Digital and Traditional Epigraphy in Context.

Francesco Mambrini

SAPIENZA
UNIVERSITÀ EDITRICE
2015

Sapienza Università Editrice
Piazzale Aldo Moro 5 – 00185 Roma

www.editricesapienza.it
edizioni.sapienza@uniroma1.it

ISBN 978-88-95814-XX-X

Iscrizione Registro Operatori Comunicazione n. 11420

Contents

Preface

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna.

Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Silvia Orlandi

Conference committee

Pietro Liuzzo
Seminar fÃijr Alte Geschichte
und Epigraphik
Ruprecht-Karls-UniversitÃďt Heidelberg
69117 Heidelberg, Germany
pietro.liuzzo@zaw.uni-heidelberg.de

Francesco Mambrini
Faculty of Mathematics
and Natural Sciences
Institute of Physics
and Semiconductor Physics
Carl von Ossietzky University
26111 Oldenburg, Germany
mario.rossi@uni-oldenburg.de

Raffaella Santucci
Faculty of Mathematics
and Natural Sciences
Institute of Physics
and Semiconductor Physics
Carl von Ossietzky University
26111 Oldenburg, Germany
mario.rossi@uni-oldenburg.de

Silvia Orlandi
Faculty of Mathematics
and Natural Sciences
Institute of Physics
and Semiconductor Physics
Carl von Ossietzky University
26111 Oldenburg, Germany
mario.rossi@uni-oldenburg.de

Claudio Prandoni
Faculty of Mathematics
and Natural Sciences
Institute of Physics
and Semiconductor Physics
Carl von Ossietzky University
26111 Oldenburg, Germany
mario.rossi@uni-oldenburg.de

Francesco Mambrini f.mambrini@gmail.com
DAI chair Raffaella Santucci raf-
faella.santucci@uniroma1.it http://digilab.uniroma1.it
Sapienza UniversitÃă di Roma
chair Claudio Prandoni prandoni@promoter.it
http://www.promoter.it, www.digitalmeetsculture.net
Promoter Silvia Orlandi silvia.orlandi@uniroma1.it
Sapienza UniversitÃă di Roma
Rainer Simon rainer.simon@ait.ac.at
http://www.ait.ac.at/ Austrian
Institute of Technology Rada Varga
radavarga@gmail.com BabeÈŹ-Bolyai
University Monica Berti monica.berti@uni-
leipzig.de http://www.monicaberti.com
UniversitÃďt Leipzig Sorin Her-
mon sorin.hermon@gmail.com http://www.cyi.ac.cy
The Cyprus Institute Charlotte
RouechÃľ charlotte.roueche@kcl.ac.uk
King's College London Leif Isak-
sen leifuss@googlemail.com http://leifuss.wordpress.com
University of Southampton Valentina
Vassallo vassallo@cyi.ac.cy The
Cyprus Institute John Bodel john$_b$*odel@brown.eduhttps* :
//research.brown.edu/myresearch/JohnBodelBrownUniversityAlisonCooleya.cooley@
//www.paregorios.orgNewYorkUniversityBryanWard −
Perkinsbryan.ward − perkins@trinity.ox.ac.ukOxfordUniversityRobertMatijasicrmatij
//www.edb.uniba.it/UniversitdiBariMarkDepauwmark.depauw@kuleuven.beKULeuv
sazu.sihttp : *//iza.zrc − sazu.siZRCSAZU, InstituteofArchaeologyAntonellaFresafres*
//www.promoter.itPromoterS.r.l.AllisonKupietzkyallison@imj.org.ilhttp :
//www.imj.org.ilTheIsraelMuseum, JerusalemDonatoFasolinidonato.fasolini@gmail.c
ClaireBeaulieumarie − claire.beaulieu@tufts.eduTuftsUniversityGianvitoPiogianvito.p
//www.di.uniba.it/ gianvitopio/UniversitdiBariAldoMoroStephenLambertlamberts@
bordeaux − montaigne.frAusonius, UMR5607, CNRSAntonioEnricoFelleae.felle@gma
lyon2.frUniversityofLyonGabrileGuidig.guidi@ieee.orghttp :
//www.english.polimi.it/PolitecnicoMilanoFabrizioFalchifabrizio.falchi@isti.cnr.itht
//www.fabriziofalchi.itCNR −
ISTIGiuseppeAmatogiuseppe.amato@isti.cnr.ithttp :
//www.isti.cnr.it/People/G.AmatoCNR −
ISTIMaurizioLanamaurizio.lana@uniupo.ithttp :
//www.lett.unipmn.it/docenti/lanaUniversitdelPiemonteOrientaleAndreaZannizann
Grunaisfrancisca.feraudi − gruenais@zaw.uni −
heidelberg.deEpigraphicDatabaseHeidelbergTassieGniadyctgniady@indiana.eduhttp :

Part I

Digital Editions of Inscriptions: methods, problems, approaches

Part II

Collaborating in Digital Epigraphy

Mapping Epigraphic Databases to EpiDoc

Liuzzo Pietro[a,*]

[a] *Ruprecht-Karls-Universität Heidelberg - Marstallstraße 6, 69117, Heidelberg (DE)*

Abstract

The Europeana network of Ancient Greek and Latin Epigraphy brings together most repositories of ancient epigraphic material and aims to provide historians not just with a useful research tool, but a curated online edition which has high quality content as well as high quality data. In this paper some of the up-conversion, alignment and enrichment tasks are presented.

Keywords: EpiDoc, XML, Vocabularies, Linked Data, Data harmonization

1. Introduction

The Europeana network of Ancient Greek and Latin Epigraphy[1] brings together most repositories of ancient epigraphic material and aims to provide historians not just with a useful research tool, but a curated online edition which has high quality contents as well as high quality data. Towards this end many steps are needed as the databases on which many institutions have worked for decades cannot be just discarded. Epigraphy enjoys a special position among digital resources as most transcriptions of existing classical inscriptions have been digitized. Many duplicates exist and a lot of the content lacks metadata almost entirely, but the work continues with dedication so that among the 10 % of digitised European Cultural Heritage, Ancient Epigraphy can certainly claim to play a strong role.[2] While the long term aim for epigraphy is on the one hand to keep up with new finds and studies and on the other to have a common and flexible entry point and backend on which to work together, this is not easy to accomplish without a careful process

* Corresponding author. Email: pietro.liuzzo@zaw.uni-heidelberg.de

1 Orlandi et al. (2014a) Orlandi et al. (2014b) Liuzzo et al. (2014) http://www.eagle-network.eu/

2 The data is taken from the Europeana 2020 strategy report available at http://strategy2020.europeana.eu/.

which involves a gradual transition from independent databases (the best option at the time when the EAGLE databases were created) to a common online data entry and editing system working on XML files in EpiDoc, as it has now become feasible. In this paper some of the mapping and content harmonization efforts undertaken by the EAGLE BPN to achieve this step towards a common epigraphic resource, will be presented as a case in which this apparently technical task involved deep revision of contents related to the discipline and required discussions and collaborative efforts by the working groups of digital epigraphists, encoders, epigraphists and developers, coming from several different background and experiences. The result of this work are the automated conversion to TEI EpiDoc XML of 90% of the text features in the EAGLE databases and the EAGLE Vocabularies, the largest existing controlled and aligned vocabularies for classical epigraphy.[3]

2. Why XML

Choosing XML for an online publication is almost obvious given the current stage of development of the editorial methodology in epigraphy. However there is room for a brief explanation of why this is preferable on a large scale to a relational database. I would like to suggest a comparison here. Coffee is produced by filtering water through a powder more or less thick. This can be accomplished by pushing water through the coffee (as in most systems, from a mocha to an espresso machine) or coffee through the water (as in a french press, where the coffee floats and is pushed down). There is pretty much the same difference when we compare the action of entering data into a database, i.e. a rigid structure of information, and marking up a text, adding tags to the text itself. The first process forces data through a grid however well thought off, while in the second case structure is added on the contents without altering them, but just enhancing their potential so that the richness and values of both contents and structure empower each other. There is just more taste and much more freedom of encoding to the level desired and to the complexity one wants. An additional advantage is the level of interoperability. An XML file is virtually software independent, it is not an Access, FileMaker or whatever database format. A third reason is that an XML file can host a complete textual description with

[3] http://www.eagle-network.eu/resources/vocabularies/

semantic markup and still have the structure of a normal text edition, while the same source can be used to produce a database or an index for interrogation purposes. For example the same markup of an abbreviation `<expan><abbr>C</abbr><ex>aio</ex></expan>` can be used to produce a database of abbreviation marks and expansions, to output in a diplomatic edition only the character actually on the stone or to print the expanded abbreviation in a critical edition.[4] To print a database to a book would be at least more cumbersome. XML provides also a better service as an archival format being explicit on the values it uses to describe contents, moreover if these follow a schema of agreed value of such elements. This said producing and sharing XML files means that more quality editions can be produced, printed editions can be produced and all sort of outputs can be supported with flexibility; it means that data is safe from software developments and support; it means you don't have to put information into a rigid table which is then hard to restructure, but that you can annotate your text as much as you want thus adding layers and layers of "possible databases" of relevant information to be extracted and used. The text is what one uses to describe situations and complexities, the markup allows also the machine to know and retrieve that information as in a database: you don't need a field or an element to state what the orientation of your text is, you can simply state and describe as you would do in a normal publication and if you want to reuse information about the text orientation then you can add elements and attributes to do so. XML allows not to worry too much on "what goes where" or about "where do I put that info" because it is not a database, so that the editor can instead focus on "what is that" and on accurately describing it. The TEI specification which is EpiDoc allows the specific epigraphic content to be described fully for what concerns the text of inscriptions, drawing on more than a decade long experimenting and development of this standard thanks to a vast and active international community. The simple fact of using the same markup to describe our texts allows us to study them together and to produce and run software for more then just one corpus of inscriptions.

[4] http://www.stoa.org/epidoc/gl/latest/trans-abbrevmark.html

3. EAGLE Mappings and Metadata modeling

The EAGLE BPN has taken 18 different content providers and has come to the decision to take the first steps to bring hundreds of thousands of ancient Greek and Roman inscriptions to TEI. This required 14 mappings of local database metadata models to TEI/EpiDoc, as well as the elaboration of XSL Transformations[5] for up-conversion[6] and the alignment of the text markup. Although this was thought to be a trivial task it turned out that whereas databases claimed to use the same conventions, a considerable amount of differentiation took place due to the data structure and to data entry procedures as well as to policies and internal decisions.

The main harmonization task undertaken has been to align the XML format of data provided for aggregation and ingestion, to the TEI specification EpiDoc.[7] This well established TEI schema, broadly used in many high quality projects, allows for a very easy alignment, for the production of an XML file compliant with international standards and for high flexibility for integration of the vocabularies and places gazetteer in use.

Moving from a database to a marked up text is a partially mechanical operation which involves a theoretical jump. As coffee is better when water goes through coffee powder rather than when coffee goes through the water (as in a french press), a database has fields to be filled, while marking up a text is a descriptive activity which is attached to information whichever textual form it takes.[8] Mapping from a database to XML forces into the XML a structure and a logic which is that of a database, whereas the freedom and flexibility achieved with XML are yet to be actually realized and exploited.

In order to offer a complete and critically structured endpoint to the user on the side of data, to describe inscriptions and their representations, EAGLE considered beside TEI also CIDOC CRM, studying and

---

5 https://github.com/EAGLE-BPN

6 This is the terminology used by Kay (2008, 906) to describe transforming without explicit structure in data which has it. In this case it is a transformation from a less explicit database structure to a fully explicit and interoperable one.

7 http://www.stoa.org/epidoc/gl/latest/

8 See above.

providing a full EpiDoc to CIDOC-CRM mapping.[9] This would have enabled a further full description in a different logic, that of the web of data, but the attempts made proved the need for more specific efforts.

There are several mappings involved in the aggregation work of EAGLE. Content providers need to map to a common Eagle Metadata Format, and the data produced will then be mapped to the Europeana Data Model[10] in order to be aggregated in Europeana, for a wider dissemination with a special eye for the general user.

4. EAGLE Metadata Model and Harmonization

The occasion of a mapping work allows also for other tasks of curation to be performed. EAGLE Members had in some case the text of a same inscription with three parallel different types of encoding conventions applied.

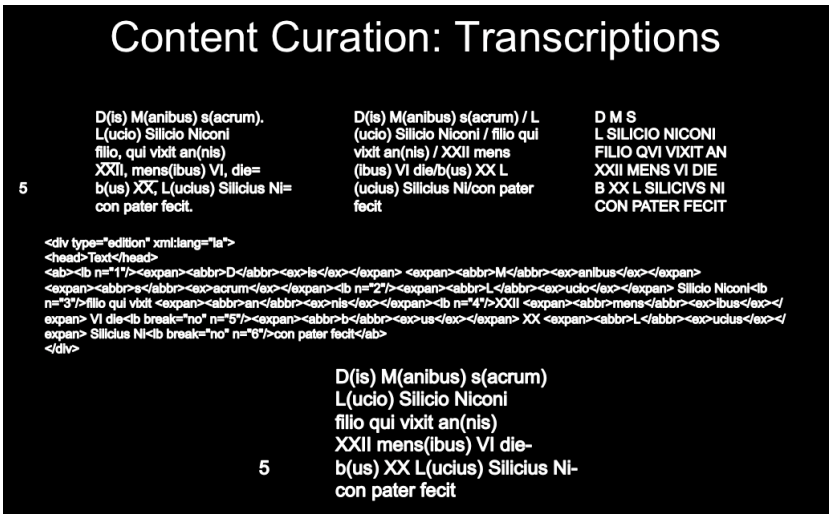

Fig. 1. Different Texts before and after harmonization though up-conversion

In the context of the mapping to EpiDoc of the metadata, the text also underwent transformation with tools which have been developed to sup-

[9] Still under development, within the ARIADNE project. http://www.ariadne-infrastructure.eu/. The need for a working group and a variant declension of CIDOC for epigraphy have been highlighted during the Nicosia EAGLE meeting in March 2015.

[10] Rivero Ruiz and Vassallo (2013)

port the alignment and harmonization of data from content providers to international standards for what concerns digital editions of inscriptions.

Given the template described in Part III, and ANNEX II of the EAGLE Metadata Schema (RIVERO RUIZ AND VASSALLO (2013)), an XSL Transformation converts from string epigraphic texts in marked up TEI-EPIDOC XML, following the EpiDoc guidelines (ELLIOTT ET AL. (2007)).

These XSLT:[11]

1. allow the conversion of epigraphic texts with various encodings and conventions from string to Epidoc markup and valid against the The EpiDoc RelaxNG schema.

2. Populate appropriate elements with available common URI from the EAGLE vocabularies[12]

This export set up will also guarantee that contents are kept aligned to the EpiDoc guidelines at all stages guaranteeing an effort free alignment to these international conventions for partners who can continue to apply local conventions for editing. But I would like to give more details on the steps of this process as an example of how it was possible to extract semantics, entities, and patterns from these text while aligning metadata format to an internationally recognized standard.

4.1. Step 1

Each project uses different conventions and therefore the regular expressions used to match particular situations are different. The process of mark up of the string text in div[@type="edition"] is accomplished in several steps to guarantee consistency and precision.

The `textstructure.xsl` looks for marker of different sections and tokenize them to apply the same XSLT to each section of the text which needs to be contained by an <ab> element. If there is only one part it applies following instructions to that only.

Each section of text is then processed by the `brackets.xsl`. Normalizing brackets is important for the following steps and splits individual semantic values. The notation [ort 3], which would mean that a supplied text is followed by a gap of three letters, is divided into [ort][3].

---

[11] Based on Chetc.txt (by Hugh Cayless, Elli Mylonas, Gabriel Bodard and Tom Elliott) and further support from the Epidoc Collaborative (especially from Gabriel Bodard)

[12] http://www.eagle-network.eu/resources/vocabularies/

The normalized string which results from this process is then passed to the `up-conversion.xsl` which works using a specific operation to search for regular expressions patterns (`xsl:analyze-string`) and substitute them with correct xml elements.

Running the transformation on the pattern <E=F> will return the following result:

```
<choice><corr>E</corr><sic>F</sic></choice>
```

The result of this template is then passed to a further template which gives consistent numbers (`insertnumbers.xsl`). Empty lines do not need to have numbers, so Xpath is used to evaluate where to put a 0 as value of the @n in the <lb> element. Starting from this

------] / e[t?] Q(---) Bl(a)e[sus?] / contub/ernalis / eius / d(e) s(uo) l(ibens) l(aetus) d(edit)

The result of this processes is then the following[13]

```
<ab>
<lb n="0"/><gap reason="lost" extent="unknown" unit="line"/>
<lb n="1"/>e<supplied reason="lost" cert="low">t</supplied>
<abbr>Q</abbr>
Bl<expan><ex>a</ex><abbr>e</abbr></expan><supplied
reason="lost" cert="low">sus</supplied>
<lb n="2"/>contub
<lb break="no" n="3"/>ernalis
<lb n="4"/>eius
<lb n="5"/><expan><abbr>d</abbr><ex>e</ex></expan>
<expan><abbr>s</abbr><ex>uo</ex></expan>
<expan><abbr>l</abbr><ex>ibens</ex></expan>
<expan><abbr>l</abbr><ex>aetus</ex></expan>
<expan><abbr>d</abbr><ex>edit</ex></expan>
</ab>
```

[13] The amount of feature supported is much higher and some example of complex text successfully converted can be seen in this presentation online

4.2. Step 2

On the elements which contain information such as Object Type, Material, Execution Technique, etc. which are typically handled with a controlled vocabulary a series of XSL transformations is passed, one for each vocabulary related to that element to match the content of the element with the vocabulary entry into the SKOS vocabulary stored in git, regularly updated and published as a self standing resource on the EAGLE portal.[14] What follows is a brief description of such vocabularies and their development.

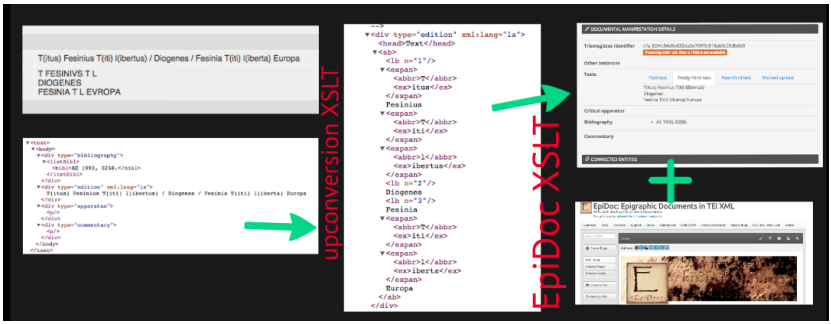

Fig. 2. Stages of transformation of a text: database, template, up-conversion, checking and display with EpiDoc XSLT.

5. Classification problems: the EAGLE Vocabularies

As Di Stefano Manzella [15] clearly explained classification is no easy issue in any field: epigraphy is no exception to this rule. Traditionally the CIL VI (Rome) classification has been used as a reference, as this typology has served as a model for all epigraphic production in the Roman Empire. There are nevertheless new glossaries and classification curated by CIL, which retain the scientifically selective constraints of a formal classification, together with the benefits of this.

Problems are various, and include also the use of terms across vocabularies and the doubts which might be generated by archaeological chance.[16]

---

14 http://www.eagle-network.eu/resources/vocabularies/

15 MANZELLA (1987, 109)

16 See PISO (2001, XI-XII), following a current of studies which has its main in G. Susini and J.-N. Bonneville.

What looks like an altar could be also the base of a statue, for example. Or perhaps it could have served two different function has an object of which we might or not have any archeological, contextual or textual trace.

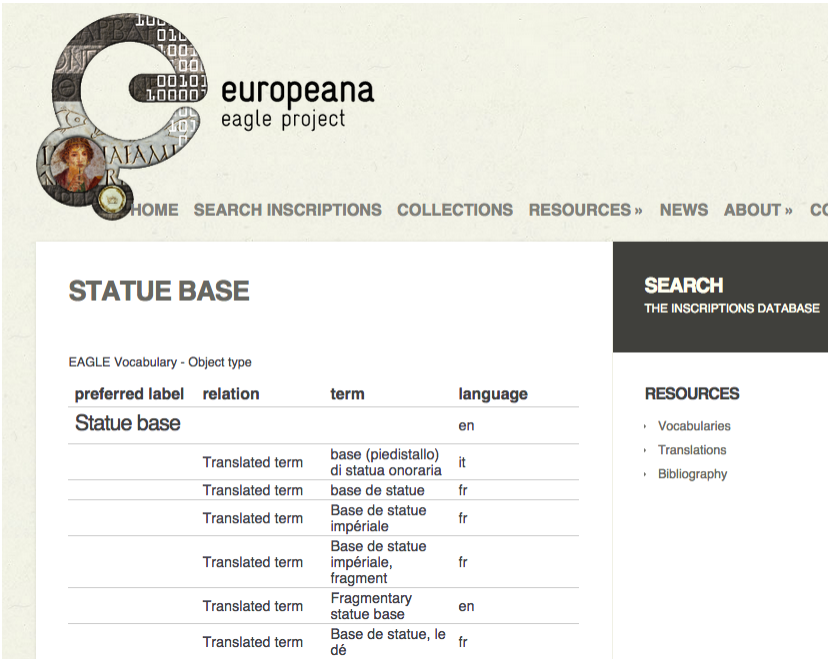

Fig. 3. Statue Base

This is also the case when we deal with techniques of execution of an inscription. Those can also be multiple and an inscription can be for example both a graffito and painted. And this is just not to mention the extreme complications of having a lot of independent vocabularies, each one for its own sake and small scope purpose. This is often the best option for correct attribution as local habits do vary. But within the framework of harmonization activities, it is important to refer to connected and interrelated definition of terms (not univocal!) and to allow for multiple values to coexist. This is one best practice in the definition of vocabularies which is very important for the nature of content in question. This is also the reason why the EAGLE decided to provide not just definitions of all main terms but also examples from different areas and in different state of preservation, and as far as possible also bibliographic reference to authoritative sources.

The EAGLE Vocabularies are the following

- Type of Inscription
- Object Type
- Material
- Writing and Execution
- Decoration
- State of Preservation
- Dating Criteria

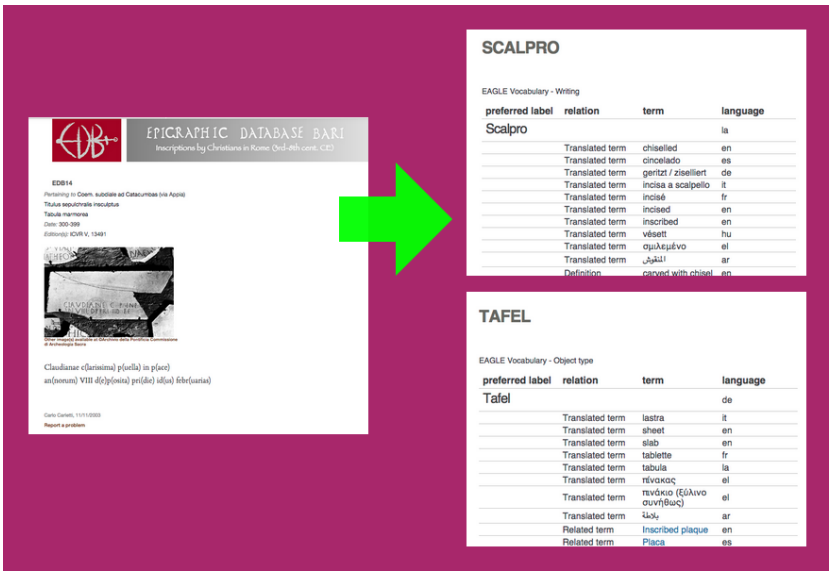

Fig. 4. EAGLE Vocabularies used by the Epigraphic Database Bari

The EAGLE vocabularies are maintained in a git repository where they are manually updated and transformed to a readable versions (with another XSLT transformation) published on the EAGLE website as a self standing resource linked to several others. A user is thus able to search for a type of object and read a definition as well as searching for more information on Wikipedia, Wikidata, Wikisource, etc. including the partners' websites. The Vocabularies perform thus both a technical function and are at the same time a living resource which can be read and used independently.

6. Linked Open Data

The Europeana LOD practice for metadata recommends the adoption of machine readable vocabularies. Within the EAGLE BPN Linked Open Data practices and approaches have been taken on board for a process which will bring current material some steps closer to that practice.[17] Addressing existing problems in classification and the publication of a machine readable vocabulary of values (controlled vocabulary) is one of these steps:[18] it involves identifying and matching values as well as giving them stable identifiers and making of them self standing resources where possible.

In facts the many problems and complexities which Manzella (1987) underlined and exposed can find a solution using the LOD approach.[19] Both the limits of choices and hierarchical organization can be bypassed using unique identifiers and relations among those. Polyhierarchic structures as well as the presence of many possible denomination with specific and self standing definition allows for both precision, consistency and sustainability over time of this approach.[20]

For example, an inscription can easily be classified both as *magistrati populi romani* and as *decuriones municipales*; a document can be classed both under the *fasti* and under the individual mentioned in the text. These major advantages can be exploited when the effort of publishing open machine readable material is undertaken and a document can be then classified chronologically and alphabetically without the need of an authorial choice.[21]

Nevertheless, the major possible advantage is that there is no evident need, with a LOD approach, to distinguish among Greek and Latin inscriptions: they can find their place together in a Linked Data edition and benefit of other efforts in this direction.

7. Crosswalking EpiDoc to EDH with XSLT

The proof of concept of usability of the XML source file is given by the reverse function being possible, i.e. to transform the XML files into a

---

17 Bizer et al. (2009)

18 Harper et al. (2012, 4-5) for the definition. See also Isaac et al. (2012)

19 Bizer et al. (2009)

20 Harpring (2010)

21 I would like to thank prof. Piso for input on this point, given during the work of the Working Group.

database data structure. A team of people from the EAGLE BPN[22] and Scott Vanderbilt, curator of the online edition of the Roman Inscriptions of Britain,[23] worked on mapping the EpiDoc contents of that project to the EDH database model,[24] facing a series of problems and challenges, which brought an interesting discussion, also on principles of "up-conversion" and cross-walking, which should be as often as possible multi-directional.[25]

Challenges included

- alignment to EDH internal vocabularies
- alignment to EDH conventions for bibliographic references
- use of internal references and retrieval of key based information
- matching of existing items
- formatting to EDH conventions differentiated based on the type of content extracted (text or names)

The result was successful and proves once more if needed that an XML source format is a preferable option whatever the variety of desired output are.

8. Conclusion

I have presented in this paper the effort of a large consortium of people and institutions belonging to the same field of interest, that of Ancient Roman and Greek Epigraphy. Mappings and harmonization of data have been for the EAGLE project not just a way to allow machines to do more, but an intellectual effort of revision of contents and establishment of best practices. This work prepares further steps which hopefully will take place soon, in which one repository of inscriptions will be accessible to all to edit, contribute and download data from several websites and interfaces. In such environments related contents

[22] Gabriel Bodard, Pietro Liuzzo, James Cowey, Frank Grieshaber, Brigitte Gräf and Francisca Feraudi Gruénais.

[23] http://romaninscriptionsofbritain.org/

[24] The work was largely based on the previous exercise of this kind, made by Gabriel Bodard and James Cowey for the Inscriptions of Roman Tripolitania.

[25] The files are available here: https://raw.githubusercontent.com/EAGLE-BPN/Epidoc2EDH/master/rib_to_edh-2.xsl

based on descriptive URIs, as well as places, personal names and other semantic information will be possible and meaningful and perhaps in a not too far future classicists will have a complete and functioning network of Linked Ancient World Data, including texts, manuscripts, papyri, coins and inscriptions.

References

Bizer, C., Heath, T., Berners-Lee, T., *Linked Data - The Story So Far*, 5 (2009), 1–22, ISSN 1552-6283,, URL http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf, this is a freely available preprint of the article later published in IJSWIS.

Elliott, T., Bodard, G., Milonas, E., Stoyanova, S., Tupman, C., Vanderbilt, S., *EpiDoc Guidelines: Ancient documents in TEI XML*, (2007), URL http://www.stoa.org/epidoc/gl/latest/.

Harper, C., Dunsire, G., Hillman, D., Phipps, J., *Linked data vocabulary management: infrastructure support, data integration, and interoperability*, 24 (2012), ISSN 1041-0031, cited by 0000.

Harpring, P., *Introduction to Controlled Vocabularies Terminology for Art, Architecture, and Other Cultural Works*, 2010, URL http://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/pdf.html.

Isaac, A., Clayphan, R., Haslhofer, B., *Europeana: moving to Linked Open Data*, 24 (2012), 34–40, ISSN 1041-0031, cited by 0000.

Kay, M., *XSLT 2.0 Programmer's Reference, 3rd edition*, Wrox, 2008.

Liuzzo, P.M., Zanni, A., Martinelli, L., Losa, L., De Nicolao, P., *The eagle mediawiki. a fully collaborative database for academics, data engineers and the general public*, in S. Orlandi, V. Casarosa, R. Santucci, P.M. Liuzzo (ed.), *Information Technologies for Epigraphy and Cultural Heritage*, 187–200, Sapienza Università Editrice, 2014, ISBN 978-88-98533-42-8, doi:10.13133/978-88-98533-42-8, URL http://www.eagle-network.eu/wp-content/uploads/2015/01/Paris-Conference-Proceedings.pdf.

Manzella, I.D.S., *Mestiere di epigrafista: guida alla schedatura del materiale epigrafico lapideo*, Edizioni Quasar, 1987, ISBN 9788885020795.

ORLANDI, S., CASAROSA, V., SANTUCCI, R., LIUZZO, P.M. (ed.), *Information Technologies for Epigraphy and Cultural Heritage*, Sapienza Università Editrice, 2014a, ISBN 978-88-98533-42-8, doi:10.13133/978-88-98533-42-8, URL http://www.eagle-network.eu/wp-content/uploads/2015/01/Paris-Conference-Proceedings.pdf.

ORLANDI, S., GIBERTI, L.M.C., SANTUCCI, R., *EAGLE: Europeana Network of Ancient Greek and Latin Epigraphy. Making the Ancient Inscriptions Accessible*, (2014b), URL http://lexicon.cnr.it/index.php/LP/article/view/408.

PISO, I., *Inscriptions d'Apulum*, Inscriptions de la Dacie Romaine - III 5), 2001.

RIVERO RUIZ, E., VASSALLO, V., *EAGLE metadata model specification*, Tech. rep., 2013.

TIGLIO. Translations and Images of Greek and Latin Inscriptions Online

Almas Bridget[c], Baumann Ryan[d], Beaulieu Marie-Claire[c], Cayless Hugh[d], Cowey James[a], Liuzzo Pietro[a,*], McCourt Finlay[b], Sosin Joshua[d]

[a] *Ruprecht-Karls-Universität Heidelberg - Marstallstraße 6, 69117, Heidelberg (DE)*
[b] *Glasgow University - Attic Inscriptions Online (UK)*
[c] *Tufts University (USA)*
[d] *Duke Collaboratory for Classics Computing (USA)*

Abstract

This paper describes the aims of a project funded by The Andrew W. Mellon Foundation to speculate on the best ways to deal with two forgotten types of content in the realm of ancient epigraphy: translations and images. Translations available online are less then 3% of the total available transcriptions online; images are often subject to policies which make extremely difficult their use for research, publication, even simple viewing. There has been no thought given to these before in an articulated manner although these are the types of content which can bring a much larger group of users to epigraphy.

Keywords: Translations, Images, Greek and Latin, Epigraphy, Online

1. Introduction

Nearly all Greek and Latin epigraphic texts are available, sometimes in multiple versions, after three decades of continuous digitization and online publication. Without counting repeated inscriptions and *instrumenta*, there are about 300.000 Latin inscriptions[1] online and about 200.000 Greek Inscriptions[2]. In some databases there is also abundant metadata and a structured bibliography. The situation for translations

* Corresponding author. Email: pietro.liuzzo@zaw.uni-heidelberg.de

1 EAGLE disambiguated total: 235.626 ; EDCS not disambiguated total without *instrumenta* and *inscriptiones christianae*: 308.581. On the definition of Roman Epigraphy, see Panciera (2012).

2 Data from the latest Integrating Digital Epigraphies's (https://youtu.be/OPfDj_hjeok) harvest from the Packard Humanities Institute, Searchable Greek Inscriptions (http://epigraphy.packhum.org/), with some duplicates, 207.964.

and images of these inscriptions (text and support) available in the digital space is nevertheless quite different. The ratio between images of inscriptions and text is of one image every two inscriptions,[3] but those inscriptions which have photographic documentation usually have many photos.[4] Translations are present only in smaller corpora edited online in most cases and the Attic Inscriptions Online project[5] is an *uniquum* in its intent to provide mainly translations of inscriptions.[6] There are many publications offering translations in print, but these are not published online. An estimate calculation, based on the 11.000 translations present in the EAGLE Media Wiki, and known collections of printed translations of inscriptions, compared to a total of texts usefully translatable of around 300,000 texts, brings to an alarming 10% of translated texts, of which only a third (slightly more than 3%) is online. Translations are perhaps not a priority for researchers who know Greek and Latin, but are a way to clarify the interpretation of a text and an invaluable tool for didactical purposes and teaching: they are the only way in which an inscription can reach a wider public in a significant way as part of cultural heritage. The same could be said for images, even more obviously, since researchers also need them because: 1) they cannot always reach the place where an inscription is stored to study it (given that the inscription is still there); 2) there are cases in which a photo might be all we are left with and these are quickly increasing as monuments get lost or are destroyed. The imbalance in the documentation is thus pressing, since translations and images are our two best controls on the constitution and interpretation of ancient documentary texts. To an extent, digital epigraphy today is the direct descendant of epigraphy's 19th century analog self: many texts, few translations, few images. This project aims to take initial steps to redress that imbalance, building resources that al-

---

3 At the time in which this paper is being written (November 23, 2015) the Epigraphic Database Heidelberg has ca. 35.000 photos and ca. 71.000 texts (0,5); The Epigraphic Database Roma has a slightly better ratio with 45.000 photos for 71.000 texts (0,6); The Epigraphic Database Bari has 34538 texts and 10341 images (0,3). In the EAGLE aggregator, the total ratio (excluding the related content of Arachne), is of 0,79 images for each text (235626 documental entities per 185999 visual entities) because smaller corpora tend to have a better photographic documentation.

4 The Epigraphic Database Heidelberg has to date ca. 14.000 records with a photo or a drawing attached, bringing the average number of images per inscription to 2,5.

5 https://www.atticinscriptions.com

6 Lambert and McCourt (2014)

low epigraphists and ancient historians to bring translations and images more closely into the suite of existing digital epigraphy resources.

2. Problems

Let's look at the data we have. The EAGLE project has gathered some insights on small collections of images of inscriptions openly published online, on Wikimedia Commons,[7] and on a set of translations of inscriptions, collected in the EAGLE Media Wiki.[8] We shall compare their impact and reach on the wider public to that of the texts of inscriptions, to underline the urgency for these materials to be produced also in order to bring epigraphy outside its restricted circles. Let's look at the visits to the Epigraphic Database Clauss-Slaby, the largest collections of texts (with minimal metadata and no directly stored image): EDCS has an average 3000 requests per day.[9] The result page is always one, containing all the results from the database, which has a total of 491353 texts. We have no means to provide better data unfortunately but for the comparison these will be enough. The images collected under the category "Media Contributed by EAGLE" on Wikimedia Commons contains instead around 8000 photos of inscriptions and we have some good insights on this data[10]. These photos have been viewed in 19 months 22,236,085 times. Another interesting information is the number of people who have worked on them, by no means only members of the EAGLE project: 7 users have made more than 1000 edits, which could be anything above the figure; 20 have made between 100 and 1000 edits; 71 have made between 10 and 100, and even more interestingly 581 have made between 1 and 10 edits. This is a critical mass of active users, uploading, editing, curating, using data they are interested in. The same situation can be noted for the 11.000 translations in the EAGLE Media Wiki, which were viewed in 18 months 1.380.000 times and have seen 280 active users, who have at least made 1 edit. The tool is not

---

7 https://commons.wikimedia.org/wiki/Category:Media_contributed_by_EAGLE

8 Liuzzo et al. (2014) http://www.eagle-network.eu/wiki/index.php/Main_Page

9 This can be easily monitored looking at the counter on the website at the end of each day. No better statistics are available.

10 Thanks to user:Fae and the authors of the BaGLAMa 2 tool. See https://commons.wikimedia.org/wiki/Category_talk:Media_contributed_by_EAGLE/reports and https://tools.wmflabs.org/glamtools/baglama2/#gid=148&month=201508

well known outside the EAGLE consortium and is a very small prototype, but the fact that it has already attracted such a mass of views is significant. What would happen if we gave 100.000 translations in the way we have given them to the people in the Mediawiki, completely openly? What would happen if those people contributing to images and translations were empowered to operate easily and intuitively to enter more and more data? Inscriptions will never get as many fans as we would like to, but perhaps their content and related resources would be a bit more accessible to non-initiated. This comparison confirms also that the usability of resources is measured at a different level when they are made open, and that images and translations have an undeniable higher relevance as an online resource, thus attracting interest also to the transcriptions, while this does not happen the other way around and only those who know what they are looking for will stumble upon an ancient inscription published online. Nothing new in these observations, but this obvious observation is in contrast with the actual situation in which photos are few and translation even fewer. Why so, then? The possible reasons are:

1. the lack of an entry point which is easy to access and use
2. people get easily worried by copyright due diligence and find diffucult sometime to trak back who is the author or the copyright owner of a photo
3. a lack of coordinated effort, planning and management of the storage of both these types of content
4. researchers working on inscriptions identify their intended audience in a very specific academic community which does not need translations and instead needs edited texts (transcriptions and metadata).
5. publication of content with (sometimes unnecessary) restrictions
6. lack of time for this effort, unrecognized in academic settings as a contribution to the progress of knowledge, as, sadly, most other digital efforts

We shall point out what has been done to solve these problems and cater for an improvement in this part of documentation and production of online content in the future.

The international group of partners, which includes University of Heidelberg, University of Cardiff, Duke University and Tufts University is holding regular workshop meetings to design and develop a suite of resources that support generation of epigraphic translations, with peer-review and publication workflows supported by Perseids' extension of the Son of the Suda On Line code (SoSOL), with publication supported by the EAGLE Mediawiki, and image management, reference ontology, geo- and other services, supported by Integrating Digital Epigraphies, and with Attic Inscription Online translations as the key content stream for development and testing.

3. Translations

To face these challenges with regard to translations, it is the opinion of the project team that tying together existing resources is a better way to tackle the issues rather than trying to superimpose a new tool or system. The available building blocks for such systematization of existing resources are the following:

- the existing local data entry point of Attic Inscriptions Online, in the process of moving to TEI-EpiDoc markup for the underlying data
- the EAGLE Mediawiki, with the Wikibase Extension, which collects translations from several users as a part of the work of the EAGLE consortium to bring epigraphy to a wider public
- the Perseids peer review system,[11] which uses the Son of the Suda Online[12]
- Leiden+,[13] a simplified markup which allows the use of normal diacritics instead of tags to enter XML markup.
- identification and disambiguation done by content providers (the epigraphic databases) and by projects such as Trismegistos[14] for

[11] http://sites.tufts.edu/perseids/

[12] Baumann (2013)

[13] http://papyri.info/docs/about and J. Sosin presentation at http://www.stoa.org/archives/1263

[14] http://www.trismegistos.org/

the members of the EAGLE consortium and IDEs[15] for Greek Epigraphy projects.

- referencing and resolution services provided by IDEs which do not just align content relating to one resource but describe the relation among them

These consider two kinds of users:

- users involved in a project with access to a data entry point in a database (using XML)
- independent contributors

With the available building blocks what can be done currently for translations is:

- a standalone javascript library to enter translations using Leiden+ (to be implemented and tested in AIO, as the best candidate for its focus on translations) which implies:
  - an enlargement of the encoding guidance and conversion capabilities of Leiden+ to EpiDoc for translations.[16]
  - recommendation on how to mark up translations for the EpiDoc guidelines

  These developments will hopefully be beneficial to the EpiDoc users community as well, which has in the past asked for more guidance on how to encode translations.
- An export of AIO in EpiDoc to the Perseids platform in which translations will be peer-reviewed for ingestion in the EAGLE wiki, bypassing the harvest process for EAGLE. This might be useful as a use case for future project willing to publish their translations with the others collected into the EAGLE Media Wiki.
- facilitate flow between existing tools and services
  - EAGLE and Perseids worked together in the past years in order to integrate the two services offered, but this had a

[15] http://blogs.library.duke.edu/dcthree/projects/

[16] https://github.com/TIGLIOPROJECT/documentation/wiki

number of limitations, e.g. the requirement for a translation to be already present in the wiki in order to be able to publish another one via Perseids. This reduced the number of items for which the integration could be used to a minimum, forcing the use of workarounds as mock text and placeholders. The integration of the EAGLE wiki with Perseids will now enable users to enter translations for any inscriptions and even a new translation from scratch with a specific new language, thus covering all possibilities for the EAGLE Mediwiki user. This requires nevertheless:

  * unique identifiers for Latin and Greek texts, which are currently provided for the first by Trismegistos, and for the second group of texts by IDEs.
  * a citation URN structure which is agreed upon and otherwise usable. This will be based on the scheme already in use for EAGLE built on CTS URN syntax as in the example:

    urn:cts:pdlepi:eagle.tm12345.perseids-translation-1

    where the structure is

    urn:cts:namespace:textgroup.work.version

– The complete workflow from data entry to publication in a website and to the common EAGLE resource via Perseids will then be a complete and replicable workflow, scalable for use from larger projects and documented to guarantee easy and sensible connection of the resources online.

The workflow for the connection of new translations and images to existing online content will be then facilitated in this way. A project with its own data entry interface should be able to use the javascript library to enter translations using Leiden+ and following the conventions already extended and public in papyri.info. They should then be able to identify with a TM or IDEs id these translations and infer a URN to push these directly to the Perseids system. Here the translations will be peer-reviewed and then sent both back to the source with an approved status and to the EAGLE Media Wiki. If this database is partner of the EAGLE project the texts will be harvested separately and the translations linked back from the Media Wiki. From the Media Wiki API they will also be available as such to external users. An independent contributor instead will be able to look up the TM text id or IDest of the inscriptions

he wants to translate and enter it to the EAGLE Media Wiki. From here this will be sent to the Perseids system and returned as described above.

The short term goal is therefore to stitch together existing resources already in development, end especially to provide ids and a clear citation syntax for all available inscriptions, which will have counter benefits also for any other digital project with these requirements.

4. Images

Most of the existing images of inscriptions are currently safely stored in private computers. Large collections of images of inscriptions are available at the major databases, and can count up to thousands of photos of inscriptions or drawings. Large collections of images are also on Flickr (e.g. Visible Words) and Wikimedia Commons (CIL and AE categories provide a good overview of what is available).

The major problem preventing publication is that people do not know if they can share the images which they have. The copyright regulations are too complicated and possible contributors opt for doing nothing instead of taking any unknown and unwanted risk. Storage of images of inscriptions in the major databases happens under very strict conditions of reuse and publication and while it is the best possible way to operate for these projects, there can be no mirroring or distribution of the resources available so that the life expectancies of content curated for decades is tied to the lifetime of these databases. A test has been done with the images of non identified items in the Epigraphic Database Heidelberg, uploaded to Wikimedia Commons and users have contributed to identify a number of those. Uploading photos on Wikimedia Commons requires an open license, which cannot always be guaranteed, whereas on Flickr it is possible to retain rights whilst publishing the photos, so that it makes a nicer tool for this kind of content, although it does not allow community editing and batch upload, which is instead possible via tools developed by Europeana and the Wikimedia Foundation for Commons.[17] A unique repository or a connection hub for all these photos would be a solution to keep the content under sight but this would still require control over resources daily published in Flickr and Commons, an involvement into the communities of users online, together with the extension of citation structures to these groups. There

[17] The GLAM Wiki toolset, https://commons.wikimedia.org/wiki/Commons:GLAMwiki_Toolset_Project.

is no easy solution for the copyright side of the problem, but a continued encouragement to share will build towards the critical mass needed for a change in perspective on this issues in the coming few years. During this project we will try to list requirements for a tool to ease out the upload of images online, which will match images and metadata provided in various formats, suggesting ids and adding them in a format compliant to the citation scheme agreed. It is in fact true that the amount of time required to use tools which ask for a one to one upload of images is another important factor which reduces the amount of content shared even from those willing to do so.

5. User involvement and expert sourcing

The problems of all digital projects looking at putting together materials from various sources and contributors are on one side to get people involved, on the other to overlook the work done and take care of the administration. The possible user scenarios are eventually many more. Two especially deserve to be mentioned here. Most input has been provided in term of new translations in the EAGLE Media Wiki during Workshops and Secondary Schools class work. Some teachers with a background in epigraphy have been contacted by the Epigraphic Database Rome to start and experiment with a didactic model which would include translating inscriptions. The students worked on a specific corpus of inscriptions, studied the text and the support and produced a translation which they entered in the EAGLE Media Wiki with the supervision of their teacher. This experience has proved successful for the students which have seen their contribution directly where it should be, together with other scholarly content. On the other side, every new translation matters in such a state as the one described above. The second example is the work done in two consecutive workshops, held in Ercolano and at the Centre for Hellenic Studies in Washington DC of the Ancient Graffiti Project,[18] which has published multiple translations for all the graffiti of Herculaneum. In this context the usability of the Wikibase software has been tested and it proved to be an extremely intuitive and powerful tool. It takes very little explanation, but there are caveats for this simplicity and namely that it is extremely easy to do things in slightly creative ways, as entering statements as source information or typing an id in a

[18] http://ancientgraffiti.wlu.edu/

slightly different way, which then need to be monitored and fixed.

6. Conclusions

Some of the tasks above have been already carried out, some are under way, but there are some general conclusions which can be summarized. There is a need to unify the resources, agree on standards for reference and citation and provide stable identifiers and citation structures, providing a comprehensive list of epigraphic publications with the relevant abbreviation in use. Although some work has been done, the amount of data makes this task continuously needed together with that of disambiguation. Other efforts need to go in the direction of flexible but harmonized standards for encoding and working on data entry giving priorities probably in a slightly different way as before, updating the tools to be able to cope. More generally, while tools are abundant and so are guidelines and cookbooks, an agreed venue for coordination of the efforts is still a desideratum, and should include not only scholarly project but also community based efforts such as those of the Wikimedia Commons users and of the Flickr user's groups. These people could be also part of the peer review process, thus contrasting the side effects of an inactive board.[19]

Aknowledgements

The Andrew W. Mellon Foundation has funded the meetings of the project TIGLIO and the work which will be carried out for Attic Inscriptions Online. We would like to thank also all the Wikimedia users who have contributed to parts of this work in various ways and especially Aubrey, CristianCantoro, Fæ, Laurentius, Magnus Manske, Sannita and Wittylama. Another special aknowledgement goes to all users of the EAGLE Mediawiki and to all the volunteers working on photos of inscriptions online.

References

Baumann, R., *The son of suda on-line*, in S. Dunn, S. Mahony (ed.), *The Digital Classicist 2013*, The Institute of Classical Studies University of

[19] The last meeting of the project, dealing with images and CTS URN structure still had to take place at the date of submission of the present contribution.

London, 2013, ISBN 978-1-905670-49-9, URL http://ryanfb.github.io/papers-BICS/sosol-bics-draft.pdf.

Lambert, S., McCourt, F., *Attic inscriptions online*, in S. Orlandi, V. Casarosa, R. Santucci, P.M. Liuzzo (ed.), *Information Technologies for Epigraphy and Cultural Heritage*, 155–166, Sapienza Università Editrice, 2014, ISBN 978-88-98533-42-8, doi:10.13133/978-88-98533-42-8, URL http://www.eagle-network.eu/wp-content/uploads/2015/01/Paris-Conference-Proceedings.pdf.

Liuzzo, P.M., Zanni, A., Martinelli, L., Losa, L., De Nicolao, P., *The eagle mediawiki. a fully collaborative database for academics, data engineers and the general public*, in S. Orlandi, V. Casarosa, R. Santucci, P.M. Liuzzo (ed.), *Information Technologies for Epigraphy and Cultural Heritage*, 187–200, Sapienza Università Editrice, 2014, ISBN 978-88-98533-42-8, doi:10.13133/978-88-98533-42-8, URL http://www.eagle-network.eu/wp-content/uploads/2015/01/Paris-Conference-Proceedings.pdf.

Panciera, S., *What Is an Inscription? Problems of Definition and Identity of an Historical Source*, 183 (2012), 1–10, URL http://www.digitalmeetsculture.net/wp-content/uploads/2013/10/Panciera-Inscription-ZPE-2012.pdf.

Part III

Users, Epigraphy and the social web

Part IV

Digital Approaches to cross-disciplinary studies of inscriptions

Romans 1 by 1. Documenting a population database for the Roman world

Rada Varga[1]

Babeș-Bolyai University, Cluj-Napoca

Abstract

The present article briefly documents the epigraphic division of a developing online population database for the Roman Empire, accessible at www.romans1by1.com. The paper presents the motivation of constructing the database and its envisioned architecture, in relation with the various sources, while emphasizing on the steps and procedures required in order to transpose epigraphical information into an ancient population database.

Keywords: Epigraphy, middle classes, provincial society, occupational inscriptions, prosopography

Romans 1 by 1 database and afferent website were created for filling in an existing gap in the study of Roman-era population. The database tries to begin answering to the need of properly cataloguing all attested inhabitants of the Roman Empire. Of course, this is a tremendous task and www.romans1by1.com is only a first step, circumscribed for now to a very specific category of sources.

1. Motivation

But why would an ancient population database be essential? Because a digital resource focused on individuals would reveal linkage possibilities that otherwise elude us, it would finally give us the complete and accurate image of the Roman attested population and, through codifications, it would open the way towards computer-assisted in-depth analyses on all relevant aspects imagined (epigraphic patterns, religiosity, migrations, onomastics, occupations, family data, etc.). A complete,

* Corresponding author. Email: radavarga@gmail.com

aggregate database will allow a longitudinal (diachronic) view on the attested Roman population from a certain area and ideally from the whole Empire, while also opening transversal (a section in time) perspectives.

The database in itself would have three components (see below Fig. 1), following the best practices in the field (Mandemakers and Dillon (2004)): the sources database (with "facsimile" transcription of the sources' text), the central database (the complete, corrected, integrated, standardized and coded form of the sources database) and the data releases (destined for on-line usage, allowing easy extraction of data in view of analyses).

Alongside these components, which actually represent various steps in the data preparation process, a population database for the Roman Empire should be built on three pillars, which we might call units or divisions of the database: epigraphic, literary and archaeological, each of them requiring different expertise, different approaches and different standards for the individual recording forms. In the end, of course, all three types of individual records will have to be integrated in a central standardized component, whose structure is to be developed by merging the three aforementioned divisions; consequently, its configuration will ideally take shape only after all three units enjoy a stable architecture at least in the sources component of the database.

The construction of the database will follow a series of steps imposed by best practices: creating a repository of sources; introducing, integrating, standardizing, coding and storing the information; enriching and disseminating the information. The codifications hold an essential part, not only in the individual linkage procedures, but in the analysing process as well. At the point when the database will comprise enough data, properly recorded and with all codifications undertaken, the usage of statistical software in order to identify trends and run comparisons over large scale geographical and administrative units might result in a better than ever understanding of Rome's social history.

2. The sources

Within this theoretical frame, whose amplitude implies a gradual and long-term approach, we have chosen to start by building the epigraphical division of the database. Thus, we began developing a project on the middle classes of the Low Danube provinces and their epigraphic attestations. The database created in this context together with the platform

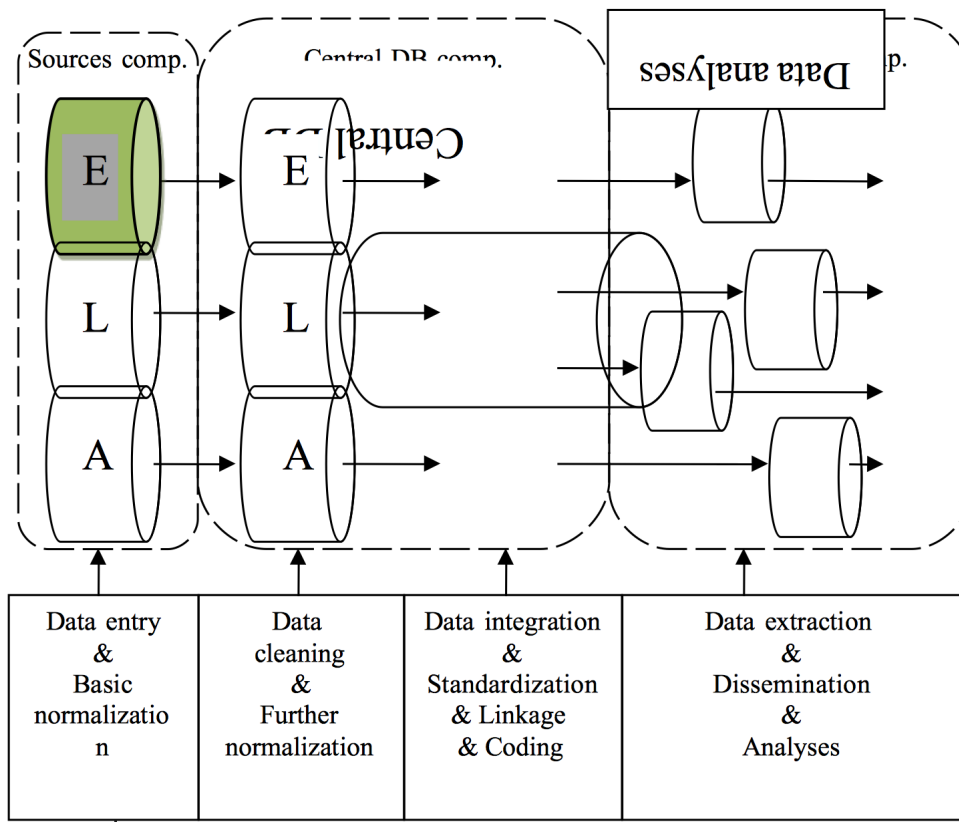

Fig. 1. Schema of the database

www.romans1by1.com represent the skeleton of the future population database of the Roman Empire.

For now, we are solely focusing on characters attested epigraphically – thus on inscriptions as sources. As constructing a metadata suitable for all social and professional categories of the provincial world is very complex, our database is created for accommodating all information epigraphically provided on members of the provincial middle classes. Terminologically, we consider all those who manifest themselves epigraphically, without being members of the imperial or provincial elite, as member of the middle socio-economical layers. Likewise, we have excluded active militaries (but not their families), as their appurtenance to the army creates their social status.

Regarding the data entry, we are trying to remain faithful to the source and to record, during the first phase, only the minimum of deduced information (gender, juridical status, ethnicity of the name). At the same time, we operate some conventional onomastic transcriptions (AEL will be Aelius from the start, etc.). All these basic normalization procedures are being thoroughly explained and documented in the data entry manual, and the reason why we chose to apply them is to speed up data standardization, run analyses and publish results on small samples, whose standardization procedures would otherwise be overwhelmingly time-consuming in relation with the results.

3. Database architecture

The core of the database's epigraphical division is formed by a table used for recording personal individual data (labelled Personal data), around which the entire network of components needed to ensure proper information recording is built. However, when starting data entry, one must begin by recording information about the source in use.

The first table to be filled in is the file of the source – Inscriptions. First of all, each inscription gets a source code, formed of 5 digits and a symbol/acronym of the province's name (D for Dacia, MS for Moesia Superior, DAL for Dalmatia, etc.) – so we have, for example 00001MS. The Inscriptions table has text fields, as well as value lists. Then, we have text strings for: Relevant expressions, Stylistic details, Atypical features, Observations, Place of discovery, Place of provenience, Ancient name provenience, Timestamp/Timeframe and External links. Although we are aware that some of the data (Timestamp/Timeframe) could have

benefitted from a standardized form, at this point we opted for vaster possibilities of expression and adaptation. Other information will be filled using value lists, as standardization is more suited: Type of inscription, Language, Material. For these we will use the Eagle vocabularies,[1] because they offer an already standardized language.

The Coordinates table links each inscription with the latitude and longitude of its place of provenience, in order to place it on a map.

The Inscription bibliography was conceived so that extracting complete or selective bibliographical lists would be possible. Thus, a normalization table includes all bibliographical titles referred to and from it; through a value list, one can select the Bibliography abbreviation. The exact reference is presented and detailed in the Details and Comments text strings. Of course, all data are linked to the Inscription code, selected as well from a value list.

Only after properly documenting the source one can access the main and most complex table, Personal data, where basically the individual file of each person is created. At this point, each new entry represents a singular epigraphic attestation of an individual, and a unique ID is generated, which will help link the respective character throughout the various components of the database and with other database entries. The person is also manually linked to the source using a value list of the inscriptions' codes. In the event of one person being attested in multiple epigraphs, each attestation will represent a new entry, and it will be assigned a new unique ID, which will be doubled, during linkage procedures, by a common ID for all instances of the same person.

The Personal data table was built to host a large variety of information offered by epigraphs, and its structure has proved, up to this day, rather stable. However, if need be, it can always be extended in order to accommodate any new kind of information. A first set of its variables are name-connected: Praenomen, Nomen, Cognomen/Personal name, Father/Master name, Agnomen, Signum, where each also has a drop-down associated for Ethnicity and the Agnomen and Signum an Observations string. While we believe the possibilities of actually identifying signa for members of the middle classes, during the Principate period, is rather reduced, we opted for facilitating their correct registration, in case they are discovered. Other data regard Natione, Ethnicity, Origo and Domus – if and how they are mentioned in the inscription. As

[1] http://www.eagle-network.eu/resources/vocabularies

acknowledged above, some information will be recorded, even if they are deductive and not literally written in the source: Gender and Juridical status (though the servile one often is literally recorded). For the latter, we have opted for a check box, which, if checked, opens all the available possibilities. The rest of the fields accommodate supplementary information, when and how there is the case: Occupation, Collegium, Deities, Age (at death), Details of life/death and Observations.

The Occupation field requires some special attention, as it has the Occupation code associated with it; as we are trying to propose a codification of Roman occupations/professions, based on and adapting the HISCO classification model[2]. While a raw classification and codification based on HISCO might only be a slight challenge, finding a theoretical model might prove to be a more elaborated task. What we aim at is constructing the "metadata" for the codification of Roman professions attested on stone and analyse how deep the classification can realistically go. Much alike other normalization fields present in the sources component of the database, Occupation code was implemented because it helps dealing faster with small and medium size samples (up to hundreds of people) in view of publishing preliminary results which are vital for dissemination and further financial support of the project.

Two important variables, for statistics and working with the data, are Dedicated by and Dedicated for, which state the relation of the recorded individual with the epigraph and with other persons mentioned by it. Both appear in the simple form of check boxes.

Another particular check box, in need of supplementary explanations, is Later. The database was conceived for attested civilians from the middle classes, but sometimes they are associated in inscriptions with militaries or representatives of the elites. In this case, we have to register a minimum of data on the later as well, in order to build a wider image of the characters that we focus on. When checking Later, it opens Status, which at its turn opens the following options in the form of check boxes: Senator, Knight, Local magistrate, Decurion, Imperial priest, Imperial slave, Imperial freedman, Military personnel. If checked, each of these boxes opens a Details text box; additionally, Local magistrate and Decurion open a City/Town value list, while Military personnel, opens Rank and Unit value lists. While elite members and military personnel will at some point be added to the database, for the currently running

[2] http://historyofwork.iisg.nl/

project purposes, their social and professional status, together with a minimum of relevant details and relation with the recorded individuals represent enough information.

Field label	Field id	Data type
Praenomen	rperson_praenomen	Text
Ethnicity praenomen	rperson_eth_praen	Value list
Nomen	rperson_nomen	Text
Ethnicity nomen	rperson_eth_nom	Value list
Cognomen/Personal name	rperson_cognomen	Text
Ethnicity cognomen/personal name	rperson_eth_cogn	Value list
Father/Master name	rperson_father	Text
Ethnicity father/master name	rperson_eth_fth	Value list
Agnomen	rperson_agnomen	Text
Ethnicity agnomen	rperson_eth_agn	Value list
Observations agnomen	rperson_remark_agn	Text
Signum	rperson_signum	Text
Ethnicity signum	rperson_eth_sign	Value list
Observations signum	rperson_remark_sign	Text
Natione	rperson_nation	Text
Ethnicity	rperson_nat_ethnicity	Text
Gender	rperson_rgender_id	Value list
Juridical status	rperson_jstatus	Checkbox. Opens checkboxes: Citizen/Libertus/Peregrine/Slave/Veteranus (opens Unit/Rank - valuelists)
Tribus	rperson_rtribus_id	Text
Origo	rperson_origo	Text
Domus	rperson_domus	Text
Collegium	rperson_collegium	Text
Occupation	rperson_occupation	Text
Occupation code	rperson_occ_code	Text
Deities	rperson_deity	Text
Age	rperson_age	Text

Details of life/death	rperson_death_details	Text
Dedicated for	rperson_dedicated_for	Checkbox
Dedicated by	rperson_dedicated_by	Checkbox
Inscription code	rperson_rinscription_id	Value list
Later	rperson_now_stat	Checkbox. Opens checkbox: Status
Status	rperson_status	Checkbox. Opens checkboxes: Senator (opens Details – text)/ Knight (opens Details – text)/ Local Magistrate (opens Details – text, City – value list)/ Decurion (opens Details – text, City – value list)/ Imperial priest (opens Details – text)/ Imperial slave (opens Details – text)/ Imperial freedman (opens Details – text)/ Military personnel (opens Details – text, Unit/Rank – value lists)

Observations	rperson_observation	Text

Based on the personal ID given to each individual, the Relationship table will solely name the relationship between individuals (A to B and B to A), choosing from a drop-down menu. The relationships have been encoded from the start, in order to make processing quicker; thus first-degree relationships have 10-codes (101-Husband, 102-Wife), second and third degree relations 20- and respectively 30-codes, non-family relations were given 40- codes and 50-s for the unspecified/unreadable relations. Male relations were given odd numbers and female ones – even numbers.

4. Conclusions

Romans 1 by 1 is a first step towards a comprehensive and exhaustive electronic resource for the attested population of the Roman Empire. The following normal steps for expanding and enriching the database is elaborating the fitted metadata for provincial elites and military personnel epigraphically attested.

Acknowledgement

This work began being developed during the period of a residential scholarship at the Hardt Fondation, Vandoeuvre. Its expansion to the area of the whole Latin language Empire is being undertaken with the support of a postdoctoral fellowship of the Fritz Thyssen Stiftung. Also important was the financial support of a POSDRU scholarship, granted by the Babeș-Bolyai University.

This work began being developed during the period of a residential scholarship at the Hardt Fondation, Vandoeuvre. Its expansion to the area of the whole Latin language Empire is being undertaken with the support of a postdoctoral fellowship of the Fritz Thyssen Stiftung.

References

MANDEMAKERS, K., DILLON, L., *Best practices with large databases on historical populations*, Historical Methods: A Journal of Quantitative and Interdisciplinary History, 37 (2004), 34–38, doi:10.3200/HMTS.37.1.34-38, URL http://dx.doi.org/10.3200/HMTS.37.1.34-38.

Part V

Posters

Author's index