COLLECTION
ÆGYPTIACA LEODIENSIA

La collection *Ægyptiaca Leodiensia* — dirigée par Jean Winand, Dimitri Laboury et Stéphane Polis — a pour vocation de publier des travaux d'égyptologie dans les domaines les plus divers. Elle accueille en son sein des monographies ainsi que des volumes collectifs thématiques.

This volume represents the outcome of the meeting of the Computer Working Group of the International Association of Egyptologists (Informatique & Égyptologie) held in Liège in 2010 (6-8 July) under the auspices of the Ramses Project. The papers are based on presentations given during this meeting and have been selected in order to cover three main thematic areas of research at the intersection of Egyptology and Information Technology: (1) the construction, management and use of Ancient Egyptian annotated corpora; (2) the problems linked to hieroglyphic encoding; (3) the development of databases in the fields of art history, philology and prosopography. The contributions offer an up-to-date state of the art, discuss the most promising avenues for future research, developments and implementation, and suggest solutions to longstanding issues in the field.
Two general trends characterize the projects laid out here: the desire for online accessibility made available to the widest possible audience; and the search for standardization and interoperability. The efforts in these directions are admittedly of paramount importance for the future of Egyptological research in general. Indeed, for the present and increasingly for the future, one cannot overemphasize the (empirical and methodological) impact of a generalized access to structured data of the highest possible quality that can be browsed and exchanged without loss of information.

Stéphane POLIS is Research Associate at the National Fund for Scientific Research (Belgium). His fields of research are Ancient Egyptian linguistics and Late Egyptian philology and grammar. His work focuses on language variation and language change in Ancient Egyptian, with a special interest for the functional domain of modality. He supervises the development of the Ramses Project at the University of Liège with Jean Winand.

Jean WINAND is professor ordinarius at the University of Liège, and currently Dean of the Faculty of Philosophy and Letters. He specializes in texts and languages of ancient Egypt. His major publications include *Études de néo-égyptien. La morphologie verbale* (1992); *Grammaire raisonnée de l'Égyptien classique* (1999, with Michel Malaise); *Temps et Aspect en égyptien. Une approche sémantique* (2006). He launched the Ramses Project in 2006, which he supervises with Stéphane Polis.

PRESSES UNIVERSITAIRES DE LIÈGE

ISBN : 978-2-87562-016-3
9 782875 620163

ÆGYPTIACA LEODIENSIA
9
Texts, Languages & Information Technology in Egyptology
Stéphane POLIS – Jean WINAND

COLLECTION
ÆGYPTIACA LEODIENSIA

Texts, Languages & Information Technology in Egyptology

Stéphane POLIS — Jean WINAND
With the collaboration of Todd GILLEN

Presses Universitaires de Liège

Texts, Languages & Information

Technology in Egyptology

Dépôt légal D/2012/12.839/17
ISBN 978-2-87562-016-3

Place du 20-Août, 7
B-4000 Liège (Belgique)
http://www.presses.ulg.ac.be

Imprimé en Belgique

Collection *Ægyptiaca Leodiensia* 9

Texts, Languages & Information Technology in Egyptology

Selected papers from the meeting of the Computer Working Group of the International Association of Egyptologists (Informatique & Égyptologie), Liège, 6-8 July 2010

Stéphane POLIS & Jean WINAND (eds.)

With the collaboration of Todd GILLEN

Presses Universitaires de Liège
2013

Table of Contents

The Digital Puzzle of the *Talatat* from Karnak

A Tool for the Three-Dimensional Reconstruction of Theban Buildings from the Reign of Amenhotep IV

Nathalie PRÉVÔT

Institut Ausonius (CNRS-Université de Bordeaux)

1. THE THEBAN *TALATAT*

The *talatat* from Karnak, or more precisely the usage of computers for the reconstruction of Atonist temples built with these *talatat* during the reign of Amenhotep IV-Akhenaton, constitutes a recurring theme in the context of the colloquia "Informatique et Égyptologie".[1] In the first session in 1984, Robert Vergnieux[2] gave a lecture on the database called *talatat*, and on his intentions and ambitions to put it online. It is now available on the Internet with protected access, and will be open to the public at the end of 2011.

The subject of *talatat* is far from being exhausted, and many aspects of its study still remain to be discussed: the conditions of their reconstruction, the restoration of their cohesion, as well as their relevance, significance, etc. Before presenting the tool of the interactive digital puzzle, which gives this article its title, it is necessary to contextualize it and so explain quickly the ATON-3D program which made its development necessary.

2. THE ATON-3D PROJECT

We are interested in a very methodological perspective of the *talatat* thanks to the revival of research on Atonist temples from Karnak, namely a vast interdisciplinary and international research project implemented in 2009 with the help of the French national agency for research (ANR-08-BLAN-0202-01). This project, called ATON-3D, aims to study the architectural policy of Akhenaton, both in Karnak and in Amarna, thanks to the tridimensional digital modelling of structures built during this reign. For Amarna, we have a lot of sources: numerous excavation reports and many reliefs from Amarna tombs are very informative, together with the tremendous advantage that Atonist temples have not been hidden by later constructions. On the other hand, most of the Amarna *talatat* disappeared, because they were in limestone and thus where burned in lime kilns during the Middle Ages.

In Karnak, there are tens of thousands of sandstone *talatat* which constitute the most important sources of documentation for the ATON-3D project. All the Atonist constructions were dismantled in the post-Amarna period, and the evidence was unfortunately scattered. We know the names of buildings, attested in some few texts found on preserved *talatat*, but we do not know their functions, plans, or locations.

1. Brocard 1994; Vergnieux 1985a, 1988, 1990, 1994.
2. Vergnieux 1985b: 223.

Figure 1. First results of the ATON-3D program:
The gem-pa-aten outside the boundaries of the Karnak precinct of Amon-Re (East; WIP)

Among the approximately 12 000 *talatat* which have been extracted from the western wing of the ninth pylon by the *Centre Franco-Égyptien d'Études des Temples de Karnak* (CFEETK), Robert Vergnieux has already systematically studied 6 666 stones of the last 24 courses in the base of the pylon.[3] However, it is necessary for us to review them today from the perspective of the three-dimensional reconstructions made by the ATON-3D project. Their importance is twofold. First, they constitute the walls of buildings which we want to restore, thus the more of it we assemble, the more we obtain dimensions, widths and heights of walls. Second, as the ornamental scheme of buildings erected with these *talatat* contains decorative scenes representing temples which were functioning during this period, the more of it we assemble, the more we obtain representations of the structures we want to restore.

3. THE *TALATAT* CORPUS

The very first database was created in the 80s on an Apple II. This database referenced 12 000 blocks which were distributed in about forty courses inside the ninth pylon. We were able to obtain the backup made on 5 ¼ inch floppy disks containing the corpus studied in Robert Vergnieux's thesis,[4] namely the lower layers of re-use (from the 24th to the 39th). We have recovered this database,[5] 25 years after the floppies were recorded, to ensure that this data be preserved and made accessible in the long term.

3. Vergnieux 1999.
4. Vergnieux 1999.
5. We had to find an Apple II and floppy disks with the programming language Pascal, because Apple II must be started under the system Pascal on floppy to be able to read the diskettes containing the old *talatat* database. The Apple II having naturally neither USB ports nor network connections had to be connected via a serial cable to a laptop still equipped with a COM1 serial port; the "modem" port of the Apple II (DIN 5 pins) and the COM1 port of the PC (V11 9 pins) were connected by a crossed cable ("null modem") made for that purpose. The application "Hyperterminal" was launched on the PC and configured to record on the hard disk all the traffic from the serial interface COM1. Finally, we used the file manager program in order to send the data to the "modem" port.

Today, with the beginning of the ATON-3D program, the *talatat* database has evolved a lot to allow integration into our information system ArchéoGRID[6] (fig. 2) and has been published online, which is a guarantee of preservation. This corpus is stored in a secure data centre[7] with a strong human and material infrastructure that will now ensure the conservation of images and metadata of the *talatat*.[8]

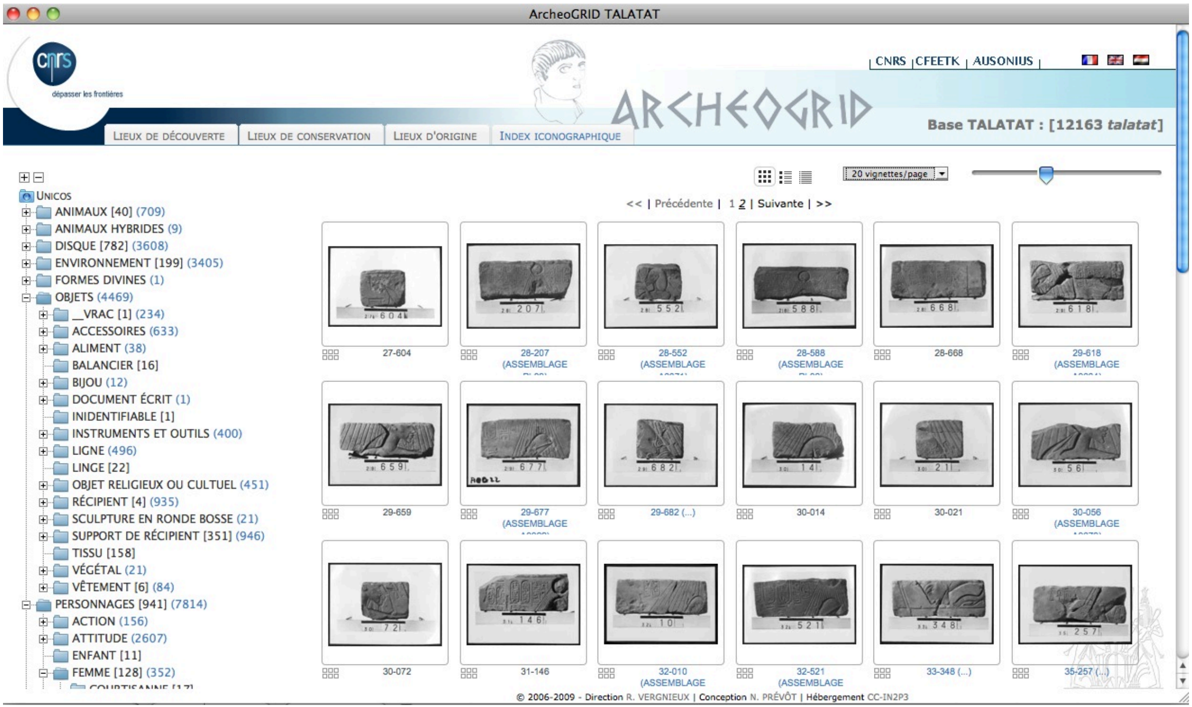

Figure 2. Archeogrid-Talatat

In partnership and in agreement with the CFEETK,[9] the *talatat* database is today available for all those involved in the ATON-3D project and should be publicly accessible in 2012.

The current database contains a little more than 12 000 documents. It is completely extensible, and we hope to add to it all other *talatat* found in Luxor, Ermant, Tod, Medamud, Hermopolis, etc. as well as establish links with other databases[10] also containing *talatat*.

For the documentation studies and the metadata entries, we use a French/English multilingual thesaurus. Each document is defined by descriptors of identification (inventory number, former

6. Archéogrid is the Information System developed by the Plateforme Technologique 3D in the Institut Ausonius, which allows the linking of the 3D reconstruction of a building or of an archaeological object to the heterogeneous information sources (photos, excavation reports, surveys, architectural drawings, historical iconography, etc.) that allowed its representation.
7. ADONIS-tge is a Major Facility (Très Grand Équipement) launched by the CNRS (French National Centre for Scientific Research) open to other partners. It aims to promote integrated access to digital data and documents in the Humanities and Social Sciences: “la grille Adonis”. It offers grid facilities (computing, storage, Web services, virtual environment, tools) and is organized within the Computing Centre of the National Institute for Nuclear Physics and Particle Physics (CC-IN2P3).
8. The engineers of the CC-IN2P3/TGE-Adonis perform with daily, weekly, monthly copies of security. Additionally, they proceed with implementation of warning devices of type technology and economic watch and with regular upgrades in order to ensure that we use software that are not only free and open source, but correspond technically to the state of the art. Moreover, they encourage us to respect systematically the standards.
9. http://www.cfeetk.cnrs.fr/index.php?page=axe-6.
10. For example, a few months ago, the ARCE presented to us a new database containing 16 000 *talatat* stored in the storehouse leaning against the temple of Khonsu in Karnak. Since 2009, all these blocks have been restored, photographed and documented. It is of the highest interest and importance to interconnect these two databases.

inventory numbers), of location (place of origin, place of discovery, place of conservation), of description (archaeological, iconographic, epigraphic), and is accompanied by bibliographical information.

To avoid the subjective nature of the iconographic indexing, we opted for an analytical description, incorporating the concept proposed by Robert Vergnieux of knowledge representation using *unicos*.[11] The aim here is to separate different types of information which we want to be able to query separately. A *unico* is the iconographic unity which corresponds to an independent iconic visual sign, e.g. human beings (royal family, priests, soldiers, courtiers, foreigners, artists, etc.), animals (cattle, horses, etc.), products of human activity (offering tables, thrones, ships, architecture, etc.), nature (sun-discs, canal and water, trees, flowers, etc.). Describing a *talatat* consists of enumerating the *unicos* present on its decorated face.

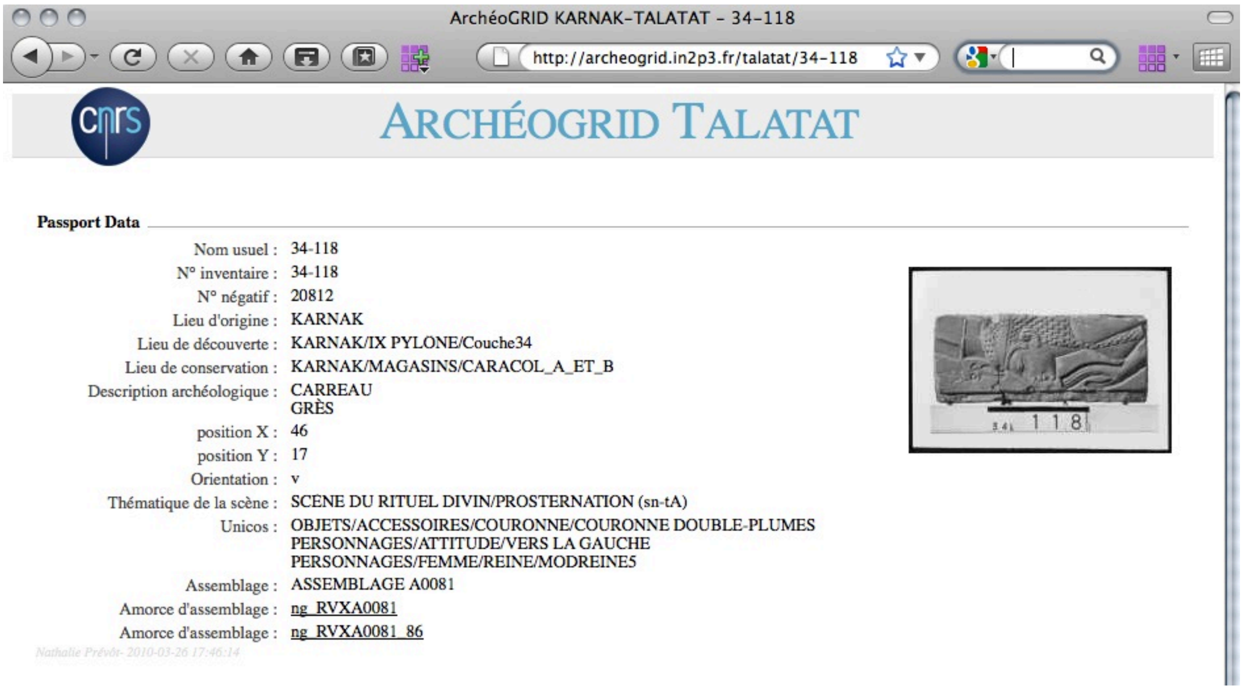

Figure 3. Metadata of a *talatat*

4. THE *TALATAT* PUZZLE

Talatat, if taken separately, are information poor. We have to reassemble them in order to get information that allows us to reconstruct the buildings.

We have developed a tool that facilitates not only the reconstruction process, but also the storage and archival of the data, allowing them to become in turn research documents that enrich the database. Such a tool has the further benefit of eliminating redundancy of effort — many scenes that have been reconstituted by draftsmen or PhD students working in Karnak using scissors and tracing paper have until now not been exploited for reconstruction purposes.

This collaborative assistant to the assemblage of blocks is interfaced with the corpus of *talatat*. It is a kind of puzzle on screen, which allows the correct placement of blocks thanks to a grid that takes into account the usual construction design of alternating rows of stretchers and rows of headers. Courses of headers are always aligned vertically in parallel. Each stretcher overlaps three headers and rows of stretchers are offset between them vertically by the length of half a header. Hence we ensure that a *talatat* matches laterally only with another one showing its decoration on the same surface as itself (e.g. headers next to headers) and above and below with a block decorated on a different side.

11. Vergnieux 1990.

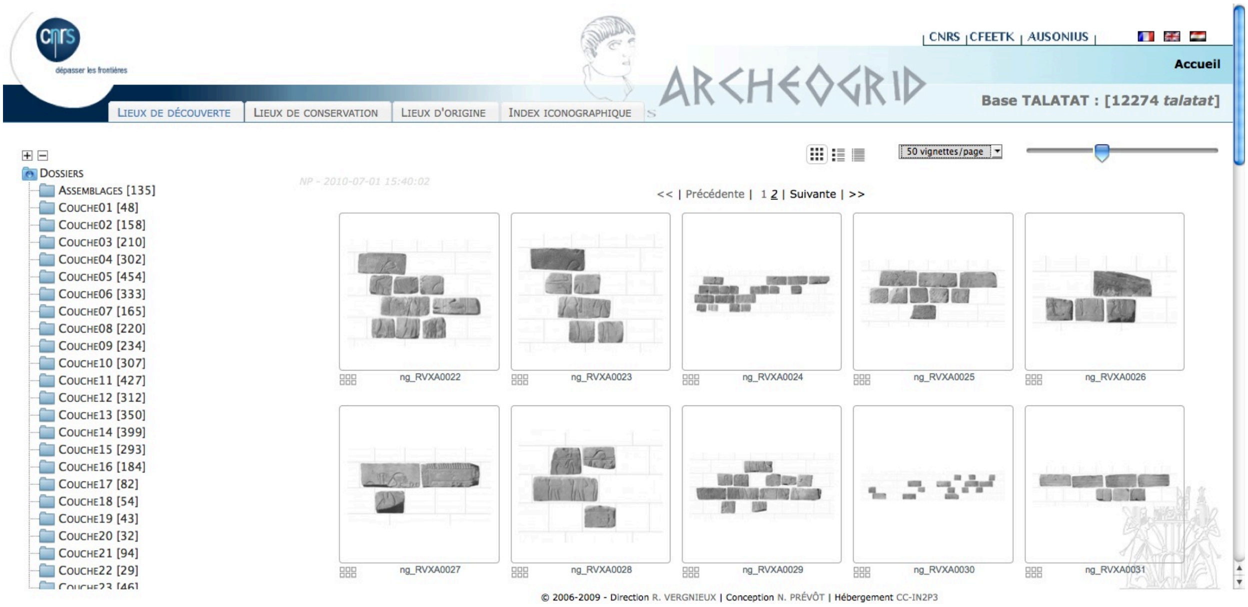

Figure 4. The assemblages in Archeogrid

On the left part of the screen, we search for potential neighbouring blocks using targeted queries (stretcher or header + the most likely *unicos*) and the results are displayed immediately; on the right part we test whether the block which seems to correspond matches with the others already in place. The photographs of the *talatat* displayed in the puzzle are cropped so that they can be set edge to edge. Of course, the images are represented at exactly the same scale.

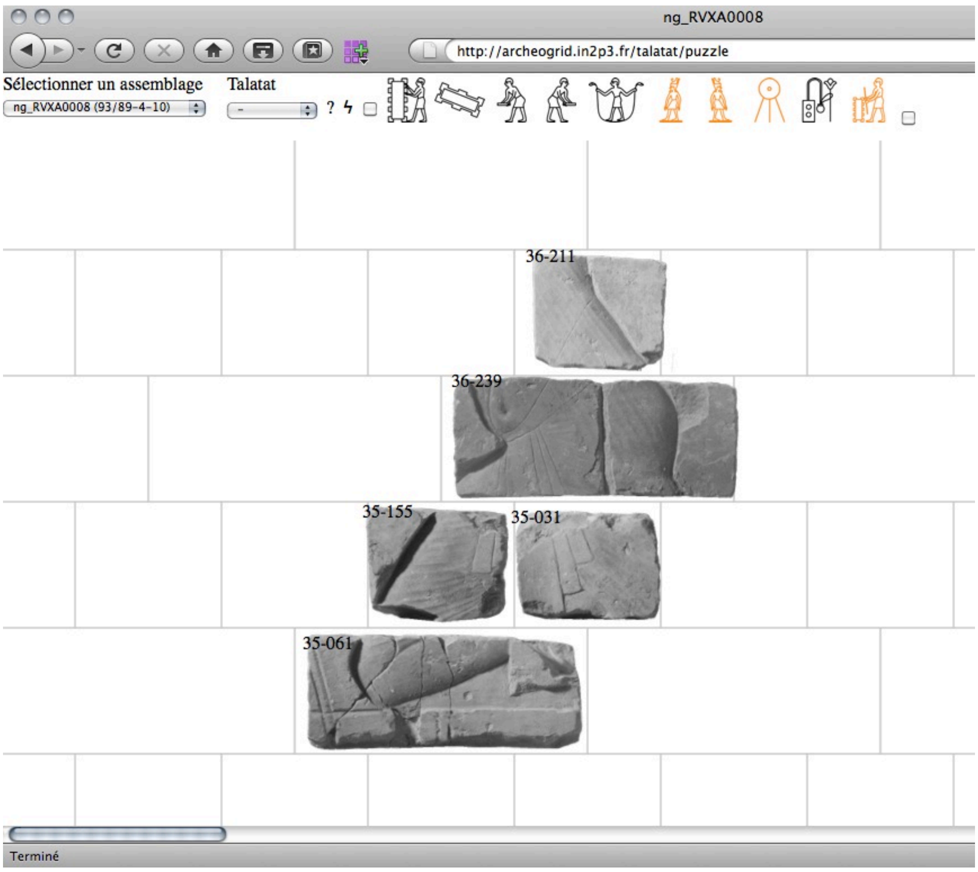

Figure 5. Assemblage A0008 with display of inventory numbers

Numerous options are available, such as zooming, rotating blocks, multiple selecting to move several *talatat* together, as well as the possibility of applying a layer of gridlines to follow the canon of proportions in the Atonist era (fig. 6). The images thus generated will be used as textures in the 3D models.

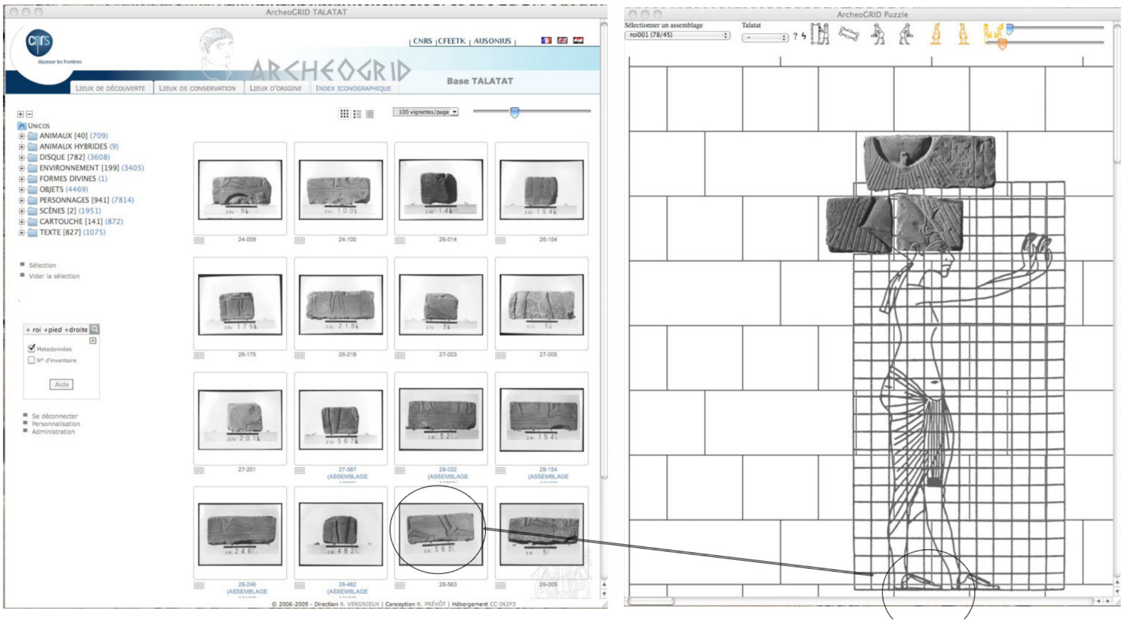

Figure 6. Using the puzzle tool

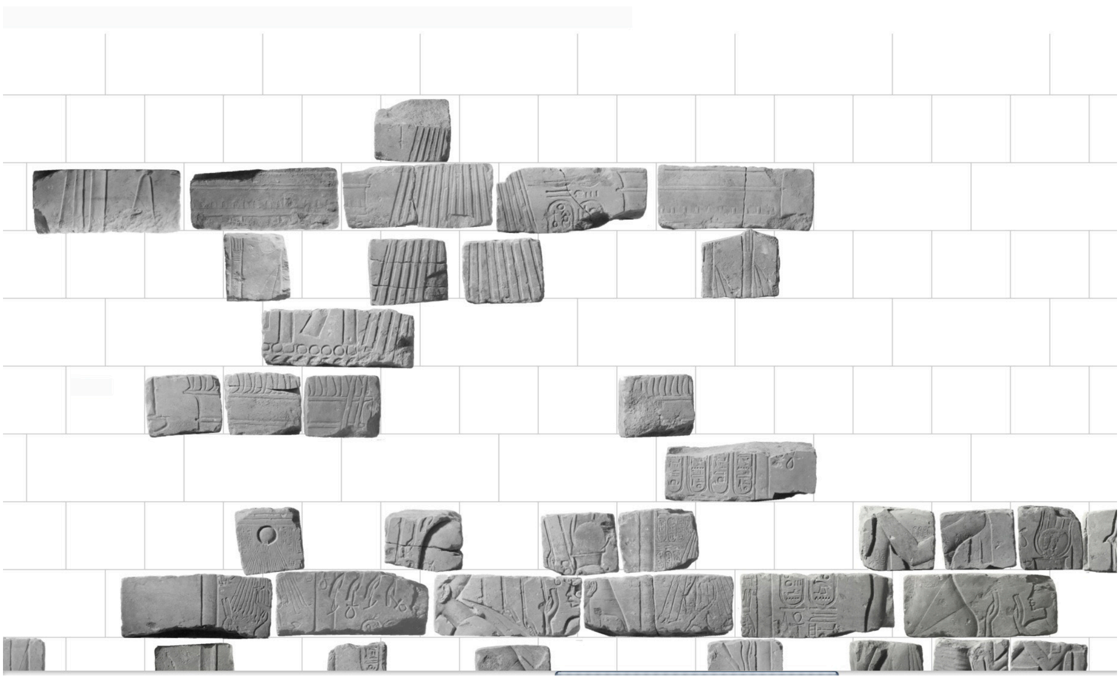

Figure 7. A portion of the assemblage A0011 published by R. Vergnieux, completed with the puzzle tool and with the help of D. Laboury[12]

The maximum height of the wall that it was possible to reconstruct with this tool is 15 cubits, or approximately 8 meters high, which corresponds to 35 courses of *talatat*. The maximum length is about 30 meters.

12. The reconstruction, still very incomplete, has 113 *talatat*; it is 11 meters long and 3,75 meters high — 17 courses in total.

The system backs up the temporarily reconstituted scenes by storing in the database the positions of every block on the grid. It is thus possible to resume a reconstruction several days after it was drafted. The current work can be even pursued by another researcher tempted by the addition of a likely block. So a number of users can participate in the collective construction of an interpretation.

This development was made possible thanks to the arrival of HTML5[13] and CSS3[14] formats, which allow the development of very ergonomic web applications and the performance of drag & drop operations and rotations of images, etc. This tool is certainly less advanced than Adobe Photoshop, but it does not require users either to download the photos of *talatat* on one's own machine, nor to install any software, providing users have a simple web browser installed.

The association between the puzzle and the database allows the immediate integration of new information in the system based on statistics obtained from the analysis of the number of reassembled stones, the number of missing stones, heights, lengths, number of courses of the reconstituted walls, etc.

Thus the database of *talatat* gets bigger and bigger according to the successful use of the puzzle combined with the subsequently obtained results. Archeogrid displays these data resulting from the research (the results of the reconstructions) together with the sources themselves (= the *talatat*).

5. INTEROPERABILITY OF THE *TALATAT* DATABASE: TOWARDS THE SEMANTIC WEB

In accord with the CFEETK, our corpus has been made machine-readable, with the double goal of making it accessible to researchers of the proto-Amarna period, and sharing it with other teams developing databases of *talatat* found in Karnak or elsewhere. This allows us to distribute resources from different research centers without gathering them physically or needing to duplicate them, which would quickly present difficulties for updating and archiving. In addition, it would be difficult to exploit resources effectively if they were stored in various locations, besides the complications associated with using a number of heterogeneous tools.

We chose to publish the data concerning the *talatat* presented on the Archeogrid website in an existing format so that they are easily reusable and immediately interoperable with data published with the same protocols. Therefore we use RDFa,[15] which allows the insertion of descriptions corresponding to the data model RDF[16] in the HTML representation of a resource. Consulting with a simple web browser the source code of the record of a *talatat* reveals a structured representation of the information according to the principles of RDF with the use of several documentary vocabularies:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML+RDFa 1.0//EN"
"http://www.w3.org/MarkUp/DTD/xhtml-rdfa-1.dtd">
<html xml:lang="fr" version="XHTML+RDFa 1.0"
xmlns="http://www.w3.org/1999/xhtml"
xmlns:foaf="http://xmlns.com/foaf/0.1/"
xmlns:dc="http://purl.org/dc/elements/1.1/"
```

13. This is the 5th major revision of the core language of the World Wide Web: the Hypertext Markup Language (HTML) for structuring and presenting content. In this version, new features are introduced to help Web application authors.

14. Cascading Style Sheets (CSS) Level 3 is a language for describing the rendering of a document written in HTML or XML.

15. RDFa, which means "RDF in attributes", is a meeting between a document format that represents the resources and a data model used by machines that describes these resources. It thus allows users to make queries on a text document with the standardized query language SPARQL just as they are done in a database with SQL.

16. As its name suggests, the RDF (Resource Description Framework) allows the description of resources (while the HTML allows the construction of representations of these resources); see Mader *et al.* (current volume) for another application of this technology within the field of Egyptology. Developed by the W3C within the framework of the activities of semantic Web, RDF is not, strictly speaking, a metadata schema. It is a model of description of the structured data inspired by graph theory. Its genericity and its flexibility offer a framework of interoperability for describing all types of resources in a networked environment like the Web.

```
xmlns:cc="http://creativecommons.org/ns#"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#
xmlns:tal="http://archeogrid.in2p3.fr/talatat/ns/1.0/"
xmlns:xsd="http://www.w3.org/2001/XMLSchema#">
```

The advantage of RDF is that it is possible to exploit the data, whatever the vocabulary used, without having to convert it, unlike XML for which it is necessary to convert the data if the user is not using the same schema. Thus it does not place the imposition on various teams to agree on a single method of structuring metadata, nor does it limit them to a lesser common denominator to ensure interoperability.[17] RDF provides a framework for great flexibility to mix and associate terms from existing vocabularies, as well as invent our own in the combination which is best suited for our particular contents. The interoperability offered means on one hand that our data are exploitable by existing tools, and on the other hand that they can really be connected with other data via the web.

To indicate keywords, we insert "dc:subject" tags that refer to Dublin Core[18] vocabulary and that we link to our thesaurus:

```
<span rel="dc:subject">
<span property="tal:unico">
<a href="[lien vers les unicos]">
   OBJETS/ACCESSOIRES/COURONNE/COURONNE DOUBLE-PLUMES</a>
</span>
```

For the geo-localization of the *talatat*, we use Dublin Core terms with the tag "spatial", which allows us to specify values of latitude and longitude while referencing in the WGS vocabulary validated by the W3C (geo:lat and geo:long).

```
<span rel="dcterms:spatial">
<span property="geo:lat" content="25.716269"></span>
<span property="geo:long" content="32.655174"></span>
</span>
```

The thesaurus used has been entirely converted to the language SKOS,[19] which, like an ontology, aims at an increased interoperability for exchanges of lexicon, the integration of a semantic Web as well as automatic machine processing.

The normalized syntax we use can be exploited in various ways to obtain information from our resources. To do this, it is simply necessary to retrieve the sitemap.xml file[20] at the site root (fig. 8) and to extract the data correctly marked with RDFa using appropriate parsers, whether written in XSLT,

17. As it is the case with OAI-PMH.
18. The Dublin Core Metadata Element Set is a vocabulary of fifteen properties for use in resource description. Its elements are broad and generic, usable for describing a wide range of resources.
19. Simple Knowledge Organization System – The language SKOS designed for representation of thesauri, classification schemes and taxonomies is an ontology allowing on one hand the representation in a multilingual context of every type of structured and controlled vocabulary, and on the other hand the alignment of various vocabularies, the objective being the machine exploitation of resources published on Web.
20. Introduced by Google, the Sitemaps protocol allows users to indicate to search engines the resources of a website to be indexed. It is an XML file which contains (for every resource): its URL, its date of last modification, the frequency of revision and the relative importance with regard to the other URLs of the site. The use of the Sitemaps protocol allows us to guide the collection of the data and metadata on *talatat* and assemblages.

JavaScript or even Python. With these tools, which already exist,[21] RDF tuples can be recreated, and exploited like any other data in this format, and can then be shown on any other website.

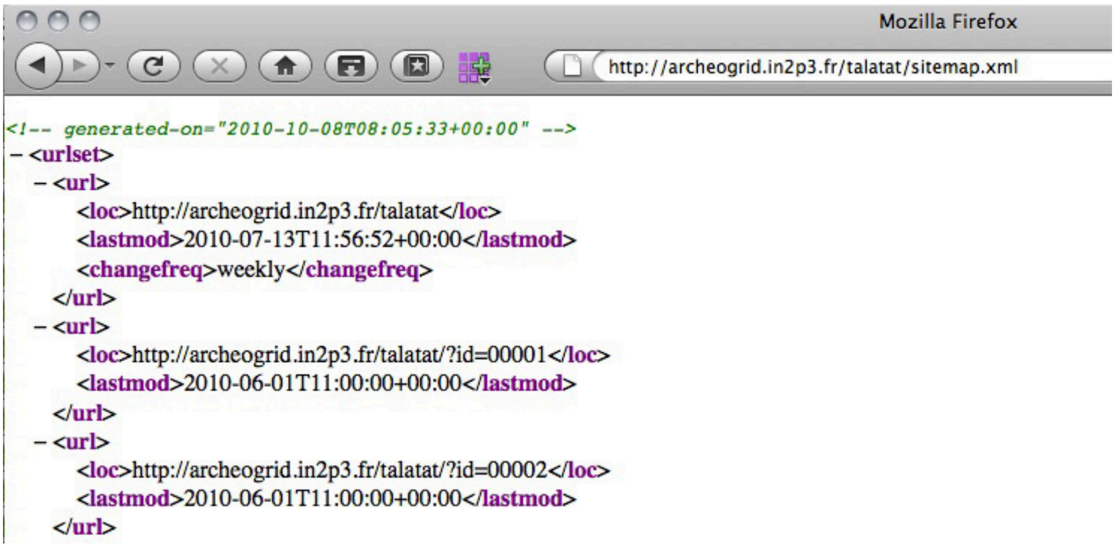

Figure 8. Sitemap.xml

In summary, all the data on *talatat* as well as the research data derived from them, such as the walls rebuilt using the puzzle, are usable and quickly interoperable, without being duplicated or moved, and always in their latest version, while they are continuously documented and interpreted by successive refinements.

BIBLIOGRAPHY

ADIDA, Ben, Mark BIRBECK, Shane MCCARRON & Steven PEMBERTON. 2008. RDFa in XHTML: Syntax and Processing – W3C Recommendation 14 October 2008, http://www.w3.org/TR/rdfa-syntax.

BILLET, Sophie. 1988. Intelligence artificielle et base de données pour lassemblage des talatat, in: *Fifth International Congress of Egyptology, October 29 — Novembrer 3, Cairo, 1988, Abstracts of Papers*, International Association of Egyptologists, Cairo, p. 23-24.

BILLET, Sophie, Michel GONDRAN & Robert VERGNIEUX. 1988. Intelligence artificielle et 'talatat', in: *Fifth International Congress of Egyptology, October 29 — Novembrer 3, Cairo, 1988, Abstracts of Papers*, International Association of Egyptologists, Cairo, p. 25-26.

BROCART, Yannick. 1996. Quelques exemples de restitutions virtuelles, in: *Informatique et Égyptologie* 10, p. 31-34.

BRICKLEY, Dan & R.V. GUHA. 2004. RDF Vocabulary Description Language 1.0: RDF Schema - W3C Recommendation 10 February 2004, http://www.w3.org/TR/rdf-schema.

MILES, Alistair & Sean BECHHOFER. 2009. SKOS Simple Knowledge Organization System Reference - W3C Proposed Recommendation 15 June 2009, http://www.w3.org/TR/skos-reference.

PRUD'HOMMEAUX, Eric & Andy SEABORNE. 2008. SPARQL Query Language for RDF - W3C Recommendation 15 January 2008. http://www.w3.org/TR/rdf-sparql-query/.

VERGNIEUX, Robert. 1985a. Karnak, Mission Permanente du CNRS. Archives documentaires, in: *Informatique et Égyptologie* 1, p. 221-223.

—. 1985b. Le fichier 'Talatat' en cours de constitution à Karnak, in: Sylvia SCHOSKE (ed.), *Akten des vierten Internationalen Ägyptologen Kongresses, München 1985* (= BSAK 1), p. 205-211.

—. 1988. Premiers exemples de résultats obtenus à l'aide du fichier informatisé sur les 'talatat' et vocabulaire de recherche, in: *Microcomputer in der Archäologie* (= WZB 37), p. 283-287.

—. 1990. La représentation des connaissances à l'aide d'unicos, in: *Informatique et Égyptologie* 7, p. 157-160.

21. See e.g. http://www.w3.org/2007/08/pyRdfa/extract?uri=[URI to be parsed].

—. 1994. Quelques aspects d'une recherche informatisée sur un lot de talatat d'Aménophis IV, in: *Informatique et Égyptologie* 9, p. 152-155.

—. 1996. Le fac-simile électronique ou les restitutions virtuelles, in: *Informatique et Égyptologie* 10, p. 134-139.

—. 1999. *Recherches sur les monuments thébains d'Amenhotep IV à l'aide d'outils informatiques*, Genève (= Cahiers de la Société d'égyptologie 4).

VERGNIEUX, Robert & Michel GONDRAN. 1997. *Aménophis IV et les pierres du soleil : Akhénaton retrouvé*, Paris.

Abstracts

Peter DILS & Frank FEDER, The *Thesaurus Linguae Aegyptiae.* Review and Perspectives

The *Thesaurus Linguae Aegyptiae* (*TLA*) represents today the largest available database of Egyptian texts and, moreover, it is worldwide accessible on the Internet with free access. It combines a text corpus of Egyptian texts from nearly all periods of Egyptian history with an electronic lexicon. Both are linked to each other and are regularly updated. The TLA provides also access to the digitalized material on which the edition of the *Wörterbuch der aegyptischen Sprache* was based (slip archive). The text corpus and the lexicon can be searched in a number of ways and for different purposes; tools for statistical analysis are provided as well. As the *TLA* is a dynamically developing database system the text corpus and the lexicon will further be expanded, especially by adding the still lacking Coptic material of the Egyptian language, and by improving the research tools gradually.

Stéphane POLIS, Anne-Claude HONNAY & Jean WINAND, Building an Annotated Corpus of Late Egyptian. The Ramses Project: Review and Perspectives

This paper reviews the experience of the Ramses Project in constructing a richly annotated corpus of Late Egyptian that consists of 300 000 words in 2011 (and is expected to grow up to more than 1 million words in coming years). During the first five years of the project, this corpus has been encoded in hieroglyphic script, translated in French or English and received annotations for part-of-speech information, lemmatization, and morphological analysis. The methodology and working tools that have been developed in order to build this corpus are here discussed and future developments are presented.

Stéphane POLIS & Serge ROSMORDUC, Building a Construction-Based Treebank of Late Egyptian. The Syntactic Layer in Ramses

This paper reports on the construction-based Treebank currently under development in the framework of the Ramses Project, which aims at building a multifaceted annotated corpus of Late Egyptian texts. We describe the specifications that have been implemented and we introduce the syntactic formalism and the related representation format that are used for the syntactic annotation. Furthermore, the annotation scheme is discussed with particular attention paid to its evolutionary nature. Finally, we explain the methods as well as the annotating tool, called *SyntaxEditor*; we conclude by addressing the question of forthcoming developments, especially the search engine and a context-sensitive parser.

Stéphanie GOHY, Benjamin MARTIN LEON & Stéphane POLIS, Automated Text Categorization in a Dead Language. The Detection of Genres in Late Egyptian

This paper is a first step in applying machine learning methods typical of Automated Text Categorization (ATC) for Automatic Genre Identification (AGI) in Late Egyptian, a language written in either hieroglyphic or hieratic scripts that is found in documents from Ancient Egypt dating from ca. 1350-700 BCE. The study is divided into three parts. After a general introduction on AGI (§1), we introduce the levels of annotation that are integrated in the Ramses corpus and can be used when performing AGI on Late Egyptian (§2). In the following section (§3) we offer a brief survey of the types of features that have been discussed in the literature on AGI, before proceeding with three case studies where we apply supervised machine learning methods — namely the naïve Bayes classifier (§4.1), the Support Vector Machine (§4.2), and the Segment and Combine approach (§4.3) — to a selection of texts in the corpus. Their respective performances are tested using lexical, part-of-speech and inflectional features.

Mark-Jan NEDERHOF, Flexible Use of Text Annotations and Distance Learning

In this paper, we discuss a framework that allows independently created annotations of texts to be combined and presented as one unified interlinear format. Applications for distance learning are also considered. As proof-of-concept, we present PhilologEg, a tool that can be used to study an Ancient Egyptian hieroglyphic text in combination with any number of translations and grammatical annotations. The tool is a fully integrated system that runs on all major platforms.

Roberto GOZZOLI, Hieroglyphic Text Processors, Manuel de Codage, Unicode, and Lexicography

This paper gives an overview of the different software available to scholars working in the field of Egyptian language, with a special focus on hieroglyphic typesetting, Unicode and lexicographical databases that systematically encodes hieroglyphs. Various problems with the *Manuel de Codage* are discussed, as well as the need for a more active interaction between computers and Egyptology. A proposal for Egyptological software is given at the end of the paper.

Mark-Jan NEDERHOF, The Manuel de Codage Encoding of Hieroglyphs Impedes Development of Corpora

In this paper, we discuss the encoding of hieroglyphic text and argue that the set of requirements for an encoding scheme depend on the intended application. Our main claim is that if this application is the development of text corpora with long lifespans and diversity of use, then encoding schemes within the tradition of the Manuel de Codage are unsuitable.

Vincent EUVERTE & Christian ROY, Hieroglyphic Text Corpus. Towards Standardization

Sharing the heritage of Ancient Egyptian written production means facing numerous technical challenges. The goal of this paper is to build a preliminary inventory of these challenges and to propose some possible solutions. After a quick overview of the topics that are possible candidate to an international standardization, the paper focuses on two aspects. (1) The ‘Multilingual Egyptological Thesaurus’ (MET), initiated in 1996 by Dirk van der Plas, has not changed since 2003. It could be updated and expanded with minimal effort under the coordination of an official body such as the Center for Documentation of Cultural and Natural Heritage (CULTNAT). (2) The ‘Manuel de Codage’ (MdC) has not benefited from developments in computer science since the third edition was published under the *Informatique & Égyptologie* mandate in 1988. Over time, each hieroglyphic software program has developed its own specific syntax to satisfy emerging needs, making it difficult for users to share ancient Egyptian texts. For these two topics, we will suggest a plan for improvement based on the Rosette Project's experience, though the input of the Egyptologists' community at large is appreciated to refine various concepts and identify the best route forward.

Christian MADER, Bernhard HASLHOFER & Niko POPITSCH, The MEKETREpository. A Collaborative Web Database for Middle Kingdom Scene Descriptions

Whilst representations, iconography and the development of scenes in private and royal tombs from the Old Kingdom have been studied extensively in the past, comparable research of Middle Kingdom (MK) representations and scene details is still underrepresented. The MEKETRE research project aims at closing this gap by systematic research of MK representations. In the course of this project, an online digital repository (the MEKETREpository) is being built that enables researchers to describe and annotate MK two-dimensional art at various levels of detail using images, free text, and controlled vocabularies. It also enables the collaborative development of semantic vocabularies for the descript-tion of these data. The MEKETREpository will publish the resulting data and vocabularies as Linked Data on the Web by utilizing Semantic Web technologies to enable their integration into other Linked Data sets such as DBpedia, Freebase or LIBRIS. The collected data is described using standardized and specialized vocabularies allowing for easy integration into existing databases and search engines. For the long-term preservation of the data, the MEKETREpository will make use of the University of Vienna's digital asset management system PHAIDRA. At its final stage the MEKETREpository will supply a platform that exposes collaboratively created, continuously evolving, and publicly available information about the MK on the Web.

Nathalie PRÉVÔT, The Digital Puzzle of the *talatat* from Karnak. A Tool for the Three-Dimensional Reconstruction of Theban Buildings from the Reign of Amenhotep IV

The revival of studies on the Atonist temples of Karnak (program of the French National Research Agency ATON-3D – ANR-08-BLAN-0202-01) required the implementation of an Information System dedicated to the Theban *talatat* that would also be accessible to the scientific community. This IS is associated with software which helps to reassemble the fragmented reliefs (a digital interactive puzzle), constituting a real tool for researchers and providing the knowledge needed to produce and validate hypotheses about the structures and dimensions of the buildings. The database is then enriched with images of the temple's extrapolated decoration, which involves 3D modelling of these extrapolations. *Talatat* indexing was based on the Multilingual Egyptian Thesaurus conventions regarding "passport" data, including iconographic description using descriptive operators called *unicos*. In the spirit of the international movement in favour of open access to scientific data, the *talatat* metadata and images are accessible online to researchers working on the proto-Amarna or Amarna periods. The *talatat* metadata is published using RDFa data model mapping for embedding RDF triples within the XHTML of our web pages, which can be extracted by compliant user agents. This corpus is stored in a secured warehouse with strong human and digital infrastructure for preservation of the images and of their metadata.

Carlos GRACIA ZAMACONA, A Database for the Coffin Texts

This article describes a database for the Coffin Texts. It was first conceived as a semantic study of verbs of motion, and for this reason many of its files are linguistically focused. Nevertheless, it may be useful for other kinds of studies, because the software employed allows integration of new files as well as modification of old ones. This is the ultimate aim of such a database: a tool appropriate for all kinds of research on this corpus. Specific features of this corpus are discussed first, followed by the database conception and structure, and finally its use, results and developments.

Azza EZZAT, The Digital Library of Inscriptions and Calligraphies

The Digital Library of Inscriptions aims at recording all inscriptions on ancient Egyptian buildings and monuments throughout the ages. These inscriptions are digitally displayed for the user, including a brief description and pictures of the inscriptions. The languages included in the Digital Library are Ancient Egyptian, Arabic, Turkish, Persian and Greek languages. Moreover, there are inscriptions bearing Thamodic, Musnad, and Nabatean scripts.

Yannis GOURDON, The *AGÉA* Database Project.
Anthroponymes et Généalogies de l'Égypte Ancienne

Since the 30s, our understanding of the ancient Egyptian personal names has been dependent on Ranke's *Personennamen*. But, because the data and its philological and sociological analysis are based on the knowledge available in the first half of the 20th century, the *PN* requires a complete revision that takes into account recent developments on the subject. Launched in 2008 at the IFAO, the *AGÉA* database project aims, eventually, to create a systematic directory of personal names for every period of the Pharaonic history, completing and modernizing Ranke's work. As a tool facilitating more efficient analysis and a better interpretation of data, *AGÉA* will focus, in its first development, on the Old Kingdom.