



Places, people, documents

Egypt:

Verse Inscriptions from CPI

Oil Smugglers in Ptolemaic Egypt

Sicily:

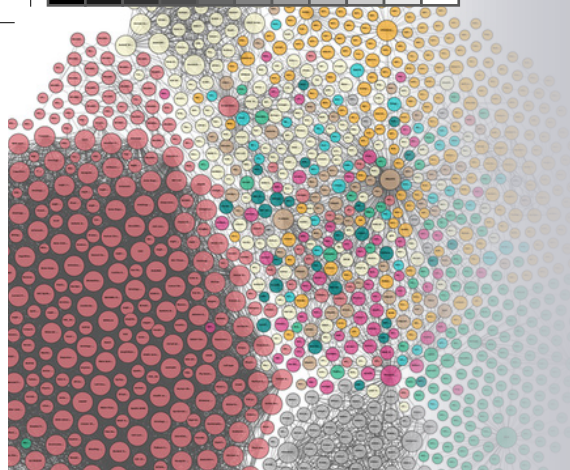
Mapping olives

Epigraphy meets petrography

Dots between words in Sicilian inscriptions

Turkey and Bactria:

News from the ARCH, BGR and CHANGE Projects



FAIR Epigraphy

Making Data Findable, Accessible, Interoperable, and Reusable

Jonathan Prag

As any reader of this newsletter will know, the world of epigraphy has been going increasingly digital in recent decades. ‘Digital’ takes many forms, ranging from the simple publishing online in PDF of previously printed material, through the presentation of new material such as digital images or searchable databases, all the way up to fully born-digital editions, sometimes rich in data linked to other projects. As a result, it is now possible to access page-scans of many volumes of the *CIL*, to search texts in text-databases such as *PHI*, *EDH*, or *Clauss-Slaby*, to search names in the online *LGPN*, or even to browse complete editions in online corpora such as *Laph2007*, *RIB online*, or *ISicily*.

However, this growing wealth of resources brings its own problems (only a few of which are highlighted here). Free online PDFs increase accessibility in a most basic sense (everyone can have a copy of *CIL* for themselves), but these are little more than screen-readable books. Text databases such as *PHI* or *Clauss-Slaby* make texts searchable in powerful new ways beyond a printed index. But what is the value or status of a text stripped of almost all context and reproduced from a frequently unspecified source, and possibly introducing new typographic errors? New databases and digital editions are powerful resources: but every digital collection looks and behaves differently and contains different types of information; texts may appear in more than one such resource, with different information (some might be new editions; others simply reproduce existing editions); and no-one can tell how long such a resource might last and whether it can be cited with security (and

all too often no-one really knows how to cite it even in its current form).

There have been several attempts to meet some of these challenges: Tom Elliott and colleagues developed the EpiDoc standard to enable projects to generate machine-readable epigraphic editions in a standardised way, and this is now widely recognised and used (but many epigraphers still find the encoding itself too rebarbative for daily use and the software to facilitate naïve human engagement requires constant development and maintenance); the Europeana-funded EAGLE project led by Silvia Orlandi attempted to unify the outputs of multiple epigraphic databases and digital corpora by collecting a minimum set of data from multiple projects and re-presenting it in a new unified database (but despite their best efforts to standardise data from different projects, the end result was weakened by the difficulty of easily comparing data between projects, and once the funding ran out, the work halted); and the Trismegistos project at KU Leuven has built upon its initial papyrological work to establish a single common identifier for every published inscription, creating the potential to link multiple editions more easily (but even Trismegistos must now charge a subscription for access to many of its services, to keep the resource alive).

However, these projects have laid the foundations for a more optimistic future. Almost two decades ago, Tim Berners Lee set out the core principles for accessible data on the internet (the five steps of Linked Open Data). Data needs to be freely and publicly accessible; but more than that, it needs to be available in formats that anyone can use (e.g. not require specialist software, which must be paid for); and it needs to be organised and published in ways that make it easy to combine with other data (e.g. the computer needs to know that ‘limestone’ = ‘calcare’, that ‘funerary’ = ‘sepolcrale’; and humans need to agree on which things are

being classified in which way). EpiDoc gives the texts standard form, similar to but richer than the Leiden conventions, and machine-readable; EAGLE began the work of aligning our vocabularies to describe things (and the Epigraphy.Info community has continued this work); and Trismegistos provides the tool to link editions and datasets for the same inscription. In the scientific world this set of ideas has been taken a step further, with the FAIR principles, that all data should be ‘Findable, Accessible, Interoperable, and Reusable’.

To make this a reality for the Epigraphic world some work still needs to be done: standards still need to be agreed in many areas; and both the guidance and the software tools need to be developed and made available to the community to make it easy for projects to publish their data according to these standards and for researchers and the wider public to be able access this data all together in an easily searchable form. But this is now a realistic prospect, and the FAIR Epigraphy project, funded jointly by the AHRC and the DFG, led by Marietta Horster (Mainz) and Jonathan Prag (Oxford), with the support of Petra Hermankova (Mainz) and Imran Asif (Oxford) is dedicated to achieving this, over the next three years. For more information (and a pilot demonstration) visit us at: <https://github.com/FAIR-epigraphy>.

Ultimately, however, for this optimistic vision to come true, something of a cultural change is also needed — published data needs to be genuinely open and freely accessible, not buried behind a paywall for consumption alone in an ivory tower. Such an approach does not preclude the maintenance of intellectual property rights, but that would need another article.



Arts and Humanities Research Council

