Ancient Egyptian texts in context
Towards a conceptual data model (the Thot Data Model – TDM)*

Stéphane Polis & Vincent Razanajao
(F.R.S.-FNRS – University of Liège)

Abstract. In this paper, we propose a conceptual data model that could be the basis for future implementations of databases and digital corpora of Ancient Egyptian texts that fully integrate the material dimensions of writing. The types of metadata that can be used for documenting the elements and relationships of this model are discussed and the resources (URIs) available for its online implementation (in the perspective of the ‘linked open data’ movement) are examined.

In Ancient Egypt, writing is everywhere: from religious and funerary buildings down 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 4th millennium BC), up until the last during the Ptolemaic and Roman periods (4th century BC-4th century AD).

Despite the inextricable link between the texts and their supports1 (whether monuments, smaller artefacts or even the physical landscape), current Egyptological text-oriented projects2 often fail to handle satisfactorily both dimensions together, and generate a significant amount of data about texts that are (at least partly) divorced from their contexts.3 For this reason,

* 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.


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: Subservience 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...
ancient Egyptian textuality is difficult to apprehend within its full (present) archaeological and (past) cultural environment, especially when using digital resources.

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. 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 archaeological object bearing text”.}

Naucratis, 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 Naucratis’, British Museum Studies in Ancient Egypt and Sudan 17 (2011) 35).

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 Pyramidentextes [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 & N. Grimal, vol. 1 [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/tn/mn) dans les Textes des Pyramides’, in Aere perennius. Mélanges égyptologiques en l’honneur de Pascal Vernus, Orientalia Lovaniensia Periodica 242, ed. Ph. Collomert, D. Lefèvre, St. Polis & 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 80 years after the first volume of inscriptions was released (Em. Chassinat, Le temple de Dendara [Cairo, 1934-1935]), all the more given the significance of spatial arrangement of texts in temples from the Greco-Roman period is known since the 1960s (Ph. Derchain, ‘Un Manuel de géographie liturgique à Edfou’, Chronique d’Égypte 37/73 [1962] 31-65).

In Egyptology, a field of study that envisions explicitly writing as a material practice has recently emerged. It provides empirical and methodological observations that are directly relevant for our purpose. For the material aspect, see e.g. K.E. Piquette, ‘Re-materializing script and image’, in Current Research in Egyptology 2008, ed. V. Gashe & J. Finch (Bolton, 2008) 89-107, as well as several contributions in K.E. Piquette & R.D. Whitehouse (ed.), Writing as Material Practice. Substance, surface and medium (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 Antefiker (TT60) with newly recorded graffiti), Studien zur Altägyptischen Kultur 42 [2013] 269-323).

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.

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

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 Text

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

- **Object**, which refers to a physically discrete material object (in its present state, based on current available evidence).
- **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.

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 markup 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) (p. 76-81). This point shall be discussed in Section 1.

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

10 A first paper on the topic was presented at the *Text Encoding Initiative conference and member’s meeting 2013* (October 28-31, Lyon – France) by L. Coulon, F. Elwert, E. Morlock, S. Polis, V. Razanajao, S. Rosmorduc, S. Schweitzer, and D. Werning.

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

12 These four elements are identified (although with other labels) by Morlock & Santin, ‘The Inscription between text and object’, 141 (see n. 12-15 below). The main originality of the TDM lays therefore in the explicit definition of all the possible types of relationships between these elements.

13 Compare with the ‘Physical object part’ defined as ‘a detachable physical part of a material object that can be physically isolated’ in Morlock & Santin, ‘The Inscription between text and object’, 341.
• **Witness**, which refers to a single occurrence of a **Text**, in its material (and more broadly philological) dimensions, on a **Document**.\(^{14}\)

• **Text**, which refers to a textual composition as it can be reconstructed from (the compilation of 1 to \(n\)) **Witness**(es).\(^{15}\)

![Diagram](image.png)

Fig. 1. The four basic elements and their relationship

Among these four elements, two refer to (nowadays) concrete dimensions, namely **Object** (as a discrete material object) and **Witness** (as the actual occurrence of a text), while two refer to abstract (or idealized) 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 here a modern perspective, but that the elements **Document** and **Text** were obviously ‘concrete’ entities at some point in the cultural environment of ancient Egypt: the metadata that describes 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**) and the elements **Document** and **Witness** (a **Document** can bear several **Witnesses**), the many-to-many relationship between the elements **Object** and **Document** requires further

---

\(^{13}\) Compare with ‘Text-bearing object’ defined as ‘a material object that bears one or several inscribed texts’ in Morlock & Santin, ‘The Inscription between text and object’, 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 to its **Writing surface**.

\(^{14}\) Compare with the ‘Inscribed entity’ defined as ‘the set of marks that were inscribed on a material support’ in Morlock & Santin, ‘The Inscription between text and object’, 341.

\(^{15}\) Compare with the ‘Abstract text’ defined as ‘an abstract entity corresponding to the ’object of thought’ that is the denotata of the inscribed entity or its intellectual content’ in Morlock & Santin, ‘The Inscription between text and object’, 341.

\(^{16}\) See D. Buzzetti, ‘Digital representation and the text model’, 76-81 and the discussion above in n. 7.
This relationship 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 Kamose\(^\text{17}\) (c. 1550 BCE), which was originally a doorjamb within a monument of Senusret I (c. 1950 BCE).\(^\text{18}\) 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 re-shaped 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 Documents (a doorjamb and a stela), each receiving specific Witnesses.\(^\text{19}\)

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 the current materiality of a text bearer (be it made of several objects or in a poor state of preservation).

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.

\(^\text{17}\) 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).


\(^\text{19}\) At a conceptual level, such cases should be distinguished from the reuse of a single Document for different Witnesses (with erasures, reinscriptions, etc.): the one-to-many relationship between one text-bearing Document and many Witnesses is perfectly able to handle such cases.
The conceptual data model under examination would thereby allow analyses of the place that a text occupies within its direct writing environment. This opens up new perspectives for studying the way texts are organized within a monument as well as the manner of their organization.

Attributes could also apply to the relationship between Objects and Document as well as to the relationship between Witnesses and Text. 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 defined in major synoptic editions of (literary) texts.

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. This Text is a rare oeuvre for which we can quite

---

20 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 & Santin, ‘The Inscription between text and object’, 341, i.e. ‘a non detachable part of an objet identified with reference to a given epigraphic or archaeological typology.’

21 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.

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

safely identify the historical author, Amennakhte, who composed it around 1150 BC.\textsuperscript{24} 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.

<table>
<thead>
<tr>
<th>Object(s)</th>
<th>Document</th>
<th>Witness(es)</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A)</td>
<td>Reconstructed original potsherd ostracon based on the three Objects</td>
<td>Occurrence of the Teaching of Amennakhte (§39-48) on O. IFAO 1255 B (\textsuperscript{v}) [\textsuperscript{+ V} lost]</td>
<td>Abstract version of the Teaching of Amennakhte, viewed here as a Synoptic edition</td>
</tr>
<tr>
<td>O. IFAO 1255 A (\textsuperscript{v}) + O. IFAO 1255 B (\textsuperscript{v}) + (O. Varille 39 [= V]; lost)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

One of these Witnesses occurs on the verso of a Document that was reconstructed\textsuperscript{25} on the basis of two fragments (Objects) from the French Institute in Cairo (O. IFAO 1255 A and O. IFAO 1255 B\textsuperscript{26}), and by using a previous transcription made by J. Černý (Mss. 17.108, \textsuperscript{f} 73-74 = KRI VII, 300,6-302,4) of the O. IFAO 1255 together with O. Varille 39, an Object which has since then disappeared but for which we have contemporary evidence. The ‘original’ Document is thus made up of three Objects (one of which is now lost) that belong to two different collections. This illustrates nicely the fact that the element Document refers to a


\textsuperscript{26} 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 an Object to be split up.
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 linking 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 (20th 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 linking together *Witnesses* with commensurable textual content (for instance, the twenty witnesses of the *Teaching of Amennakhte*).\(^{27}\)

This first illustration of the data model is admittedly simple, since it just consists of an ostracaon 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 describing precisely 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 (§2.1). In a second step we examine metadata that are specific to a single element (§2.2).

\(^{27}\) 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*).
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 a Document), allows us to divide the metadata pertaining 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.

<table>
<thead>
<tr>
<th>EVENT</th>
<th>OBJECT</th>
<th>DOCUMENT</th>
<th>WITNESS</th>
<th>TEXT</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLACE</td>
<td>Place(s) of provenance</td>
<td>Place(s) of fabrication &amp; use</td>
<td>Place(s) of inscription &amp; emendation</td>
<td>Place(s) of composition &amp; revision</td>
</tr>
<tr>
<td>DATE</td>
<td>Date(s) of discovery &amp; change(s) of custody</td>
<td>Date(s) of fabrication &amp; use</td>
<td>Date(s) of inscription &amp; emendation</td>
<td>Date of composition &amp; revision</td>
</tr>
</tbody>
</table>

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, 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.

---

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

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).
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.

![Diagram](image)

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

We shall not discuss at length all the types of events that can apply to the four elements of the TDM. Beside 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 re-used 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.

2.1.1. Space related metadata

We understand a place as a geographical ‘Thing’ in a very similar way to the Pleiades Data Model,\(^\text{30}\) that is to say that it can be defined by its association to 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.\(^{31}\) either 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; or it is something alluded to in texts, and it has, \textit{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 modeled through a class \textit{Location}. By \textit{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 a part of it. Such entities may or may not have a precise, georeferenced location, which is documented in \textit{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 may 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 \textit{Modern Place} in which lies the archaeological site and \textit{Modern Place} Tell el-Far‘un, aka Nebesheh. The remains of the \textit{Ancient Place} known as Imet in ancient Egyptian texts, are situated here; it is also mentioned as Bourtô in Herodotus (II, LXXV).\(^{34}\) There are, therefore, two distinct types of places, ancient


\(^{33}\) 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.

and modern, and the pieces of information about them, i.e. location and place names, have to be documented in separated records. In semantic web terms, this would imply that the *Modern Place* Tell el-Farʿun would bear a specific URI (and the same for its place names), whereas the *Ancient Place* named Imet would be identified by another URI.\(^{35}\)

### 2.1.2. Time related metadata

As we envision *Object* as an artifact 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 the 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*,
- which 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 re-use* 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:

---

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-Farʿun) and 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).
• In our data model, *Object* embodies the materiality of an inscribed artefact and can be enriched with metadata related to e.g. *Dimensions*, *Material*, 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, ostracon, 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.

• To *Text* are attached metadata referring e.g. to *Author*, *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, 3D images are likely to become more and more common for several types of objects.

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

---

37 K.E. Piquette, ‘Re-materializing script and image’, 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, paste, inlay, etc.), subtraction (carving, incision, etc.) or transformation (impression in clay or past, fired and unfired, etc.).

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

40 See e.g. the toolbox developed by the ‘Digital Epigraphy & Archaeology’ project for the squeezes of Greek and Latin inscriptions: http://www.digitalepigraphy.org/toolbox/info.html.
ideal starting point for locating a place based on inferences from distribution patterns. Additionally, the TDM would allow to map relationships between ancient places by drawing links between the ancient places and the occurrences of their names.

A second perspective is toponymy as a linguistic field, and the opportunity offered by the data model to correlate 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.\textsuperscript{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 about the movements of (text bearing) objects in Ancient Egypt, a fundamental aspect of archaeological research, which are hardly quantifiable at the moment.

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

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. 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 Museum.\textsuperscript{42}

\textsuperscript{41} Using for instance the API ‘FindNearBy’, see http://www.geonames.org/export/web-services.html#findNearby.
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 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 Fig. 6 and Section 1.1).

![Fig. 6. UML view of the Main CIDOC-CRM Elements mapped on to the TDM](image)

No strict equivalent to the element Witness can be found in the current version of CIDOC-CRM, despite the existence of a 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 a class TX1 Written Text.43

The TDM class Event and its four subclasses are close matches of CIDOC-CRM E92 Spacetime volume and extension CRMgeo44 SP1 Phenomenal Spacetime 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 than 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”, and is therefore more in line with Pleiades, which allows places to be un-located, an obvious requisite in the Ancient World context (see above, Section 2.1).

An interesting point regarding possible connexions 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 onto CIDOC-CRM Archaeo, an ontology devoted to the documentation of archaeological resources.

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

![Diagram](image)

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

### 4.2. TEI/Epidoc

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

---


46 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 CRMARcheo (http://www.ics.forth.gr/isl/CRMext/CRMarchaeo/docs/CRMarchaeo_v1.4.pdf).


48 [https://sourceforge.net/p/epidoc/wiki/Home/](https://sourceforge.net/p/epidoc/wiki/Home/).
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.\(^\text{49}\) three of the four core elements of the TDM find their place in the TEI Header, whereas one, namely the \textit{Witness}, can be associated with the TEI header as well as with the TEI body — as a view (i.e., an edition or interpretation\(^\text{50}\)) 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:

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

![Fig. 8. Mapping of TDM with TEI elements](image)

4.3 A similar conceptual model: Eagle Europeana

Eagle Europeana\(^\text{51}\) 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 \textit{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,

---

\(^{49}\) Compare with the previous suggestions of mapping made by Morlock & Santin, ‘The Inscription between text and object’, 343-344.  
\(^{50}\) For the crucial distinction between \textit{Witness} and the text encoded, which involves a lot of interpretative and editorial decisions, see Morlock & Santin, ‘The Inscription between text and object’, 340-341.  
it logically uses the CIDOC-CRM elements mapped on to 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 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 Fig. 6).

Beside 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 class Eagle Object and Artifact 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 on to TEI elements, which means that any implementation of this data model would enable creation of TEI compliant .xml files.

An implementation of the TDM in 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 instatiate 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),52 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.griffth.ox.ac.uk) will certainly become an essential resource. The core aim of this project is the production of a Digital Topographical Bibliography of

52 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.
Ancient Egypt, a digital version of a more than one-hundred year old bibliographical tool. 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 a site,\(^{53}\) to that of a room or section of a room, and even to that 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*).\(^{54}\) This international project is 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 of the Humboldt Foundation.\(^{55}\) Thot is conceived as a multilingual repository of resources for metadata enrichment and data exchange. The resources are *SKOS*\(^{56}\) 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,\(^{57}\) or Ariadne resources.\(^{58}\)

---

\(^{53}\) 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).

\(^{54}\) http://thot.philo.ulg.ac.be/


\(^{56}\) For a primer about Simple Knowledge Organization System, essential for building up thesauri, see https://www.w3.org/TR/skos-primer/.

\(^{57}\) http://www.getty.edu/research/tools/vocabularies/aat/.

\(^{58}\) http://www.ariadne-infrastructure.eu/