Linking digital tools for the Ancient World through Trismegistas
stable identifiers

Trismegistas [TM] (www.trismegistos.org) started in 2005 as a platform to facilitate access to information about published papyrological texts in all possible languages and scripts from Graeco-Roman Egypt, in a first instance Greek, Latin and Egyptian (including hieroglyphic, hieratic, and demotic). The inclusion of Egyptian soon dissolved the disciplinary boundary with epigraphy, broadened the chronological window, which was eventually set to 800 BC – AD 800, and led to the inclusion of further languages such as Coptic, Aramaic and Arabic. In 2010 the idea grew to expand the geographical scope to include the Ancient World in general, starting off with texts in ‘smaller’ indigenous languages such as Etruscan, Italic, Messapian, Lepontic or even Runic, and by 2012 the decision was taken to include so-called ‘unpublished’ texts for which information is available in online repositories. This means that TM increasingly wants to be a platform pointing to places where information can be found about all texts from antiquity, thus facilitating cross-cultural and cross-linguistic research. This will of course only be possible through cooperation with all players in the field, since our aim is to lead people to the partner websites, where more information, often including photographs, transliterations and translations of the texts, can be found.

Disciplinary boundaries as well as the distinction between published and unpublished are increasingly less productive in the holistic approach stimulated in and facilitated by a digital environment. Nevertheless current scholarship is still rife with this fragmentation of knowledge. Text publications often remain grouped per discipline, with different volumes or series for papyri, ostraca, or inscriptions; literary texts are treated separately from documentary texts; and in the past bilingual texts have been split up according to language. On the other hand the practice of text publication has separated the physical objects from the ‘dematerialized’ text, which has grown to lead a life of its own. In this way the holders of the documents, such as museums and libraries, have in many cases even become alienated from the knowledge produced by scholarship, while the scholars often have difficulties locating the objects on which the texts were written. This disjunction between the various scholarly disciplines and the institutional owners is also reflected in the various conventions for identifying texts: sometimes one of the publications is referred to, sometimes the inventory number is used, both of which are in an informal setting often replaced by a name if the document and text are well-known. Thus the trilingual inscription crucial for the decipherment of hieroglyphs can be referred to as OGIS 1 90 (a publication) or as British Museum, EA 24 (its inventory number), but it is generally known as the ‘Rosetta Stone’.
Because all of these systems have drawbacks, certainly in a digital environment, TM uses serial numbers to identify texts. The Rosetta Stone, for example, is available under TM 8809. This number is in itself arbitrary and meaningless, but www.trismegistos.org/text/8809 directs users to the relevant metadata: publications (for Greek, Demotic and hieroglyphs), the museum inventory number, associated date(s) and places, the writing surface, the textual genre, etc. By not discriminating between scholarly disciplines nor between published and unpublished documents, and by gradually incorporating all into one standardized database, the TM text serial number aims to facilitate communication between the different fields of research as well as between the scholarly world and the universe of museums and libraries. All this should improve the accessibility and the quality of information about historical sources, both to scholars and eventually also the general public.

An illustration of what becomes possible in a digital world where texts are unambiguously identified through TM text serial numbers are our recent advances in dealing with places and people. Combined with the availability of the full text in open access repositories, and the development of Named Entity Recognition [NER], TM has made significant progress in creating gazetteers of anthroponyms and toponyms that occur in the ancient sources. Starting out with Greek papyri, where TM could build on the Prosopographia Ptolemaica (a Who’s Who of Ptolemaic Egypt), we tackled the full text of some 50,000 documents found in the Duke Databank of Documentary Papyri [DDbDP], and distilled over 350,000 Greek (and Latin) personal names from the texts from Egypt dated before AD 500.

In 2014 then, TM joined the SNAP-project (snapdrgn.net), which focuses on how to link together large databases of persons and names, to make our own dataset more accessible in new standards such as RDF and to explore ways of exchanging information through Linked Open Data. Also as a part of this project, and benefiting from its disambiguation role in the current EAGLE project (www.eagle-network.eu), TM is currently applying NER to the almost half a million Latin inscriptions (outside Egypt), with very promising first results. Similar pilot projects have been set up to filter out (ancient) abbreviations in the text or editorial corrections (mainly for grammatical irregularities).

An open access digital world with stable TM and other identifiers offers many exciting possibilities: we have discussed above how communication and cooperation between various digital projects, both from the scholarly world and from museums, will be easier; on the basis of the TM gazetteers the underlying identifiers could be added in PDFs of scholarly articles and books, which could even lead to automated bibliographies; a collaboration could be set up with Wikipedia, to make specialized knowledge more accessible. Further down the line interesting things should be
possible with Social Network Analysis [SNA]. In the two years that we have been experimenting with this tool, it has proven valuable both for data visualization and new insights in the structure of the evidence. It can be used as a disambiguation tool, for identifying people across texts, but also for the reconstruction and analysis of actual (social) networks, for example to detect groups in the witnesses appearing in contracts. Our blog Six Degrees of Spaghetti Monsters (spaghetti-os.blogspot.be) aims to be a guide for those wanting to tackle SNA for the first time.