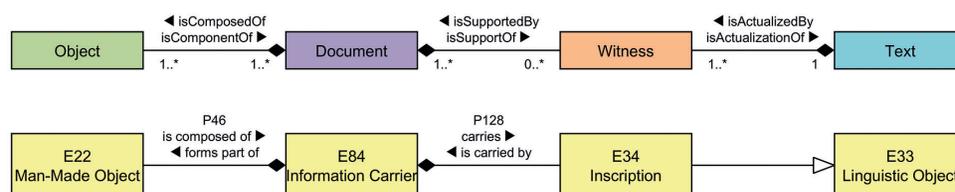


Figure 6: UML view of the Main CIDOC-CRM Elements mapped on to the TDM.

No strict equivalent to the element *Witness* can be found in the current version of CIDOC-CRM, despite the existence of the CRM class *E34 Inscription*. The scope note attached to this class speaks against an identification, as ‘this class comprises recognisable, short texts attached to instances of *E24 Physical Man-Made Thing*’. An adjective such as ‘short’ is not a suitable qualifier when referring to the *Witnesses* in our field, where inscriptions can be made of up to hundreds of lines of hieroglyphs. Nevertheless, an extension of the CIDOC-CRM devoted to epigraphy is currently under development and seems to refer to the concept attached to *Witness* more adequately, by providing the class *TX1 Written Text*.⁴⁴

The TDM class *Event* and its four subclasses are close matches of CIDOC-CRM *E92*

⁴² The links between the TDM and the FRBRoo ontology (http://www.cidoc-crm.org/frbr_inro.html) shall be discussed in a forthcoming paper dealing with the question of intertextual relationships in digital textual editing.

⁴³ <http://www.cidoc-crm.org/>. Mapping was made with version 6.2 (http://www.cidoc-crm.org/docs/cidoc_crm_version_6.2.pdf). For a comprehensive overview, see D. Oldman, *The CIDOC Conceptual Reference Model (CIDOC-CRM): primer*, version 1, July 2014 (http://www.cidoc-crm.org/docs/CRMPrimer_v1.1.pdf).

⁴⁴ See A. Felicetti *et al.*, ‘CIDOC CRM and Epigraphy: a Hermeneutic Challenge,’ at the 34th joined meeting of the CIDOC-CRM SIG, Heraklion, Crete, 6–9 October 2015 available at http://www.cidoc-crm.org/docs/34th-meeting-presentations/CRM_epigraphy.ppt.

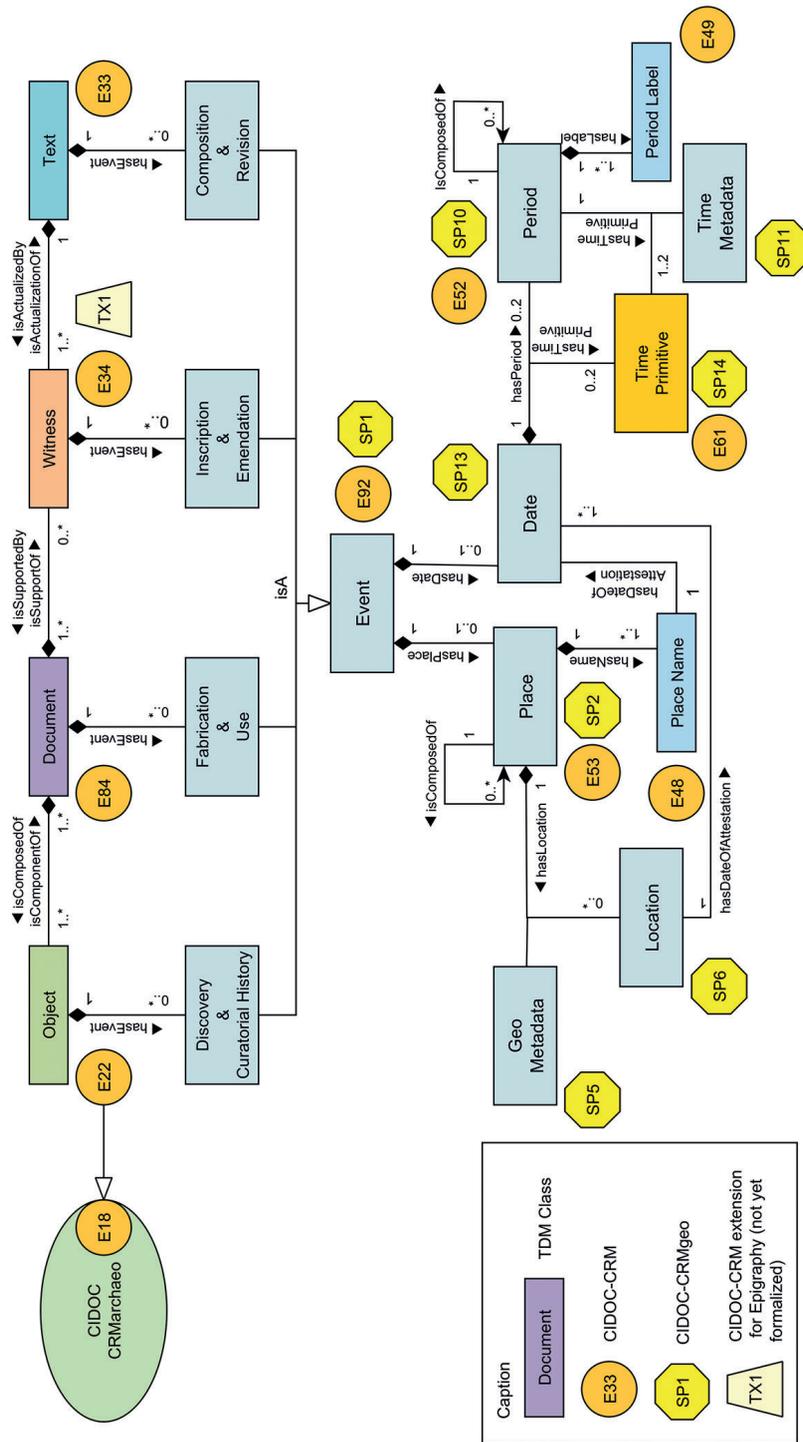


Figure 7: Mapping the TDM with CIDOC-CRM: A general overview.

Space-Time Volume and extension CRMgeo⁴⁵ *SP1 Phenomenal Space-Time Volume*, as they also encompass both spatial and temporal dimensions. SP1 is an interesting candidate as it models material, real-world phenomena (which would be, in our context, the discovery and history of the *Object*, the fabrication and use of the *Document*, the writing and emendation of the *Witness*, and the composition and revision of the *Text*). The only reservation that one can have is the fact that *Place* in the TDM is a concept slightly different from *E53 Place*, which is a ‘purely geometric notion’,⁴⁶ and is therefore more in line with Pleiades, which allows places to be unlocated, an obvious requisite in the ancient world context (see above, section 2.1).

An interesting point regarding possible connections between the TDM and CIDOC-CRM is that the mapping of TDM *Object* onto CIDOC-CRM *E22 Man-Made Object* would allow a precise contextualization of any written *Document* in its archaeological context by plugging the TDM into CIDOC-CRM *Marchaeo*,⁴⁷ an ontology devoted to the documentation of archaeological resources.

Further mappings with CIDOC-CRM elements are illustrated in Figure 7.

4.2. TEI/Epidoc

Since ancient texts are at the core of the TDM, TEI⁴⁸ (and its Epidoc specification⁴⁹) is the second obvious candidate in terms of data model mapping (although neither TEI nor Epidoc comes with an explicit data model). Furthermore, as a well-established set of recommendations for text encoding, mapping onto TEI is a good way to investigate how the TDM could find concrete implementations.

The four core elements of the TDM, as well as most of the associated classes relating to space and time, are easily accommodated within the TEI environment:⁵⁰ three of the four core elements of the TDM find their place in the TEI Header, whereas one, namely the *Witness*, can be associated with the TEI header as well as with the TEI body — as a view (*i.e.* an edition or interpretation⁵¹) of an inscription, it naturally finds its place within the TEI body itself. The four core elements are documented in the TEI header section `/TEI/teiHeader/fileDesc/sourceDesc/msDesc`. The fact that bits of information pertaining to a single class of the TDM are spread over several sections shows, however, that the data model underlying TEI is quite different. The overall mapping can be summarized as follows:

⁴⁵ <http://www.ics.forth.gr/isl/CRMext/CRMgeo/docs/TR435-CRMgeo.pdf>.

⁴⁶ See *Definition of the CIDOC conceptual reference model*, version 6.2, 15, s.v. E27 site.

⁴⁷ Connection made possible by the fact that *E22 Man-Made Object* is a subclass of *E19 Physical Object* and subsequently of *E18 Physical Thing*, which appears in CRM *Marchaeo* (http://www.ics.forth.gr/isl/CRMext/CRMMarchaeo/docs/CRMMarchaeo_v1.4.pdf).

⁴⁸ <http://www.tei-c.org/>.

⁴⁹ <https://sourceforge.net/p/epidoc/wiki/Home/>.

⁵⁰ Compare with the previous suggestions of mapping made by Morlock and Santin, ‘The Inscription between text and object’ (n. 7, above) 343–344.

⁵¹ For the crucial distinction between *Witness* and the text encoded, which involves a lot of interpretative and editorial decisions, see Morlock and Santin, ‘The Inscription between text and object’ (n. 7, above) 340–341.

| TEI element | TDM Class or metadata |
|---|--|
| /TEI/teiHeader/fileDesc/sourceDesc/ msDesc | <i>Document</i> , as a whole |
| /TEI/text/ body | <i>Witness</i> , as an interpretation/view |
| /TEI/teiHeader/fileDesc/sourceDesc/msDesc/physDesc/ handDesc and / script | <i>Witness</i> , or metadata dealing with handwriting and script |
| /TEI/teiHeader/fileDesc/sourceDesc/msDesc/ msFrag | <i>Object</i> , as physically discrete parts (objects, blocks, artefacts, etc.) of a <i>Document</i> |
| /TEI/ sourceDoc | <i>Writing surfaces</i> , on which <i>Witnesses</i> are inscribed |
| /TEI/teiHeader/fileDesc/sourceDesc/msDesc/physDesc/objectDesc/ layoutDesc | <i>Text Layout</i> |
| /TEI/teiHeader/fileDesc/sourceDesc/msDesc/msContents/ msItem | <i>Text</i> |

Figure 8: Mapping of TDM with TEI elements.

4.3. A similar conceptual model: Eagle Europeana

Eagle Europeana⁵² is an initiative that aggregates data sets from many Greek and Latin epigraphic projects in order to provide a single user-friendly portal to the inscriptions of the ancient world. Interestingly, this project uses CIDOC-CRM (section 4.1) and TEI/Epidoc (section 4.2) as standards and *lingua franca* for ingesting the heterogeneous resources.

The CIDOC-CRM Conceptual Model for Eagle has been developed based on an analysis of several mappings between CIDOC-CRM and databases of the projects' partners. As such, it logically uses the CIDOC-CRM elements mapped onto the TDM in section 4.1, namely *E22*, *E84*, *E34*, and *E33*. As noted above, the hierarchical structure between these elements implies that the overall structure of the Eagle model differs from the TDM, as *E34 Inscription* is carried by (CIDOC-CRM property *P128i*) *E84 Information Carrier*, which is itself part of *E22 Man-Made Object* (see Figure 6).

Besides the CIDOC-CRM conceptual Model for Eagle, it also developed an Aggregator Conceptual Model (EACM) so as to enable the integration of metadata from different data sets. As regards the core elements, the EACM is not very different from the TDM: the classes *Eagle Object* and *Artefact* appears to be a good equivalent for the TDM *Document* and *Object* (especially because the former is linked to a class *Inscription* that matches well with the TDM *Witness*).

5. Conclusions and perspectives

The goal of the TDM is first and foremost to allow an accurate description of the ancient Egyptian textual material within digital projects, by using the set of metadata that apply to the elements and to their relationships respectively. As illustrated above (section 4.2), the elements of the TDM can be mapped onto TEI elements, which means that any implementation of this data model would enable creation of TEI compliant .xml files.

⁵² <http://www.eagle-network.eu>; see in particular 3.1.2, M.-A. Sicilia *et al.*, *Eagle Metadata Model Specification*, version 1.1, [2015] (http://www.eagle-network.eu/wp-content/uploads/2013/06/EAGLE_D3.1_EAGLE-metadata-model-specification_v1.1.pdf).

An implementation of the TDM with a linked open data perspective is certainly the best way to go. However, without a set of well-identified resources, it will be difficult to instantiate it properly in this semantic web perspective: the ability to refer to (or to create) *Unique Resource Identifiers* (URIs) is a key issue, which has to be handled in parallel with implementation of the model. Among the resources that are already (or will soon become) available are:

- The stable identifiers provided by the *Trismegistos* project for *Documents*. Even if *Trismegistos*' definition of *Document* does not completely match the definition provided here (see section 1.1 and n. 6 above),⁵³ the role of this platform in studies of written documents from the ancient world makes it a mandatory point of reference.
- For places and place names, the *TopBib* project of the Griffith Institute, University of Oxford (<http://topbib.griffith.ox.ac.uk>) will certainly become an essential resource. The core aim of this project is the production of a *Digital Topographical Bibliography* of Ancient Egypt, a digital version of a bibliographical tool that is more than one hundred years old. It implements the geographical concepts described in this paper (section 2.1), and provides the framework and URIs to point to places of provenance and creation, from the top-level granularity of a region or site,⁵⁴ to that of a room or section of a room, and even to those of artefacts.
- A third project will soon provide strong support for the full implementation of the data model: the *Thot* project (*Thesauri and Ontology for documenting ancient Egyptian textual resources*).⁵⁵ This international project has been jointly developed by the Berlin-Brandenburg and Leipzig Academies of Sciences and Humanities, and the Department of Egyptology at the University of Liège (where it is hosted), thanks to a grant from the Humboldt Foundation.⁵⁶ *Thot* is conceived of as a multilingual repository of resources for metadata enrichment and data exchange. The resources are SKOS⁵⁷-compliant thesauri aimed at providing URIs for most of the metadata elements required by the conceptual data model under discussion here. Supplying descriptors in English, French, German, and Arabic, *Thot* concepts will also be mapped to various existing thesauri and online resources, such as Eagle, Trismegistos, the Getty Art and Architecture Thesauri,⁵⁸ and Ariadne resources.⁵⁹

F.R.S.–FNRS, University of Liège

⁵³ The main IDs attributed by the *Trismegistos* project are indeed similar to what is referred to as a *Document* in the TDM: 'in principle all texts written on what was in antiquity a single writing surface belong together and form one document receiving a single Trismegistos number (...)' (http://www.trismegistos.org/about_identifiers.php). See, however, n. 6.

⁵⁴ The *TopBib* project is in the process of implementing georeferenced data for archaeological sites, which will thus provide easy access through maps and open up new perspectives in terms of text analysis (see above section 3).

⁵⁵ <http://thot.philo.ulg.ac.be/>.

⁵⁶ The Anneliese Maier Forschungspreis awarded to Jean Winand (<https://www.humboldt-foundation.de/web/amp-2015.html>).

⁵⁷ For a primer about Simple Knowledge Organization System, essential for building up thesauri, see <https://www.w3.org/TR/skos-primer/>.

⁵⁸ <http://www.getty.edu/research/tools/vocabularies/aat/>.

⁵⁹ <http://www.ariadne-infrastructure.eu/>.