Datenbanken – Abstracts

Yanne Broux, Mark Depauw, Herbert Verreth: Trismegistos goes global: inscriptions, places and people

Trismegistos has long been a database project focusing on Egypt, 800 BC – AD 800. Its origins lie in the papyrological world rather than in epigraphy. Yet since a couple of years both the disciplinary focus and the geographical limitations have been set aside. This paper hopes to demonstrate that Trismegistos has now become a useful tool for epigraphers as well.
At the core of Trismegistos lies the texts database. This now counts nearly 700,000 entries, and over 500,000 of these are epigraphic. Almost all of the Latin inscriptions have been included through a cooperation with the EAGLE project, and we are actively exploring an expansion to Greek epigraphy. Our final goal is to map all textual evidence surviving from the Ancient World as a whole, as exhaustively as possible. Therefore Trismegistos (already) includes texts in the local vernacular, e.g. in the Italic languages, Etruscan, Iberian, or Demotic.
At the same time we are expanding our geographical and onomastic databases to include the evidence from Latin epigraphy. Named Entity Recognition has assisted us to filter out proper names from the Latin inscriptions. As a result Trismegistos Places added over 80,000 toponyms found in the nearly half a million Latin inscriptions. The personal names in these texts amount to almost 500,000 name clusters. Together with those from Greek inscriptions, they will be the subject of a separate project called APEX (for Ancient Profiles EXploited). With the help of graph visualization and network analysis, co-occurrence of individuals will be tagged to lay the foundations for a true 'Facebook of the ancient world'. On the basis of the database, we plan a large-scale investigation of the changes in personal identification and of the construction of identities through local and Graeco-Latin names in the Roman Empire.
Named Entity Recognition has also enabled us to create a database of abbreviations in Latin inscriptions. This is turn has laid the foundations for an automated retrieval of formulaic language. Both are available on Trismegistos Abbreviations (& Formulae).

Ulrich Huttner, Julien M. Ogereau: The *Inscriptiones Christianae Graecae* Database: Towards a Digital Corpus of Early Christian Inscriptions from Greece and Asia Minor
This paper will review the latest development of the *Inscriptiones Christianae Graecae*, a relational database of early Christian inscriptions from Greece and Asia Minor (http://www.epigraph.topoi.org), which is currently being developed by the Forschungsgruppe TOPOI B-5-3 under the direction of Prof. Drs. K. Hallof and C. Breytenbach at Humboldt-Universität zu Berlin (http://www.topoi.org/project/b-5-3/).
In the first part of this presentation, authors will introduce the new interface of the database, its improved search functionalities, geo-mapping tools, as well as its open-access, citable, published version in Edition Topoi collections (forthcoming at: http://repository.editiontopoi.org/). They will also offer critical reflections on the process (and challenges) of updating the old database software, of integrating old data content into a new database structure, and of migrating data into a citable, scholarly, publication platform. In the second part, they will provide a brief overview of the epigraphic data from western Asia Minor (Caria, Lycia, Mysia, etc.), central and northern Greece (Attica, Corinthia, Macedonia), as well as Thracia, which have been recently entered into the database. The paper will conclude by highlighting the significance of the material for our understanding of Early Christianity, and, more generally, for our understanding of the interaction between indigenous epigraphic cultures and Graeco-Roman culture and the transformation of the eastern provinces of the Roman Empire in late antiquity.

Chiara Lasagni: "The Epigraphic Landscape of Athens": Preliminary Results for a Semantics of Athenian Topography

The Epigraphic Landscape of Athens, entirely founded by the Italian Ministry of University and Research (*SIR*: *Scientific independence of young researchers* 2014, duration: 2016-2018), is a project lead by a team the University of Turin (Department of Historical Studies). The research group is composed by Chiara Lasagni (coordinator), Enrica Culasso Gastaldi, Daniela Marchiandi, and Francesca Rocca.

The main aim of the project is to perform a mapping of the places of discovery (*fundorte*) and of original display (*standorte*) of the Athenian public inscriptions. With "public inscriptions" we mean all those epigraphic documents that can be understood as an expression of the political and institutional life of Athens, i.e. decrees, laws, interstate agreements, sacred laws, financial accounts, inventories, honorific dedications, etc. The display of such inscriptions in certain places of the ancient city establishes an ongoing dialogue with its various monumental areas, and constitutes an essential element of the "epigraphic communication" between the Athenian *polis* and its citizens. In fact, the arrangement of the epigraphic monuments completes, and sometimes enriches, the very message of the inscribed text. Public epigraphy intertwines aspects both of technical and practical order (i.e. the epigraphic monument safeguards the stability of the related text and ensures its publicity), and of ideal order (the same artefact substantiates and celebrates the very functioning of the *polis* as a community). For this reason, a wide-ranging analysis on the *standorte* of the Athenian public inscriptions (that has never been undertook so far) is able to provide us new and unbiased elements on these aspects in their development over time.

Moreover a parallel mapping of the inscriptions' findspots is performed with a twofold aim: first, to tidy up a tangle of data, often scattered and controversial (therefore constituting a useful tool for researchers interested in Attic epigraphy); second, to identify groups of epigraphic evidences clustered in specific points of the urban area, and to relate them. By crossing the information about the places of discovery and that on the original positions of the stones, the mapping will highlight the pathways followed by the reused inscriptions, thus identify the monuments or the areas of the asty where they more likely come from.

We have opted for the panel "Databases" since *The E.L.A.* project has at its core the creation of an on-line collection of data on the topography of the Athenian public inscriptions.

In our lecture, we are going to present the general contents of this project (which, at the time of the 15° CIEGL will be still underway), and the main features of the related database*, also focusing on a number of case studies, dealing with the theme of the "semantics of the Athenian topography". Our major aim is to demonstrate the heuristic potentialities of this mapping for the study of the epigraphic communication of the Athenian state, showing the E.L.A. database "at work" in performing some concrete case studies.

Silvia Orlandi, Pietro Liuzzo: From EAGLE to Koinè

This paper will address the next steps in Digital Epigraphy. The EAGLE project has carried out an enormous amount of work in order to harmonize, aggregate, and facilitate use of data in online epigraphic archives, fostering and supporting numerous databases and projects. For the first time it is possible to access records in different databases with one search and find in one place both greek and latin inscriptions. Nevertheless the three years of the EAGLE project benchmarked the need for a much longer effort and for a deeper change in the long standing online database traditions in epigraphy. An aggregator can only serve the temporary function of displaying the amount of data collected, but its update and maintenance fall into the pitfalls of always having to deal with legacy databases. So what will be the next step? Could we rather than exporting to a common aggregator, export from it? The advantages of a common source data entry and editing point, visualized in different places and by different services is self evident and would offer to the users a much more reliable service not only from one access

point but from all the access point serving that data. Three years ago there was no common standard for data encoding, but there is now and this could be taken forward to prepare a different architecture and coordination of epigraphic projects capable also of giving researchers a view of what would need more work and effort, for example by displaying that 5 digital edition of one inscription exist while none is available for another. The many existing Linked Open Data tools for annotation could be integrated into such service to foster easy editing, annotation and enrichment of the common data, benefiting especially the numerous small scale projects, occasional users, contributors but especially end users.

Neel Smith: Morphological parsing of Greek inscriptions in multiple dialects and alphabets

Epigraphists are interested in questions about societies "which have several languages and scripts existing simultaneously in their epigraphic culture," to quote the central theme of this conference, but in working with digital texts are often hampered by inadequate technical standards. ISO 639-3 (language codes) can distinguish only between "Mycenean Greek" and "Ancient Greek before 1453"; ISO 15924 (codes for scripts) identifies a single "Greek" script. One consequence is that computational systems analyzing literary Greek texts cannot be applied to epigraphic texts in local dialects, or in epichoric scripts, even if the dialect is familiar (e.g., texts in classical Attic dialect but written in the Attic alphabet). The Kanōnes system for building parsers addresses these issues directly. It cleanly isolates the definition of writing systems so that the same engine can be used to parse texts in different epichoric alphabets. It focuses explicitly on tailoring sets of morphological stems and inflectional rules for specific corpora, and allows epigraphists to define these data sets without programming, by editing simple tables.

In keeping with its corpus-linguistic orientation, Kanōnes identifies lexical entities, forms, stems and inflectional patterns with canonical identifiers. Analyzing a corpus, of whatever size, therefore yields machine-actionable data not only about what vocabulary appears in the corpus, but also about what specific forms of each lexical entity are attested, and what grammatical rules underlie the identification of each form. This is valuable information per se, and provides rigorously structured data for higher-order analyses (e.g., of syntactic or rhetorical structures).

This paper will introduce the Kanōnes system, and illustrate how it can be applied to epigraphic texts in various alphabets and dialects.

1. The technical design of the Kanōnes system is presented in "Approaches to morphological analysis of historical languages"
(*BICS*, 59[2], December 2016). The source code is freely available on github: see
http://neelsmith.github.io/greeklang/morphology/.

Ernest Suyver: SEG Forward. Further development of SEG as a digital publication

At present, SEG is published both in print and online. The differences between the two are slight.

This is a problem, because SEG contains a wealth of information. This wealth is now suboptimally accessible. A user would have to read everything to access it, just as with the print. This is both impossible and unnecessary, given the possibilities that the digital medium affords us.

How to improve? To answer this question, we need to go back to the use cases. How do epigraphists and others use SEG? What types of questions are they asking? What kind of answers do they expect?

SEG is a bibliography. Every entry discuss a monograph or journal articles that in their turn discuss Greek inscriptions. But it is a *bibliographie raisonnée*. The publications and the inscriptions are discussed by the editors of SEG. Newly found inscriptions and new readings are represented and discussed. It is here that the unique value of SEG can be found. (This is

what the “S” in SEG stands for: it supplements the existing epigraphic corpora).
It follows that first, the bibliographic information must be identified and made machine-readable. This is done in an automated process that Brill has developed, called *Bibliographic Entity Recognition*. In a first phase, books and articles will receive DOIs and made into openURLS, so that users can refer to the literature much more easily. Additional functionality also becomes available, such as browse, search, or download such data. In a further phase, bibliographic data will be offered in a network tool as a web of publications, citations, and references.
Second, the information about the inscriptions themselves will be made machine-readable. For example, the provenance of an inscription is captured and, through the addition of coordinates, plotted on an interactive map. Similarly for the inscription’s date. This may be called data visualization and will be applied to a few metadata categories determined by international standards.
In a further stage, the contents of the inscription itself can be captured. Proposographical information, for example, can also be made available in a network tool. (Indeed, the same as the bibliographic network mentioned above, thus adding the dimension of “subject” to the web). Brill created the *Brill Text Tool* that offers grammatical and lexical information to users reading source texts. This can be applied to SEG as well.
There is much that can be done, but we would like to do it with epigraphists, classicists, historians, linguists and other scholars and students. We actively seek collaboration and welcome feedback. Continuous and preferably increased usage of SEG Online is good for epigraphy and good for Brill. Together we go forward and develop new ways of making accessible the wealth of information that is SEG.

John Traill: The ATHENIANS Project and Databases
The ATHENIANS Project, soon to be made available in electronic format, offers researchers a vast body of epigraphical, topographical, and prosopographical information. Its range of sources is wide and they include literary, inscriptional, papyrological, archaeological, geographic, ceramic, and numismatic data which have been combed by numerous scholars over many years, then verified, analysed, classified, entered, and stored in relational databases.
This initiative, an 85-year-old research venture, which began with the commencement of the modern phase of the Agora Excavations in 1931 as a simple hand-written card catalogue of personal names found on newly discovered inscriptions, was transformed and expanded four decades later into a computer-based project located at Victoria College and the Computer Systems Research Group (later "Institute"), and finally in the Department of Classics of the University of Toronto. ATHENIANS is now experiencing a renaissance with the support of its long-term software supplier, EMPRESS Embedded Database, and, more recently, the Lassonde School of Engineering at York University. The purpose of the project is to disseminate via a sophisticated electronic relational database management system (RDBMS) the masses of accumulated information concerning more than 100,000 known residents of ancient Attica. Over the past four decades most of the prosopographical information has been entered into two EMPRESS databases and then formated and published in 21 volumes of the series "Persons of Ancient Athens." A good deal of this material, combined with new topographical and epigraphical databases, will be useful to scholars of virtually every aspect of ancient Athens.
As an example I cite the ancient economy, a topic which has been receiving much recent attention and was an important part of the original conception of the ATHENIANS project. Complete references to Davies' "Athenian Propertied Families 600-300 B.C.," for instance, were included in the main relation of the database and provision was madevia a series of more than 100 different kinds of joins to otherpersons both for constructing stemmata or family trees but alsofor establishing groups or sets of persons sharing a common theme or context. All references to Athenian financial officials, e.g. treasurers of Athena, trierarchs, members of the

cavalry, mint magistrates, persons cited in building accounts, lessees, lessors, etc., i.e. the sort of people who are treated in Davies, are included. Our material, however, covers a larger span, namely all ofantiquity down to the Late-Roman and Early-Byzantine periods,and our register embraces not just the wealthy but all classes of Athenian society.
The documentation provides full citations of searchable texts including a complete set of monetary and other special notations in ancient inscriptions, papyrii, and literary texts. I supply several examples of how the organization of the prosopographical databases will facilitate searching. Of the 16 attributes in the Main relation one, "stat", provides several dozen levels for classifying the citizenship and social stutus of an individual. Of the seven attributes in the References relation one, "class", allows a scholar to select material from a 100 different categories of documents. Our system of classifying grave monuments according to 8 levels of increasing elaboration may also be exploited for a number of purposes. In addition, data may be treated with tools of statistical research, some already within the EMPRESS system, for example the five aggregate functions of COUNT, MAX, MIN, SUM, and AVG, or with other tools from external statistical software. Dates have been provided as presented in standard publications such as the Corpus, but we have added two additional formated interpretative attributes, viz "datefrom" and "dateto",creating windows to facilitate computer searching.
The large body of prosopographical information stored in the Main and References relations has been linked through an Interactive Map to topographical and epigraphical databases. This map has standard features such as tiling, hovering, panning, dragging, and zooming, and one can easily move from Marathon on the Map, for example, to relevant information in the topographical database and to other on-line maps like traditional Karten von Attika and modern Google, then to all the Marathonians in the prosopographical database or to the inscriptions found at Marathon, and finally to citations of the deme or the demotai in the epigraphical database of digital images of squeezes. These aspects of the project continue to be expanded and enhanced.
Information may readily be retrieved and formated via a simple user-friendly search form, while more advanced users will exploit the full resources of the highly reputed EMPRESS RDBMS through both traditional methods and also through new techniques of data mining with the goal of making this electronic tool more useful for all scholars. We are creating a Google-like search engine and are redesigning the database, in the words of one of our senior computer science research associates, "to transform it into a research-oriented information system allowing arbitrary and personalized query and exploration of it to answer structured and semi-structured questions, and to do data analysis and inference to discover unknown relations among persons, places, and events."

Rada Varga: *Romans 1 by 1*. New developments in the study of Roman population
The proposed paper documents *Romans 1 by 1*, a population database working with Roman-era inscriptions. The first general presentation of the database and its metadata were published in the proceedings of the EAGLE 2016 international conference on *Digital and Traditional Epigraphy in Context*, but as this is a very dynamic tool, its configuration has undergone important mutations since August 2015 (when the article was sent for publishing).
While the database architecture hasn't changed, the metadata has increased significantly: we began with a metadata adapted for registering members of the middle classes of the Latin language provinces of the Empire, but expended it in the fall of 2015 to accommodate all people attested (at least) in Roman Dacia and Moesias. So, the epigraphic categories diversified, the fields required for registering life details/*cursus honorum* became more complex and the linguistic span extended (the Greek inscriptions of the Black Sea shore of Moesia Inferior) – bringing forth new epigraphic habits and patters as well. Besides the present structure, we will present the difficulties faced and questions raised when expanding and diversifying the metadata, as well as the solutions we opted for and our motivation(s) in doing so. For most of

our problems, scientific or technical in nature, as we have faced both, there usually was more than one solving option and it is relevant for which one we have opted, as it reflects our goals and envisioned future uses of the database.
The last section of the presentation will focus on some applications of the database. The most obvious ones, which were the focus of our interest so far as well, refer to prosopographical reconstructions (linking people which have not been linked/identified as the same person throughout more inscriptions and reconstructing relatively fluent life courses) and network analyses (we have focused on professional networks in the Low Danube provinces so far and the visualization results are quite interesting).
Finally, we intend to present our vision on the future of the database, plans for linking it to the big on-line prosopographical resources (SnapDrgn, Trismegistos), as well as its immediate future scientific employments.