DIGITAL CLASSICS ONLINE

Table of contents

A Selection of Papers from the Digital Classicist Seminar Berlin (2012-2015)

Matteo Romanello, Martina Trognitz, Undine Lieberwirth,
Francesco Mambrini, Felix Schäfer

Introduction

The Digital Classicist seminar series was established in 2006 in London as a platform for discussion aimed at bringing together people working at the intersection between Classics and computing science. This seminar format has found a fertile ground also in Berlin – a city warming with activities related to the Digital Humanities (DH) and Digital Archaeology – where a comparable series was launched in 2012. Since then, more than forty papers were presented, a selection of which is published in this special issue.

In the context of this volume the term 'Digital Classics' is a placeholder for a much broader range of disciplines. 'Digital Classics' should not be misunderstood as a simple transfer of an existing subject into the digital realm, because this does not do justice to the possibilities provided by digital methods. Rather it should be understood as wide and open as possible.

Thus the scope of this special issue is the ancient world at large, covering not only the usual research areas of Classics and 'Altertumswissenschaften', but including also for example Babylonia, the Near East and China. The selected papers highlight inter- and cross-disciplinary approaches, which show how the digital component can lead to overcome disciplinary boundaries and to establish closer connections among philology, archaeology, ancient history and information science. Moreover, the contributions focussing on texts are not limited to classical Greek and Latin literature, but deal as well with the Indian, Ancient Egyptian and Arabic literary tradition.

The Digital Classicist Seminar Berlin (henceforth DCSB) is informed by the ideals of de-centralisation, international collaboration and openness that characterise the Digital Classicist community as a whole.[1] As such, the seminar series is organised by a group of researchers belonging to various key institutions in Berlin – the excellence cluster TOPOI, the German Archaeological Institute (DAI) and the Berlin-Brandenburg Academy of Sciences and Humanities (BBAW) – and is made possible by the generous financial support of these institutions, as well as of DARIAH-DE. As part of our commitment to openness, the videos and slides of the seminar, and also the publications originating from it, are made available under a Creative Commons licence.[2] The open publication of the (seminar) outputs aims at

---

1   The description on the Digital Classicist website reads: "The Digital Classicist is a decentralised and international community of scholars and students interested in the application of innovative digital methods and technologies to research on the ancient world. The Digital Classicist is not hard-funded, nor owned by any institution", <http://www.digitalclassicist.org/index.html> (15/03/2017).

2   Abstract, slides and videos of each seminar are accessible from the DCSB website, <http://de.digitalclassicist.org/berlin> (15/03/2017).

making them as accessible as possible and potentially allows for reuse of these materials in other contexts (e.g. as learning and teaching resources).[3]

Since 2014 the DCSB is also part of the academic programme of the Freie Universität under the name "Digitale Methoden in den Altertumswissenschaften" and in this way contributes to establish in Berlin a DH curriculum with a specific focus on the ancient world. In addition to gaining familiarity with a broad array of methods for the study of the past, students can put into practice and improve their communication skills. This is achieved through the choice of English as official language for the seminars and a final presentation. This presentation consists of preparing a poster – a medium of essential importance for scholarly communication – on one of the topics that were touched upon during the seminar. It turned out that not only preparing a poster is a very useful soft skill for students to acquire (and rarely taught at undergraduate level) but it also helps students to engage more actively and more in-depth with the subject.

As for the organisation of materials, the wide variety of topics covered and methods employed by papers in this issue – which well reflects the nature of contributions that are presented at our seminar – made it impossible to arrange its content thematically. Therefore, we decided to present the papers going from the general to the particular, and to arrange them into three somewhat coherent blocks.

The first block features two methodological reflections on digital media, and on the effects they have on our scholarly practices. In the first contribution Haciguzeller focusses on cartography and collaborative mapping. After putting collaborative mapping into the context, she presents four possible scenarios for its use in archaeology, namely psychogeography, local community involvement, online collaborative mapping for archaeology and spatial narratives. In the second contribution, Trachsel considers how the philological work entailed by editing fragmentary texts (i.e. works that are known to us only through quotations contained in other texts) changes when moved to a digital environment. In particular, Trachsel reflects on how the new possibilities for the presentation of textual content that are offered by the digital medium, as opposed to printed editions, may lead us to a new understanding of what a fragment is and how it can (or should) be presented to the reader.

The second block gathers contributions about research projects that lead to the development of digital tools or services, which can be seen as individual components of an emerging digital infrastructure for the study of our past.

The papers by Bodard et al. and Varga form a prosopographical dyptich. The former presents the SNAP:DRGN project, where the acronym stands for the extended project's name *Standards for Networking Ancient Prosopographies: Data and Relations in Greco-Roman Names*. SNAP:DRGN aimed at building a virtual authority list for ancient people through Linked Data collection of common information from existing resources. On the basis of three case studies the authors discuss the various challenges that are encountered when prosopographical data is merged into a single reference system. Some of these issues are due to the peculiarities and the fragmentary status of ancient sources, while others are related to technical aspects.

The latter paper, while remaining in the realm of prosopography, focusses on one specific geographical region, the Roman provinces of Dacia and Moesia. In this paper Varga discusses two aspects of the *Romans 1 by 1* project. First, she describes the structure of the

---

3   The first publication originated from this seminar series is Bodard & Romanello (2016).

prosopographical database being built in the project and then explains how its structure allows for asking complex questions about the Roman population.

In the last paper of this block Henny et al. take into consideration the whole lifecycle of a digital resource – the digital edition of the *Book of the Dead* – and reflect on the experiences made and the lessons learned with regards to the various stages of a digital resource's lifecycle. The authors address important issues, which are not often enough discussed in Digital Humanities scholarship, such as the costs of maintaining and continuously curating a platform like the Book of the Dead that is intended to be a living digital resource.

Finally, the third block consists of papers presenting the results of explorative research, conducted by using various methods ranging from imaging techniques to visualisation to network analysis.

Piquette describes the setup and results of capturing unrolled Herculaneum Papyrus fragments with Reflectance Transformation Imaging (RTI) and Near-infrared RTI (NIR RTI). Previous to this the fragments have been documented with NIR photography, which produced clear, but static images. The main advantage of RTI is that the resulting images can be relit digitally from any desired position, thus making it possible to discern holes and distinguish different papyrus layers, and to increase the contrast between ink and writing surface.

In his paper Roeder explores the use of various visualisation methods in order to further our understanding of the linguistics transformations occurring when ancient Greek is translated into Classical Arabic. He presents and discusses various visualisation examples drawn from the *Glossarium Graeco-Arabicum*, a parallel corpus of about 100,000 words.

Elwert et al. present the results of applying social network analysis methods on ancient Egyptian and Indian texts. They demonstrate how these methods can provide insights into the relations existing between different protagonists, their possible grouping into communities and the identification of the most important relations. Among the examples discussed by the authors there is also a plot network representing the main content and the sidelines of "the contendings of Horus and Seth", which proved to be a useful resource for distant reading.

Finally, Da Vela examines the use of social network analysis to understand how local communities in Hellenistic Northern Etruria were transformed during the process of Romanisation, especially with regards to their perceived cultural identity. Her proposed approach tries to relate identity markers detected from archaeological and epigraphic sources with the diachronic evolutions in the structure of the social network between the local communities.

Bibliography

Bodard & Romanello (2016): Gabriel Bodard & Matteo Romanello (Eds.), Digital Classics Outside the Echo-Chamber, London, DOI: 10.5334/bat.

Authors[4]

Undine Lieberwirth M.A.
Exzellenzcluster 264 – TOPOI
FORUM Spatial Data
Freie Universität Berlin
Hittorfstr. 18
14195 Berlin
http://www.topoi.org

Email: undine.lieberwirth@fu-berlin.de

Dr. Francesco Mambrini
Deutsches Archäologisches Institut
IT-Referat
Podbielskiallee 69–71
14195 Berlin
http://www.dainst.de

Email: francesco.mambrini@dainst.de

Dr. Matteo Romanello
Deutsches Archäologisches Institut
IT-Referat
Podbielskiallee 69–71
14195 Berlin
http://www.dainst.de

Email: matteo.romanello@gmail.com

Dr. Felix Schäfer
Deutsches Archäologisches Institut
IT-Referat
IANUS Forschungsdatenzentrum
Podbielskiallee 69–71
14195 Berlin
http://www.ianus-fdz.de

Email: felix.schaefer@dainst.de

Martina Trognitz M.A.
Deutsches Archäologisches Institut
IT-Referat
IANUS Forschungsdatenzentrum
Podbielskiallee 69–71
14195 Berlin
http://www.ianus-fdz.de

Email: martina.trognitz@dainst.de

Collaborative Mapping in the Age of Ubiquitous Internet: An Archaeological Perspective

Piraye Hacıgüzeller

Abstract: Over the past decade, collaborative mapping projects have become widespread, allowing for and promoting voluntary participation in cartographic processes. A major factor in the increasing popularity of collaborative mapping in recent years has been the developments in digital cartographic media in general and internet mapping in particular. In this paper the aim is to discuss the possibilities of online collaborative mapping in archaeology. Following an overview of collaborative mapping and its current state in today's increasingly online and digital world, four potential modes are introduced through which collaborative mapping in archaeology can be carried out: psychogeography, local community involvement, an online archaeology map system and spatial narratives.

1. Introduction

In the last ten years or so, collaborative mapping projects have become popular, enabling and promoting voluntary participation in cartographic processes. Such projects are considered to be inherently emancipatory and pluralistic since they typically allow different groups of people to collaborate equally in a cartographic process.[1] Unlike traditional cartographic projects, collaborative mapping involves more than one person with the power to decide what to put on or exclude from the map. Therefore, archaeological collaborative mapping potentially provides a new set of practices to render archaeology more prolific in the sense of being more multivocal, emancipatory and open to multiple realities.

A major factor in the increasing popularity of collaborative mapping in recent years has been the developments in digital cartographic media in general and internet mapping in particular. Specifically, the user-friendly and efficient nature of digital cartographic media has democratized mapping processes since the 1990s.[2] Such media serve to capture, store, manipulate, analyze and display cartographic information not only by professional cartographers but also others, thereby challenging the authority of professionals in the cartographic process.[3] In the case of internet mapping, many online collaborative mapping platforms (e.g. OpenStreetMap, Wikimapia and Google Earth) use a revolutionary online mapping logic today, mapping 2.0., which provides the means for the visitors of these platforms to participate in the online mapping processes rather than being passive users of the information presented.[4]

In this paper the aim is to introduce the concept of collaborative mapping to archaeologists with the possibilities provided by digital media and internet. As recently highlighted, collaborative

---

1   Grasseni (2004); Perkins (2014); Sletto (2009).

2   Rød et al. (2001).

3   Dodge (2011).

4   Perkins (2014).

mapping „could be a powerful tool for investigating an archaeology of the present that has political and social meaning; an accessible and richer archaeology that allows everyone to meaningfully contribute".[5] Such mapping projects can indeed challenge established power structures in archaeology and sustain multivocality[6] as well as helping the management of archaeological projects in ways tuned into the concerns, rights and interests of local population.[7] With the increased accessibility of cartographic information and processes provided by digital tools and internet, the stakes are even higher for archaeological collaborative mapping today. The task for archaeologists now not only involves exploring the promises of collaborative mapping for archaeology, which is long overdue, but also to reflect on these promises in relation to the digital and online transition in cartography.

2. Online Collaborative Mapping: An Overview

Western cartography is known by its historians and theoreticians as a practice that historically serves the intentions of powerful individuals and institutes with, for instance, militarist, capitalist, colonialist or nationalist interests.[8] Geographer Brian Harley[9] suggests for the case of colonialism that „[A]s much as guns and warships, maps have been the weapons of imperialism". As elaborated by Kitchin and others[10], Western mapping has indeed played an instrumental role:

> „in imperial exploitation through the erasure of indigenous peoples from the coloni-sers' maps ... In the partition of India, the annexation of Palestinian land, or the ‚*terra nullius*' of Australia, cartography has been integral to colonial practices, providing both spatial justification and a rationalising tool for colonisers, past and present".

Being such a powerful practice, it is not surprising that Western mapping has remained an elite enterprise throughout modern times.[11] It was not before the 1990s that cartographic skills became considerably easier to acquire (a phenomenon referred to as „democratisation of cartography"[12]) with the coming of age of digital mapping and later on the internet providing increasingly easy access to mapping tools and relevant data as well as the media to circulate maps and cartographic information.[13] This democratization process has improved further in the 1990s in response to the „maps and power" critique within the critical cartography and critical GIS literature.[14] In sum, collaborative mapping practices have their roots in the fertile intellectual soil of critical cartography and critical GIS that started to gain depth in the early 1990s; these practices got further fed in the same period by technical developments in digital computing and computer network science.

---

5   Lee (2016).

6   See Conkey / Gero (1997), 429; Joyce / Tringham (2007).

7   Atalay (2008); Duke / Saitta (1998).

8   Kitchin et al. (2011).

9   Harley (1988), 282.

10   Kitchin et al. (2011), 389.

11   Jacob (2006).

12   Rød et al. (2001).

13   Dodge et al. (2011).

14   Crampton (2010).

Collaborative mapping started to materialize in the 1980s, with guides explaining how to carry out such mapping projects[15] and the initiation of projects such as the Parish Maps Project (see below). In the years that followed, collaborative mapping was practiced widely especially in the context of advancing territorial and cultural claims of indigenous and local populations, highlighting their knowledge, and designing and carrying out locally-led projects.[16] In fact, the ubiquity of indigenous and „counter" mapping projects that aim to subvert established power structures have caused collaborative mapping to be largely understood in relation to indigenous and minority rights.[17] Two examples of indigenous mapping projects are presented by Wainwright and Bryan[18] who review the collaborative mapping projects of the Maya and Mayangna communities of southern Belize and eastern Nicaragua respectively from a critical perspective. In order to avoid displacement, dispossession and destruction of their livelihoods through state-sanctioned practices, these communities created maps that express and document their territorial claims and used these maps in legal battles with the assistance of organizations such as the World Wildlife Fund (in the case of Mayanga community) and Toledo Maya Cultural Council. As Wainwright and Bryan rightly point out in their study, however, the maps created by the Maya and Mayangna communities fail to transcend typical issues of modern politics, namely territory and property rights. Collaborative mapping with political ambitions has in fact often been criticized along these lines of not managing to subvert established power structures but merely reworking them along the lines of more traditional mapping practices (see below).[19]

Participatory GIS (PGIS) (also known as Public Participation GIS or community-integrated GIS) practices have formed a significant branch within collaborative mapping in North America where GIS technology is largely accessible in the public sphere.[20] The key idea of PGIS is to turn GIS use into a more democratic practice through the active participation of people and GIS communities in the mapping process.[21] In other words, PGIS practices celebrate „the multiplicity of geographical realities rather than the disembodied, objective and technical 'solutions' which have tended to characterize many conventional GIS applications".[22] Among the successful applications of PGIS are Elwood's examination of the impact of GIS use by a Minneapolis (Minnesota) community-based neighborhood organization in urban planning and neighborhood revitalization;[23] Kyem's investigation of how innovative participatory applications of GIS can manage conflicts through a case study on forest resources in Southern Ghana;[24] and Walker et al.'s collaborative GIS project in rural Australia for community-based decision making about sustainable resource use.[25]

Many online applications of collaborative mapping today share characteristics with what is known as „Web 2.0" identified by rich and variable content, enhanced interactivity and user participation. Online manipulation of cartographic data with a participatory map 2.0 logic

---

15   E.g. Aberley (1993); King / Clifford (1985).

16   Cf. Chapin et al. (2005); Chambers (2006); Perkins (2007).

17   Parker (2006).

18   Wainwright / Bryan (2009).

19   Parker (2006); Perkins (2007), 127; Wainwright / Bryan (2009).

20   Chapin et al. (2005); Perkins (2007), 127.

21   Elwood (2006).

22   Dunn (2007), 616.

23   Elwood (2002).

24   Kyem (2004).

25   Walker et al. (2002).

is referred to as „Web mapping 2.0"[26] or „maps 2.0".[27] Gartner defines Web mapping 2.0 as „Web 2.0 applications that have a spatial frame of reference".[28] Among the examples of Web mapping 2.0 applications Gartner counts are geotagging which is geo-referencing digital objects such as photographs, videos, audio files, websites; geoblogging which is geo-referencing information in blog entries and mashups which involves combining cartographic information collaboratively on a web-served base map. While all these applications of Web mapping 2.0 can be considered online applications of collaborative mapping, Web mashup applications are worth highlighting here due to their recent popularity. A mashup application, as Gartner explains, „is a Web application that combines data from more than one source into a single integrated tool" and is composed of three parts: the content provider, the mashup site itself, and the client Web browser.[29] For the Web mapping 2.0 mashup applications the content provider (e.g. Google, Yahoo, Microsoft) would typically provide geographically referenced raster or vector data, an application programming interface making this data available (e.g. Google Maps, MS Virtual Earth, Yahoo Maps) using various Web-protocols (e.g. GeoRSS and KML) and a standardized user interface.[30] The mashup site will then facilitate mapping of spatial information fed by multiple sources onto the provided base map. The pioneering application for web mashups was called HousingMaps combining all the housing locations fitting certain criteria from Craig's List on Google Maps, an essentially collaborative mapping application.[31]

3. Online Collaborative Mapping for Archaeology

Four particular modes in which online collaborative mapping in archaeology can be carried out are discussed in this section. These modes are not mutually exclusive but each has a distinct focus: the psychogeographic exploration of archaeological places, community involvement in archaeological online collaborative mapping, the establishment of an online archaeology map system and archaeological spatial narratives.

Psychogeographic explorations of urban places and related mapping practices have its origins in the practices of the surrealism-inspired movement Situationist International (SI).[32] As one of the founders of SI, Guy Debord introduced the idea of psychogeography and related playful city strolling practices (known as *derivé* (fr.) or drift (eng.)) in the 1950s. *Drifts* involved walking in a city, often not alone but in small groups (hence its collaborative nature), letting oneself to be drawn by attractions and influenced by the ambiances of the city, all the while mapping one's personal tracks to study „the precise laws and specific effects of the geographical environment, consciously organized or not, on the emotions and behavior of individuals".[33] Since the 1990s, there has been a revival of psychogeographic urban exploration and related mapping practices

---

26   Gartner (2009).

27   Crampton (2009).

28   Gartner (2009), 71.

29   Gartner (2009), 71–72.

30   Gartner (2009), 72.

31   Ratliff (2007).

32   Cf. Perkins (2007), 128; Pinder (1996); Pinder (2005); Wood (2010a); Wood (2010b), 171–177.

33   Debord (1955) cited in Wood (2010a), 186.

in Europe and North America.[34] As Perkins[35] explains, participants in such explorations „walk the city in new [playful] ways, following algorithmic patterns (first left, second right, third left etc), solving puzzles, reclaiming places from commerce or surveillance by staged performances, navigating new routes and constructing new maps". More specifically, there are many map art projects[36] that can be discussed as forms of psychogeography today as they aim to map the emotional and sensory engagement of a group of people with a given place.[37] These art projects do not always have an online or even digital component but are nonetheless inspiring for such digital or internet-based applications. Interesting in this context, for instance, is artist Christian Nold's work in which he collaborates with a group of primary school children in Bristol (UK) to map their sensory journeys during their commute between home and school.[38] The children were provided with global positioning systems (GPS) units and asked to press a button on the unit when events that they found important occurred along their path. Each child recorded five locations as the scene of such experiences and these daily travel experiences were collected on a map together with corresponding doodles drawn by the children. In another project, Nold created communal emotion maps of urban areas (e.g. East Paris, San Francisco, Greenwich) using a GPS unit and a „bio mapping device" the artist invented himself. While the bio-mapping device measured and recorded galvanic skin responses (GSR) of the wearers (like a lie-detector), the GPS device recorded people's locations when these responses to the environment took place.[39] In the project, communal emotion maps were created by bringing together tracks and GSR of various people on a single map.

Drifting in archaeological places in a way inspired by psychogeography would open up opportunities for making new types of maps that attempt to catch ambiances of such places as well as collaborating in new types of mapping processes. However, in the today's age of digital media and ubiquitous internet, there are many more possibilities for archaeological practices to expand the psychogeographic collaborative experience. For instance, thanks to new open source map renderers such as Tangram, psychogeographic maps can be created through real-time streaming of automatically geo-referenced images, audio and video data recorded during collaborative drifts at an archaeological place. Other participants around the world can participate in these digital cartographic events through the internet. What is worth noting here is that even though collaborative psychogeographic maps of archaeological places would not necessarily represent a shared vision of these places in terms of their human emotions and behavior, they would provide the opportunity to talk about them in new ways that pay attention to the personal or shared experiences.[40]

A second mode of archaeological online collaborative mapping might have the local community at its core. Community mapping in archaeology can serve to further enhance or at least experiment with the democratization process in archaeology.[41] That is, local communities living in areas of archaeological interest can get involved in mapping which then can serve as a means to have a say in the archaeological process. For instance, Daniel Lee recently carried out a community mapping project in Orkney which he refers to as a counter map „made by residents and visitors from their everyday journeys, favorite walks, island tours, encounters,

---

34   O'Rourke (2013); Pinder (2005).

35   Perkins (2007), 128.

36   See Obrist (2014); Harmon (2004); Harmon (2009).

37   See Pinder (2005).

38   Nold (2009).

39   Nold (2009); Perkins (2007), 128.

40   Nold (2009), 7.

41   Lee (2016); cf. Parker (2006).

and significant places and objects".[42] The participants were carrying a handheld GPS device or were using their smart phone with a GPS app installed to map their journey and record a „site" (heritage or non-heritage related) of their choice. The result was not a „final map" but several maps showing the selected „site" and the trajectory taken by each participant throughout the day. Participants had the option to write a summary of their day, and record the site with a photo, a sound recording and/or video to add an extra dimension to the mapping process.

While the idea of community mapping in archaeology is new, the Parish Maps Project in the UK can serve as an important source of inspiration for archaeologists in this context. The Project was launched by the charity Common Ground in 1985 to celebrate „local distinctiveness" by encouraging local people map their own parish on the basis of their own set of values rather than that of a, for instance, professional cartographer.[43] The Parish Maps Project typically avoids central standardization encouraging local people to „employ whatever skills were available to create a map of their own place" and „decide what is mapped, who is involved, how mapping should be carried out, the form of the map and its medium".[44] While this quality of Parish mapping is commendable for the considerable freedom it gives to the non-experts participating in the mapping process, digital and online community mapping projects in archaeology may require the involvement of standard media and methods because of their high technicality.

One online application that would serve community mapping projects in archaeology well are the „online mapping sessions" where participants would tag and annotate existing maps on the basis of a very specific (set of) question. Organized by a moderator, these well-defined mapping sessions can help people to think and communicate through a map and would allow recording of the valuable cartographic process through which cartographic realities are negotiated and the community map is brought about.[45] Importantly, such mapping sessions have already been carried out in non-digital environments where participants of the mapping process were divided into groups and received stickers to place on the maps in relation to the cartographic discussion subject.[46] The advantage of online collaborative mapping sessions, however, is the accessibility they provide to different parties across the world who may still not be living in or in the vicinity of the area of cartographic interest but may have the knowledge, experience and memories about the mapped place and, hence, may still be considered as part of the local community.

A third mode for collaborative mapping in archaeology could involve establishing an online Archaeology Map System acting as a global archaeological cartographic movement building on local archaeological knowledge. Much like the Green Map System energizing eco-cultural movements at different parts of the world since the 1990s,[47] a global Archaeology Map System may use common symbols and practices in order to gather locally generated information about archaeological places (e.g. surface remains, location and length of touristic routes, places of accommodation close to archaeological sites). Archaeology maps could then typically represent a variety of archaeology-related points of interest on a map.[48] An Archaeology Map System can be supported by different partners, projects organizations and grants, and maps produced can meet different local needs. As Perkins explains for the case of London's Green Map:

---

42   Lee (2016), 1.

43   Clifford (1996); Wood (2010b), 143–155.

44   Perkins (2007), 128.

45   See also Kitchin / Dodge (2007).

46   Dreessen et al. (2012).

47   Perkins (2007), 132–133.

48   Williamson / Connolly (2009), 98.

> „the mapping is employed as part of a London-wide Local Agenda 21 initiative promoting local green activity and communicating green issues, by engaging Londoners to build local sustainability and capacity for sustainable future action. This map shows: food growing projects, food coops, farmers' markets, community gardens, scrap banks, computer, furniture and white goods reuse projects, walking or cycling projects, residents [sic] groups with green approach, locations of different local groups, recycling points, health food shops, green businesses, and various council services".[49]

In the case of the Archaeology Map System, the maps can be produced by team members of archaeology projects communicating the results of the project as well as local knowledge about the area gathered from local communities. Such maps can also communicate upcoming events in relation to the archaeological project and local area and, as such, encourage and promote participation to these events by non-local and non-expert groups and individuals. Even though the maps created as a result of the Archaeology Map System may not speak to the interests of all parties engaged with archaeological places, the System would at least serve as a platform where theoretically everybody can have their say on places in relation to archaeology.

An Archaeology Map System would benefit significantly from free and open source software (FOSS).[50] Its availability at minimal or no cost, the experience of substantive online technical support for problem troubleshooting and the possibilities of sharing geospatial data easily are among the main advantages for these software. In cartography, FOSS can avoid discrimination against groups and individuals that lack resources to access commercial cartography and spatial data management software yet want to contribute to an Archaeology Map System. A fully open source cartographic portal is OpenStreetMap (OSM) which was founded in 2004 and provides free geospatial data that could facilitate archaeological collaborative mapping applications.[51] A collaborative mapping project itself, OSM provides free base maps which archaeologists can use to bring together their knowledge of archaeological places from across the world. Among the free and open GIS and spatial data management software and web-based map servers that can be useful for archaeological collaborative mapping applications are PostGIS, QGIS and GeoServer. These software packages would support the online production of, visualization of, interaction with and service of archaeological spatial information collaboratively in an Archaeology Map System. It is certainly worth noting here, however, that despite their „openness" in terms of software development and use, these software programs can be considerably inaccessible to many archaeologists due to the advanced technical knowledge required to use them. However, they can still serve as a better alternative in comparison to their commercial counterparts given that commercial software is not necessarily more accessible than FOSS in terms of technical knowledge. If anything, commercial software maybe even less accessible in many cases due to the monopoly tried to be established for user training by the profitable software development companies.

A fourth mode through which collaborative mapping in archaeology can be carried out is narratives. As Caquard[52] underlines, the relationship between maps and narratives has never been as fully embraced and fruitfully explored as it is today. One of the reasons for the strengthened ties between maps and narratives has been the critical turn in cartography since the 1980s which „has dramatically modified the relations between maps and narratives ... by deconstructing and exposing the metanarratives embedded in maps, and by envisioning

---

49   Perkins (2007), 132.

50   Crampton (2009), 93–95.

51   Perkins (2014).

52   Caquard (2011).

maps as a compelling form of storytelling".[53] Archaeologists can rely on textual narratives in collaborative mapping practices in the cases where such narratives are, due to their familiarity, a more convenient format to present an archaeological place collaboratively. It is technically possible today to automatically geo-reference place names appearing in texts, images or audio and video files through the process called geoparsing.[54] Places whose locational information is retrieved automatically from collaboratively written or told out archaeological narratives can be mapped out and such a collaborative map would be especially meaningful with its narrative companion. Another possible collaborative mapping practice on the basis of narratives can be carried out using online multimedia storyboard environments (e.g. Prezi, Chronozoom) where collaborative narrative maps of places can be created through access to the same map by multiple users through internet. Prezi zooming presentation platform is worth special attention here. It is increasingly used for telling spatial narratives in humanities thanks to the fact that it allows inserting texts, images, video and sound to the narrative with relative ease.[55]

4. Conclusions

Collaborative mapping remains an unexplored practice for archaeologists despite archaeology's clear aspirations regarding multivocality and involvement of local communities to archaeological processes. With the increasing digitalization in cartography and ubiquity of online mapping since the 1990s, promises of collaborative mapping for archaeology today is even more exciting. Digital transition in cartography made it easier and quicker to produce, manage, explore and circulate spatial information. A major impact of this transition in archaeology has been the introduction of geographical information software, which seems to have finally established itself in the discipline successfully after years of debate. Archaeological internet mapping, on the other hand, is in its early days at the moment and it is likely that the coming years for archaeologists will involve several trials and errors, and related discussions on the topic. It can only be hoped that the increasing incorporation of Web 2.0 to archaeological practices through, for instance, blogging, social media and wiki entries[56] will be mirrored for Web mapping 2.0 and other collaborative mapping concepts, media and practices in the discipline in near future.

---

53   Caquard (2011), 137.

54   Caquard (2011), 138.

55   Harris (2015), 46.

56   Perry / Beale (2015).

5. References

Aberley (1993): D. Aberley, Boundaries of home: mapping for local empowerment, Gabriola Island and Philadelphia.

Atalay (2008): S. Atalay, „Multivocality and indigenous archaeologies", in: Junko Habu, Clare Fawcett, and John M. Matsunaga (eds.), Evaluating Multiple Narratives: Beyond Nationalist, Colonialist, Imperialist Archaeologies, New York, 29–44.

Caquard (2011): S. Caquard, „Cartography I: Mapping Narrative Cartography", Progress in Human Geography 37 (1), 135–144.

Chambers (2006): R. Chambers, „Participatory Mapping and Geographic Information Systems: Whose Map? Who is Empowered and Who Disempowered? Who Gains and Who Loses?", Electronic Journal of Information Systems in Developing Countries 25 (2), 1–11.

Chapin et al. (2005): M. Chapin/ Z. Lamb / B. Threlkeld, „Mapping Indigenous Lands", Annual Review of Anthropology 34, 619–638.

Clifford (1996): S. Clifford, „Places, people, and parish maps", in: Sue Clifford and Angela King (eds.), From Place to PLACE: Maps and Parish Maps, London, 3–7.

Conkey / Gero (1997): W. Conkey / M. Gero, „Programme to Practice: Gender and Feminism in Archaeology", Annual Review of Anthropology 26, 411–437.

Crampton (2009): J.W. Crampton, „Cartography: Maps 2.0", Progress in Human Geography 33 (1), p. 91–100.

Crampton (2010): J. W. Crampton, Mapping: a critical introduction to cartography and GIS, West Sussex.

Debord (1955): G. Debord, „Introduction à une Critique de la Géographie Urbaine", Les Lèvres Nues 6, 11–15.

Dodge et al. (2011): M. Dodge / R. Kitchin / C. Perkins, „Introductory essay: technologies of mapping", in: Martin Dodge, Rob Kitchin, and Chris Perkins (eds.), The Map Reader: Theories of Mapping Practice and Cartographic Representation, West Sussex, 116–121.

Dreessen et al. (2012): K. Dreessen / L. Huybrechts / T. Laureyssens / S. Schepers / S. Baciu, MAP-it: a participatory mapping toolkit, Leuven.

Duke / Saitta (1998): P. Duke / D. Saitta, „An Emancipatory Archaeology for the Working Class", Assemblage 4, 1–7.

Dunn (2007): C.E. Dunn, „Participatory GIS – a People's GIS?", Progress in Human Geography 31(5), 616–637.

Elwood (2002): S. Elwood, „GIS and Collaborative Urban Governance: Understanding Their Implications for Community Action and Power", Urban Geography 22, 737–759.

Elwood (2006): S. Elwood, „Critical Issues in Participatory GIS: Deconstructions, Reconstructions, and New Research Directions", Transactions in GIS 10 (5), 693–708.

Gartner (2009): G. Gartner, „Web mapping 2.0", in: Martin Dodge, Rob Kitchin, and Chris Perkins (eds.), Rethinking Maps, London and New York, 68–82.

Grasseni (2004): C. Grasseni, „Skilled Landscapes: Mapping Practices of Locality", Environment and Planning D: Society and Space 22, 699–717.

Harley (1988): J.B. Harley, „Maps, knowledge and power", in: Denis Cosgrove and Stephen Daniels (eds.), The Iconography of Landscape, Cambridge, 277–312.

Harmon (2004): K. Harmon, You are here: personal geographies and other maps of the imagination, New York.

Harmon (2009): K. Harmon, The map as art: contemporary artists explore cartography, New York.

Harris (2015): T.M. Harris, „Deep geography - deep mapping: spatial storytelling and a sense of place", in: David J. Bodenhamer, John Corrigan, and Trevor M. Harris (eds.), Deep Maps and Spatial Narratives, Bloomington, 28–53.

Jacob (2006): C. Jacob, The sovereign map: theoretical approaches in cartography throughout history, T. Conley (trans.), E.H. Dahl (ed.), Chicago and London.

Joyce / Tringham (2007): R.A. Joyce / R.E. Tringham, „Feminist Adventures in Hypertext", Journal of Archaeological Method and Theory 14 (3), 328–358.

King / Clifford (1985): A. King / S. Clifford, Holding your ground: an action guide to local conservation, London.

Kitchin / Dodge (2007): R. Kitchin / M. Dodge, „Rethinking Maps", Progress in Human Geography 31 (3), 331–344.

Kitchin et al. (2011): R. Kitchin / M. Dodge / C. Perkins, „Introductory essay: power and politics of mapping", in: Martin Dodge, Rob Kitchin, and Chris Perkins (eds.), The Map Reader: Theories of Mapping Practice and Cartographic Representation, West Sussex, 388–394.

Kyem (2004): P.A.K. Kyem, „Of Intractable Conflicts and Participatory GIS Applications: the Search for Consensus Amidst Competing Claims and Institutional Demands", Annals of the Association of American Geographers 94 (1), 37–57.

Lee (2016): D. Lee, „Map Orkney Month: Imagining Archaeological Mappings", Livingmaps Review 1 (1), 1–25.

Nold (2009): C. Nold, „Introduction: emotional cartography - technologies of the self", in: Christian Nold (ed.), Emotional Cartography: Technologies of the Self, emotionalcartography.net/EmotionalCartography.pdf (26.02.2017), 3–14.

Obrist (2014): H.U. Obrist, Mapping it out: an alternative atlas of contemporary cartographies, London.

O'Rourke (2013): K. O'Rourke, Walking and mapping: artists as cartographers, Cambridge (MA) and London.

Parker (2006): B. Parker, „Constructing Community Through Maps? Power and Praxis in Community Mapping", The Professional Geographer 58 (4), 470–484.

Perkins (2007): C. Perkins, „Community Mapping", The Cartographic Journal: The World of Mapping 44(2), 127–137.

Perkins (2014): C. Perkins, „Plotting Practices and Politics: (Im)mutable Narratives in OpenStreetMap", Transactions of the Institute of British Geographers 39, 304–317.

Perry / Beale (2015): S. Perry / N. Beale, "The Social Web and Archaeology's Restructuring: Impact, Exploitation, Disciplinary Change", Open Archaeology 1, 153–165.

Pinder (1996): D. Pinder, „Subverting Cartography: the Situationists and Maps of the City", Environment and Planning A 28(3), 405–427.

Pinder (2005): D. Pinder, „Arts of Urban Exploration", Cultural Geographies 12, 383–411.

Ratliff (2007): E. Ratliff, „The Whole Earth, Catalogued. How Google Maps is Changing the Way We See the World", Wired 15 (7), 154–159.

Rød et al. (2001): J.K. Rød / F. Ormeling / C. Van Elzakker, „An agenda for democratising cartographic visualisation", Norsk Geografisk Tidsskrift (Norwegian Journal of Geography) 55 (1), 38–41.

Sletto (2009): B.I. Sletto, "'We Drew What We Imagined': Participatory Mapping, Performance, and the Arts of Landscape Making", Transactions of the Institute of British Geographers 50 (4), 443–476.

Wainwright / Bryan (2009): J. Wainwright / J. Bryan, „Cartography, Territory, Property: Postcolonial Reflections on Indigenous Counter-Mapping in Nicaragua and Belize", Cultural Geographies 16, 153–178.

Walker et al. (2002): D.H. Walker / A.M. Leitch / R. de Lai / A. Cottrell / D. Pullar, „A community-based and collaborative GIS joint venture in rural Australia", in: William J. Craig, Trevor M. Harris, and Daniel Weiner (eds.), Community Participation and Geographic Information Systems, London and New York, 137–152.

Williamson / Connolly (2009): D. Williamson / E. Connolly, „theirwork: the development of sustainable mapping", in: Martin Dodge, Rob Kitchin, and Chris Perkins (eds.), Rethinking Maps, London and New York, 97–112.

Wood (2010a): D. Wood, „Lynch Debord: About Two Psychogeographies", Cartographica 45(3), 185–200.

Wood (2010b): D. Wood with J. Fels / J. Krygier, Rethinking the power of maps, New York and London.

Author[57]

Dr. Piraye Hacıgüzeller
Ghent Centre for Digital Humanities and Archaeology Department of Ghent University (Belgium)
Sint-Pietersnieuwstraat 35
UFO, 9000 Gent

Email: piraye.haciguzeller@ugent.be

DIGITAL CLASSICS ONLINE

Presenting Fragments as Quotations or Quotations as Fragments

Alexandra Trachsel

Abstract: In this paper, we shall present some reflections on issues that arise when scholars prepare new editions of fragmentarily preserved works from antiquity. Our focus will be on those fragments that are transmitted as quotations in other extant texts from antiquity. In such cases, the textual sequences considered as fragments have to be dealt with in a particular way, as they are the results of several selection processes that have not been taken into account adequately previously. A digital approach provides better means to represent this special form of fragments, so that the complexity of the transmission of these pieces of evidence can be understood more accurately.

1. Introduction

The issues dealt with in this contribution originated from questions that were at the center of a research project, which was mainly carried out at the University of Hamburg, but which could also benefit, during a Marie Curie Fellowship, from much input provided by the Department of Digital Humanities at King's College London. The aim of the project consisted in investigating a fragmentarily preserved work from antiquity with the ultimate goal to prepare a new digital edition of the remaining evidence. A large part of our investigations turned out to be dedicated to the question of what a fragment is and how the answer to this question would influence the process of editing fragmentarily texts in the future. Moreover, this question had to be approach from a particular angle, as the work, on which our project focused, was preserved only through quotations made by later authors. The preserved evidence from the lost work did not have the form of a broken piece that can be studied, as for instance an inscription engraved on a stone that later was broken. It should, on the contrary, be seen as a textual sequence that has been embedded as quotation in a second text and was then transmitted as part of this second text. This situation is of particular interest with regard to two editorial issues, for which a digital approach brings new solutions. The first, as mentioned, touches on the definition of the term „fragment" and its status as object of research, whereas the second deals with the arrangement of these items in a new digital edition and the consequences this has for our understanding of the work under discussion. We shall discuss both aspects, but before doing so, we would like to add a few preliminary remarks about the author we are dealing with, as the peculiarities of our case study may have some influence on our approach, even if the two questions, which are at the center of our investigation, have a much broader scope.

2. The historical context of our case study

Demetrius of Scepsis is a scholar from the 2nd century BCE, who is roughly contemporaneous with Aristarchus of Samos and Crates of Mallos, two of the most influential scholars in Hellenistic times.[1] He lived and worked in Asia Minor, either in his hometown Scepsis or in the Library of Pergamon.[2] His main field of research was Homeric scholarship and he wrote a detailed commentary on a special part of the Homeric text, the so-called Catalogue of Trojans, the enumeration, given in book 2 of the *Iliad*, of the warriors fighting on the Trojan side in the famous Trojan War.[3] According to ancient testimonia Demetrius' work was a huge exegetical treatise, extending over 30 books.[4] However, today only 76 rather short fragments are preserved, which is one of the difficulties in our task here.[5] In order to understand Demetrius' work today we have to find a way to bridge the gap between the few items, which were preserved, and the voluminous original work, which they should represent. The title of the work, however, is fairly well attested as ὁ Τρωϊκὸς διάκοσμος (*ho troikos diakosmos*) and gives a good starting point, as it refers to the part of the Homeric text, on which Demetrius was commenting. The two Homeric catalogues (the Catalogue of Ships and the Catalogue of Trojans) are indeed either refers to by ὁ κατάλογος (*ho katalogos*) or by ὁ διάκοσμος (*ho diakosmos*), so that the title of Demetrius' work emphasizes the link between his own work and the text he wanted to explain. With this in mind, we may now proceed to our two questions: (1) what kind of witness of the original work is a fragment when this textual sequence is a quotation preserved in a subsequent work and (2) how should these pieces of evidence been arranged, so that a modern reader my have an idea about the original work, despite the indirectness and fragmentation of our modern access to the ancient work.

3. The status of quotations as witnesses about an ancient work

When a work, like the one by Demetrius, is preserved indirectly, which means only in form of quotations made by later authors, the editorial work starts from a different point than when scholars work on pieces of evidence that came down, even in a broken form, from Antiquity (e.g. inscriptions and papyri, or entirely preserved medieval manuscripts of a given work). In the case of a fragmentarily preserved work, the editorial process does not start with the transcription of a set of (medieval) manuscripts, or of an original document from Antiquity (inscriptions and papyri). In the case of collections of quotations, the editorial work most often starts with the study of previous editions of the texts, in which the quotations were embedded. This means that the analysis is based on critical editions of extant ancient texts, from which only very few parts are relevant.[6] Therefore, there is often more than one source-text involved and the fragments of a given work have to be gathered from several source-texts. In this case, it is not possible to study the history of transmission of all of the source-texts involved, before

---

1 Str. 13.1.55 [C 609], and Montanari (2015) 641–672 and Montana (2015) 60–183.

2 Montanari (1993) 651–652.

3 Il. 2.816–877. This part of the Homeric text is considered as the smaller counterpart to the Catalogue of Ships, which is the enumeration of the Achaean participants in the Trojan War (Il. 2.494–759).

4 Str. 13.1.45 [C603].

5 See the collection made by Biraschi (2011) (= FGrHist 2013).

6 These critical editions are those that are based on the deciphering of medieval manuscripts and of other objects from Antiquity.

focusing on the fragments of the given author, in which one is interested. For Demetrius, this would for instance mean that the editor would have to edit first works with extremely difficult histories of transmission such as Strabo's *Geography*, Athenaeus' *Deipnosophistai* or the scholia to the *Iliad* before focusing on Demetrius, which would be a completely different task, and this not only because of the length of the task.[7] Therefore, when a scholar plans to establish a new edition of a fragmentarily preserved work – and in particular in the case of a digital edition –, two questions should be addressed with regard to this special state of preservation:

(1) How to represent texts which are not based on the transcription of a set of manuscripts or on original documents from Antiquity, but on previous editions themselves based on critical editions of ancient texts.[8]

(2) How to represent as accurately as possible the links between the preserved quotations in a given source-text and the source-text itself, as these textual sequences are more or less close renderings of an initial lost piece.

This second question raises a very important point we should bear in mind when dealing with fragments of works that were transmitted through quotations. Up to now, editions of such kind of works frequently created the illusion that they are reconstructions of a lost work. Nonetheless, we need to remain aware of the fact that, however complete these editions may be, they are only collections of either statements about these works, which are gathered under the heading of *testimonia*, or more or less close renderings of passages from the original work, which passed through the mind and/or writing-process of the second, quoting author.[9]

However, a further aspect a scholar has to take into account, when dealing with fragmentarily preserved texts, has to be added to these two initial questions. Most of the time, when studying an author whose work is lost, the editorial work has also to take into account one or several previous modern editions of the same work.[10] However, modern editors do often not agree with regard to editorial choices about a given textual sequence, which is accounted for as a fragment of a given work. Therefore, if several modern editions of fragmentarily preserved works exist, they often differ considerably with regard to the appearance or non-appearance in the collection of a given textual sequence, the way these pieces of evidence should be edited and presented or about the length the textual sequences should have. Therefore, a further question should be addressed when dealing with fragments:

(3) How to keep visible for the modern reader the choices made by modern scholars/editors in the editorial process.

---

7   The fragments of Demetrius' work are preserved in: Strabo (1st century BCE–1st century CE); Athenaeus, Pausanias and Harpocration (2nd century CE); Hesychius (5th or 6th century CE); Stephanus of Byzantium (6th century CE); Eustathius and Tzetzes (12th century CE) and the scholia to the *Iliad*, those to Pindarus, to Euripides, to Apollonius Rhodius and to Theocritus.

8   The *digital Athenaeus* (http://digitalathenaeus.org) is for instance a project that starts from an extant text, which contains many quotations (Athenaeus' *Deipnosophistai*) and edits it first before focusing on the quotations from other texts that it contains.

9   For a project addressing this issue, see the *Sharing Ancient Wisdoms (SAWS) Project* from King's College London (http://www.ancientwisdoms.ac.uk/).

10   For Demetrius of Scepsis, the list of previous editions/collections of the preserved fragments is composed by: Stiehle (1850), Stiehle (1851), Gaede (1880), Biraschi (2011) and Pagani (2016).

This third question should also remind us of another characteristic of fragments preserved in form of quotations. Their existence depends most of the time not only on the choices of ancient authors who decided to use pieces from them in their own works as quotations, but also on selections modern scholars made in ancient texts. They read the works, in which the quotations have been transmitted and decide which part of the initial text would be a fragment from a given author. Therefore, we have actually two selection processes to deal with when trying to edit fragmentarily preserved authors: first we should focus on the choices the ancient authors made. They had for instance Demetrius' entire work at their disposal and could select those passages from his book, which seemed relevant for their own argumentation; in a second step, we should also analyze the decisions modern editors make when extracting these pieces again out of their transmitted context.[11] Therefore, we should remain aware of the fact that, when taking the form of quotations, fragments have no independent existence and should be represented within their context, so that the reader can take account of this intermediary status. From this point of view, it becomes obvious that the link between the textual sequence, which an editor decided to consider as evidence from a lost work, and the source-text, in which it was preserved as a quotation should be maintained, when one attempts to gather all pieces of evidence from a lost work. This procedure would indeed enable an editor to express, not only the ancient authors' choices, but also the ones made by his modern predecessors.

To sum up, with regard to the three issues concerning editions of fragmentarily preserved works that we just mentioned, two characteristics about the way the evidence from these works has been preserved should be made visible when establishing a new (digital) edition. First, it should remain clear that their existence depends on two kinds of choices: the one from ancient authors who selected passages from a previously written work to use them in their own text and the one from modern scholars who decide which part from an entirely preserved text could be interpreted as a quotation from another work. Secondly, as the source-text, in which the quotations are made, is crucial for the transmission of the pieces of evidence from the lost text,[12] the establishing of the wording of the fragments depend on two steps, which are carried out by two different scholars, or group of scholars: the editors (1) dealing with the fragments will build on the work the scholars (2) who edited the source-text have done. This situation creates some problems with regard to these two methodological steps, as an „edition" of a fragmentarily preserved text does not mean the same as an „edition" of an entirely preserved text. There is for instance less need for the deciphering of the original document or for a diplomatic transcription of the writing it bears. The work for an edition of a fragmentarily preserved work starts, as mentioned above, directly with the critical editions of the source-texts, as previous scholars have established them. But, as a matter of fact, there is often more than one edition of a work from Antiquity, so we have the following situation:

---

11   Trachsel (2012) 415–429.

12   This is particularly visible when scholars use the concept of „cover text" to define a source-text, in which the quotations are preserved. For this concept (a cover text conceals, protects and encloses a textual sequence from a previous work) and its implications for collections of fragments, see Schepens (1997) 144–172, Schepens (2000) 1–29, Lenfant (2002) 415–477, Lenfant (2013) 289–305, Berti (2012) 439–458, and Berti (2013) 269–288.
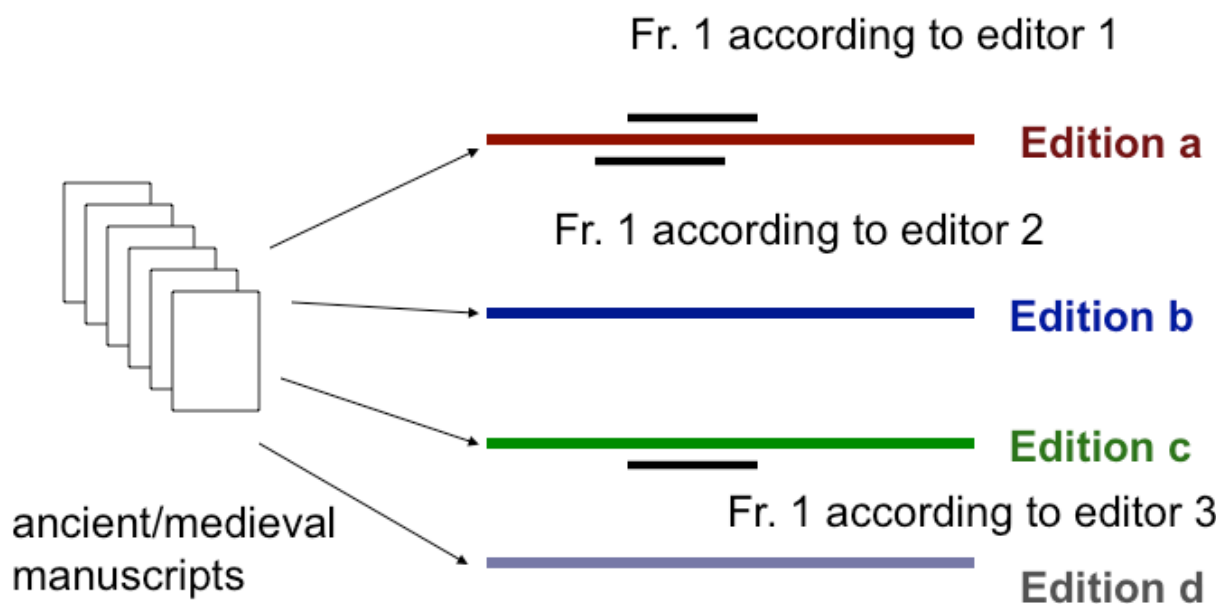
Figure 1: Edition a, b, c, d: the different modern editions of an extant work (source-text); Fr. 1 (according to editor 1, 2, 3): the choices made by the editors of a collection of fragments about a textual sequence that could count as fragment of a lost work.

This representation shows the most basic case, when editor 2, in this case not the editor of the source-text, but the one of the collection of fragments, may have used the same edition of the source-text as editor 1 (who made a previous edition of the same collection of fragments), but did not agree about his or her choice when it come to the decision about which part from the source-text belongs to the quoted author. Then, editor 3, who again works on the same collection of fragments, but a few years after editor 2, may agree with either editor 1 or 2[13] about the choices they made in the text about the part, which belongs to Demetrius, but he or she may have used a different edition of the source-text[14] and then still have a rather different text for fragment 1.

Often, however, the situation is more complex, as for instance an utterance, which is believed to belong to a lost work, is composed of two textual sequences that have been transmitted in two different ancient works. In this case, because the passages from the two different source-texts are almost similar in their content or wording, editors believe that it was the same passage from the original work, which was used in two different ways in Antiquity, most of the time by two different scholars, (which we call ancient work 1 and ancient work 2 in the illustration below) and therefore got, from the beginning, two different renderings.

---

13   In our schema editor 3 agrees with editor 1.

14   Here it is edition c, which is used. This may be a more recent edition of the extant source-text, in which the quotations are preserved, or one that editor 3 considers more accurate, because of his or her own scholarly background, than the other editions.
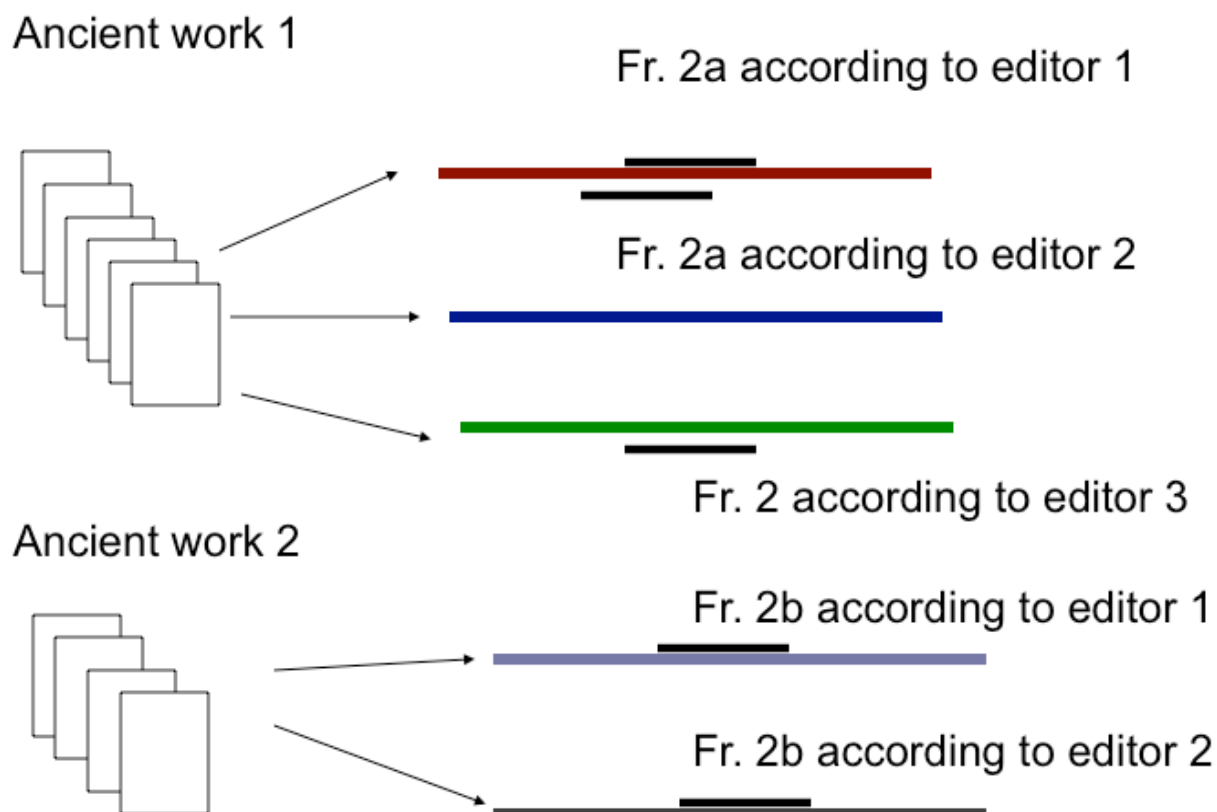
Figure 2: Editor 1 and editor 2 compose the fragment from the lost work from two textual sequences taken from two different ancient works (Fr. 2a and Fr. 2b). Editor 3, on the contrary, argues that only the passage from ancient work 1 is a witness of the lost work. He or she dismisses the evidence from ancient work 2.

This situation is, however, not the same as saying fragment 1 is from source-text A whereas fragment 2 is from source-text B. That would mean that, even if the two fragments (or quotations) are from the same lost work, they come from different places within this book. For instance one comes from the beginning whereas the other comes from the end. In this case, the two fragments can be quite different in their content and not be linked to each other. In the case that is illustrated by the schema, the two pieces are believed to illustrate exactly the same passage of the lost work, but preserved in different renderings.

In some cases, the link between the wording of fragment 2a and fragment 2b is obvious and there is no difficulty to believe that they come from the same passage. In other cases, however, the link one editor sees between two texts is more controversial, especially when there is, for instance, no name appearing in one of the fragments, or if fragment 2b is rather short or comes from a work for which the history of transmission is in itself complex. This aspect is represented in the schema by editor 3, who dismisses the evidence from the second ancient work in opposition to editor 1 and editor 2. He or she only considers the textual sequence from ancient work 1 as a witness from the lost work.

This complexity of the modern choices about the accuracy of a preserved textual sequence to represent an extract from a lost work makes it even more relevant to represent these modern choices in a new digital edition of a collection of fragments. Moreover, the two cases just presented should also have shown how peculiar the status of a fragment of a lost work is. It become obvious from these considerations that a fragment, if it has been preserved as a quotation, becomes a cluster of different renderings which are sometimes assembled from very different ancient works in which they have been preserved in very different forms. Moreover, the relations between the supposed wording of the original work and the several preserved

renderings, as well as the relations between the different renderings, may be quite different and has to be defined by the editor in order to establish how accurately the preserved textual sequence may represent the original work. All these questions challenge the more traditional view that a fragment of a lost work is a well-defined textual sequence, which is supposed to reflect the original state of this work and which can be detached from the context, in which it has been preserved, and printed as an independent body of text in a collection of fragments. These questions could, however, be taken into account when using new digital tools to establish such editions of fragmentarily preserved works.[15]

4. The arrangement of the fragments once they have been defined

Once the textual sequences have been defined as belonging to a given lost work, the editor has to face still another difficulty, namely the one of arranging the selected passages in the collection he or she wants to establish. This further choice, even if this seems trivial, may have considerable effects on the way the work will be perceived by modern readers and has therefore not to be neglected.

In the case of Demetrius of Scepsis, this question of arrangement or ordering of the fragments in a collection is particularly important. First, as mentioned above, we have kept only very few fragments from his work and, because of this small number of evidence, we are no longer able to reconstruct the original structure of the work. We rely, therefore, entirely on modern opinions with regard to this question. Secondly, because of this lack of evidence, the ways modern editors arranged the fragments differ widely from one collection to the other and they provide completely different perceptions of the work. In our case of the fragments of Demetrius' work, the first editor, Richard Gaede,[16] arranged most of the fragments according to a geographical principle. His choice is explained by the fact that the German scholar took Strabo's *Geography* as his starting point, as the largest amount of fragments from Demetrius' work was preserved in this source-text. Strabo was a geographer, who described the regions he was interested in according to a spatial representation. For instance, during his description of the Troad in book 13 of his *Geography*, the part, in which Demetrius is most frequently quoted by Strabo, the geographer follows an outline, which goes from North to South and from East to West. Moreover, he first describes the seashore before giving details about the interior of the land. In this description, Strabo quotes Demetrius as a valuable source, whenever he finds his work relevant for the point he wants to make. Gaede took account of this situation, when preparing his collection of the fragments from Demetrius' work. He collected and numbered the passages from Strabo's text defined as quotations of Demetrius' work according to their occurrence in the Strabonian account. Moreover, the modern scholar gave them a somewhat disproportional weight, as he decided to subordinate fragments coming from source-texts other than Strabo to this structuring principle. This means that the other witnesses about Demetrius' work were added rather independently either before the core elements from Strabo or after.[17] This procedure highlights the geographical content of Demetrius' work and led subsequent scholars

---

15   See Romanello (2009) 155–174, Romanello (2011) 147–155, and Berti (2009) 259–262 for the present state of research.

16   Gaede (1880).

17   For instance all the fragments coming from Athenaeus' *Deipnosophistai* are given before those coming from Strabo's *Geography*. This is explained by the fact that, despite their smaller number, they all contain some indications about the place within Demetrius' work from where they were taken.

to classify Demetrius' work as a mainly geographical treatise.[18] However, other arrangements are possible: some scholars have ordered the fragments from Demetrius' work according to the chronological order of the authors who quoted Demetrius.[19] Finally, a third solution would highlight the link between Demetrius' work and the Homeric text, which he wanted to explain. Such an approach would allow us to take into account the fact that Demetrius' work was most probably a commentary-like exposition of the scholar's statements, which followed, in its structure, the outline of the Catalogue of Trojans.[20] Moreover, such an additional layer, which is based on the outline of a well-defined part of the Homeric text rather than on a map, would enable us to explore a richer way of interpreting Demetrius' goals, which would probably offer a more accurate understanding of Demetrius' work. However, with such an approach, there are again several difficulties, which must first be resolved. For instance, we would need to decide which fragments could be linked to which word or verse of the Homeric text. For some of the fragments such a link is easy to find, especially for those mentioning explicitly toponyms occurring in the Catalogue of Trojans or those about textual criticism for a given line.[21] For other fragments, the situation is more difficult. Sometimes the line or the word, on which Demetrius is commenting, does not come from the Catalogue of Trojans.[22] In still other cases, the topic dealt with in the preserved fragment can not be linked to any of the lines from the Homeric texts, neither to those of the *Iliad* nor to the ones of the *Odyssey*.[23] In order to explain both situations, we must consider two hypotheses. In the first, we must assume that in some of his comments Demetrius quoted passages, which came from other parts of the Homeric texts than the Catalogue of Trojans. In the second, we should postulate that only the parts of his comments, which did not contain the link to the Homeric text, were preserved. Indeed, to explain the first case, we must hypothesize that a later author, who reused Demetrius' work, extracted from Demetrius' long comment on a given line of the Catalogue of Trojans a textual sequence, which deals with a passage that came from a different part of the Homeric text. In Demetrius' comment this may only have been a secondary point, but for the quoting author these parts were more interesting, so that he or she selected it for his own text. The explanation for the second hypothesis is similar. In this case, the author, who quoted from Demetrios' work, was interested in a detail from Demetrius' comment, which came for instance from an excursus made by Demetrius in his comment. This portion of the text may have been quite independent from the main point of Demetrius' comment and completely unrelated to the Homeric text. But, as we kept in the quotation once again only the issues selected by the quoting author, we do no longer see how these elements were used in the original comment.

Another difficulty arises when more than one word in the Homeric text can be linked to the topic developed in one fragment, or if a word, which is dealt with in a fragment, is mentioned more than once in the Homeric text. In both cases, even if the link between the word, or the place name, and the content of the fragment is clearly established, there still remains some ambiguity about which of the occurrences in the Homeric text should be taken as entry for the comment

---

18   This is the reason why Demetrius appears for instance in volume 5 of the continuation project of Jacoby's *Fragmente der griechischen Historiker* (= FGrHist) dedicated to geography. See Gehrke (2016).

19   Müller (1851) 382 (= FHG) and Pagani (2016). This procedure does not take account of the structure of the original work, but emphasizes the history of transmission.

20   This is the ordering pattern Stiehle (1850) 535 follows.

21   E.g. fr. 20 [Gaede/Biraschi] = Steph. Byz. s.v. Ἀρίσβη (α 426 [Billerbeck]) about *Il*. 2.836; fr. 44 [Gaede/Biraschi] = Str. 14.5.28 [C680] and fr. 45 [Gaede/Biraschi] = Str. 12.3.20–23 [C550–552] about *Il*. 2.856–857; fr. 55 [Gaede] (= fr. 55a [Biraschi]) = Str. 8.3.6 [C339] about *Il*. 2.839.

22   E.g. fr. 17 [Gaede] (= fr. 17a [Biraschi]) = Str. 10.5.19 [C489] about *Il*. 2.676–677, a line from the Catalogue of Ships; or fr. 75 [Gaede/Biraschi] (= Tzetz. *ad Lyc*. v. 530) about *Il*. 2.701, another line from the Catalogue of Ships.

23   E.g. fr. 6 [Gaede/Biraschi] = Athen. 15.697c; fr. 7 [Gaede/Biraschi] = Athen. 4.155b; fr. 9 [Gaede/Biraschi] = Athen. 3.80d; fr. 10 [Gaede/Biraschi] = Athen. 4.173f.

Demetrius gave. Furthermore, in some cases it is also important to take into consideration personal names, not only place names, as sometimes heroes are mentioned by a word that refers to the place where they are living, especially in the Catalogue of Trojans. For instance the Trojans and the Dardanians are people that are named by the place where they are living (e.g. Trojans means those from Troy and Dardanians those from Dardania). Furthermore, it is often only possible to link a fragment to the Homeric line through an action a hero performed in a given place, which is not mentioned in the Catalogue. For instance, the hero Pandarus, mentioned in *Il.* 2.827 of the Catalogue of Trojans, goes to the river Scamandros, which is not mentioned in the Catalogue, and encounters there another hero. Demetrius may have been interested in explaining the setting of the river Scamandros, but does it when the name of Pandarus occurs in the Catalogue of Trojans. In this case, in order to find a link between the preserved textual sequence and the Homeric text, we have not only to look for the name of the Scamandros, which may or may not occurring in the Catalogue of Trojans, but also for the one of Pandarus. Finally, there will certainly be cases, where it is simply not possible to decide. Especially in the fragments, which come from source-texts other than Strabo, the topic is often completely unrelated to the Homeric text.[24] Therefore, for a large majority of the fragments, it will be problematic to establish such a link to the Homeric text and the degree of certainty about such attributions should be expressed and made visible to the reader who may decide for himself if he wants to accept the editor's choice or not. In a digital edition, the presentation of such competing arrangement is certainly possible. In this case, not only the choices underlying this third way of arranging the fragments would remain visible, but also the two other options concerning the ordering of the fragments, the one emphasizing the geographical aspect of Demetrius' work (Gaede) and the one highlighting the history of transmission (Müller/Pagani).

5. Conclusion

In this discussion, we attempted to show, by presenting two issues that are at the centre of the editorial work about fragmentarily preserved texts, that, because of the many assumptions that guide an editor when establishing a collection of fragments, a new edition should explore nowadays the additional possibilities of digital tools in order to take account of these scholarly choices and to present them to the future reader. In the first part, we focused on the question of how a textual sequence preserved in a source-text may be interpreted as a piece of evidence from a lost work and what status this textual sequence may have when seen as a fragment from this lost work. In this context, the editorial choices concern either the length of the passage defined as fragment, or the wording of this passage, as the textual sequence may be rendered differently in various editions of the source-text. In the second part, the editorial choices about the content were dealt with. Depending on how a modern scholar interprets the content of a passage defined as a fragment he may opt for another arrangement of the fragments and creates, in doing so, a different image of Demetrius' work. Therefore, because of these many choices, on which an edition of a fragmentarily preserved text is based, we would gain much if it would be possible not only to give the most accurate reconstruction of this work, but also to present the scholarly choices, which lead to this description. This combination would certainly be one of the biggest advantages of a digital edition in comparison to a more traditional one.

---

24   E.g. fr. 3 [Gaede/Biraschi] = Athen. 14.658b; fr. 4 [Gaede/Biraschi] = Harpocr. s.v. Θυργωνίδαι; fr. 69 [Gaede/Biraschi] = Harpocr. s.v. Ἴων; fr. 72 [Gaede/Biraschi] = Athen. 2.44e; fr. 73 [Gaede/Biraschi] = Athen. 4.167d.

6. Selective Bibliography

Berti (2009): Berti, M., Romanello, M., Babeu, A. and Crane, G., „Collecting Fragmentary Authors in a Digital Library", in: Proceedings of the 2009 Joint International Conference on Digital Libraries, *Austin* TX, New York, 259–262.

Berti (2012): Berti M., „Citazioni e dinamiche testuali. L'intertestualità e la storiografia greca frammentaria", in: V. Costa (ed.), Tradizione e Trasmissione degli Storici Greci Frammentari II. Atti del Terzo Workshop Internazionale, Roma, 24–26 febbraio 2011, Tivoli (Roma), 439–458.

Berti (2013): Berti M., „Collecting Quotations by Topic: Degrees of Preservation and Transtextual Relations among Genres", Ancient Society 43, 269–288.

Biraschi (2011): Biraschi A. M., „Demetrios von Skepsis (2013)", in: H.-J. Gehrke (ed.), Die Fragmente der griechischen Historiker, Part V, Brill Online [first published online in 2011 at <http://referenceworks.brillonline.com/entries/fragmente-der-griechischen-historiker-v/demetrios-von-skepsis-2013-a2013>].

Gehrke (2016): Gehrke H.-J., „About Die Fragmente der Griechischen Historiker continued Part V (about_v)", in: H.-J. Gehrke (ed.), Die Fragmente der Griechischen Historiker Part V, Brill Online [Consulted online on 08 September 2016 at <http://dx.doi.org/10.1163/1873-5363_jcv_aabout_v>].

Gaede (1880): Gaede R., Demetrii Scepsii quae supersunt, Greifswald.

Lenfant (2013): Lenfant D., „The Study of Intermediate Authors and its Role in the Interpretation of Historical Fragments", Ancient Society 43, 289–305.

Lenfant (2002): Lenfant D., „Les citations de Thucydide dans les scholies d'Aristophane: contribution à l'analyse de fragments d'historiens", in: S. Pittia (ed.), Fragments d'historiens grecs. Autour de Denys d'Halicarnasse, Rome, 415–447.

Montana (2015): Montana F., „Hellenistic Scholarship", in: F. Montanari, S. Matthaios, and A. Rengakos (eds.), Brill's Companion to Ancient Greek Scholarship, Leiden-Boston, 60–183.

Montanari (1993): Montanari F., „Pergamo", in: G. Cambiano, L. Canfora, and D. Lanza (eds.), Lo spazio letterario della Grecia antica, tom. 1,2: L'Ellenismo, Roma-Salerno, 639–655.

Montanari (2015): Montanari F., „Ekdosis. A Product of the Ancient Scholarship", in: F. Montanari, S. Matthaios, and A. Rengakos (eds.), Brill's Companion to Ancient Greek Scholarship, Leiden-Boston, 641–672.

Müller (1851): Müller C., Fragmenta historicorum graecorum (vol. 4), Paris.

Pagani (2016): Pagani L., „Demetrius [2]", in: F. Montanari, F. Montana, and L. Pagani (eds.), *Lexicon of Greek Grammarians of Antiquity*, BrillOnline [consulted online on 04 March 2017 at <http://dx.doi.org/10.1163/2451-9278_lgga_urn:cite:cidocCRM.E21:lgga.Demetrius_2_it>].

Romanello (2009): Romanello M., Berti, M., Boschetti, F., Babeu, A. and Crane, G., „Rethinking Critical Editions of Fragmentary Texts by Ontologies", in: S. Mornati (ed.), Rethinking Electronic Publishing: Innovation in Communication Paradigms and Technologies – Proceedings of the 13th International Conference on Electronic Publishing held in Milano, Italy 10–12 June 2009, Milano, 155–174.

Romanello (2011): Romanello M., „The Digital Critical Edition of Fragments: Theoretical Problems and Technical Solution", in: P. Cotticelli Kurras (ed.), Linguistica e Filologia Digitale: Aspetti e Progetti, Alessandria, 147–155.

Schepens (1997): Schepens G., „Jacoby's FGrHist: Problems, Methods, Prospects", in: G.W. Most (ed.), Collecting Fragments – Fragmente sammeln, Göttingen, 144–172.

Schepens (2000): Schepens, G., „Probleme der Fragmentedition (Fragmente der griechischen Historiker)", in: C. Reitz (ed.), Vom Text zum Buch, St. Katharinen, 1–29.

Stiehle (1850): Stiehle R., „Der Τρωϊκὸς διάκοσμος des Demetrios von Skepsis", Philologus 5, 528–546.

Stiehle (1851): Stiehle R., „Nachtrag zu der Abhandlung „Der Τρωϊκὸς διάκοσμος des Demetrios von Skepsis", Philologus 6, 344–347.

Trachsel (2012): Trachsel, A., „Collecting Fragments Today: What Status Will a Fragment Have in the Era of Digital Philology?", in: C. Clivaz, J. Meizoz, F. Vallotton, and J. Verheyden (eds.), Lire demain – Reading tomorrow, Lausanne (e-book version), 415–429.

7. Digital Projects

Berti M., and Cumani, G., „The digital Atheaneus" <http://digitalathenaeus.org> (consulted online on 08 September 2016).

Roueché C., Hedges, M., Dunn, S., Jordanous, A., Lawrence, F., Lawrence, K. et al., „The Sharing Ancient Wisdoms (SAWS) Project" <http://www.ancientwisdoms.ac.uk/> (consulted online on 08 September 2016).

Author[25]

Alexandra Trachsel
University of Hamburg
Department of Greek and Latin Philology

Email: alexandra.trachsel@uni-hamburg.de

Standards for Networking Ancient Person data:
Digital approaches to problems in prosopographical space

Gabriel Bodard, Hugh Cayless, Mark Depauw, Leif Isaksen,
K. Faith Lawrence, Sebastian Rahtz[†]

Abstract: Prosopographies disambiguate names appearing in sources by creating lists of persons, but the progress of scholarship now makes these lists difficult to maintain. In a digital context unique stable identifiers can be reshuffled ad libitum when searching and ordering information. Digital data increasingly brings together complementary research outputs: the Standards for Networking Ancient Prosopographies project takes on the challenge of creating an aggregated resource, adopting a Linked Open Data approach. In this paper we shall present three case studies highlighting the promise and problems of encoding unambiguous identities, titulature and other disambiguating information, and treating divine figures as person-data, respectively. Digital approaches are tools for research, assisting rather than replacing the historian, who remains central to the research endeavor.

1. Introduction

Digital methodologies, especially the use of Linked Open Data, are being used to encode, explore, share and open to computational analysis many areas of ancient world data, especially at very large scales. We hope to address some of the promises and concerns with such an approach to the particular case of ancient prosopography, namely the aggregation of multiple sources of person-data into a single virtual person authority.

The term 'prosopography' commonly refers to a scholarly method, investigating the communalities of a specific group of people to learn more about the social and political background of events and evolutions.[1] Like biography, the traditional focus is on the well-known political elite. Prosopography, however, is interested in what people have in common, rather than in what makes them stand out as individuals. For that reason, modern prosopography also studies 'ordinary people' to map longer-term social evolutions.

Prosopography also refers, however, to the tools, which the scholarly method produces and uses. In this meaning a prosopography is a list of people sharing a specific characteristic: geographical, chronological, or thematic. This limitation may be implicit, as in the case of the *Prosopographia Imperii Romani* (PIR),[2] which only includes (important) office holders, or the *Prosopographia Ptolemaica* (PP),[3] which only includes people whose title or activity places them in a specific social context. In many cases, however, it is explicit, e.g. in Devijver's

---

1   See Keats-Rohan (2007), and especially the contribution of Verboven et al. (2007).

2   PIR² (= Groag et al. (1933–2016)); indices are also searchable online at <http://pir.bbaw.de/>.

3   The printed volumes by Leuven scholars all appeared in the series Studia Hellenistica, vol. 1–6 between 1950 and 1968, vol. 7 (index) in 1975, vol. 8–9 (addenda et corrigenda) in 1975 and 1981. For a brief history of the later history of the PP, see Mooren (2001). Add also vol. 10 (ethnics) published in 2002.

prosopographical work on equestrian officers or that of Janiszewski and colleagues on Greek rhetors and sophists.[4]

Whatever the selection criterion, the people belonging to the group must be unambiguously identified in the prosopographical lists. Names are an excellent tool to do this, but unfortunately homonymy is rife in certain groups and some of the group's members remain anonymous. This is traditionally where numbers come in: they are in ready supply as unique identifiers, either in combination with the first letter of the name (e.g. PIR2 H 234: second edition, names with H, no. 234) or on their own (e.g. PP V 13081: fifth volume, no. 13081). Together with an alphabetic ordering system, they allow easy referencing and navigation across the multiple volumes an extensive prosopography can consist of. In combination with indices, they even allow a thematic classification, e.g. for various social groups in the PP.

In a static environment, this would be a perfect system. As the progress of scholarship, however, leads to corrections, deletions and additions, the original ordering system is difficult to maintain in a new, updated version: if holders of an office are numbered in chronological order, the discovery of a new incumbent disturbs this sequence; if the beginning of a name of a fragmentary preserved person can be reconstructed thanks to new information, the alphabetical order may be disrupted; or if someone eventually turns out never to have been qualified for listing, this person's removal can cause a gap (think of the alleged pharaoh Ptolemy VII).[5] Creating a new order with new numbers is potentially confusing and therefore problematic. A possible solution would be to assign meaningless consecutive serial numbers to all individuals in a random order, and to provide all relevant information for that person under that serial number. This system is better suited to cope with change, but it has the drawback that the user would always need to pass through various elaborate indices in order to find information. In a non-digital context, this would be too cumbersome and user-unfriendly, and the resulting list would no doubt also be considered too chaotic. But in a digital context this modus operandi becomes plausible. Unique stable identifiers can create a firm skeleton to be reshuffled ad libitum when searching and ordering information. If a number is in itself meaningless, it is no longer difficult to change information provided under that number, or to add a new person to the selection that in non-digital times would have disturbed the order.

The advent of computers in the seventies and eighties thus provided a perfect solution for many practical problems of prosopography, but it took some time before this insight seeped through. Prosopography did embrace the computer, and people did start work on the digitization of paper volumes, but mainly because it made additions and corrections much easier. The *Prosopographia Ptolemaica* was such an early adopter, and moved to a relational database structure to hold its information.[6] Making the database available to people was difficult, however, and perhaps because CD-ROMs were far from perfect, practically no databases for the ancient world were ever produced on this medium.[7]

The arrival of the Internet in the nineties and early two-thousands greatly facilitated digital publication of prosopographies. And it was in this new scholarly context that the importance

---

4 Devijver (1989–1992) and Devijver (1973–2001); Janiszewski et al. (2014).

5 For Ptolemy VII Memphites, see the late Chris Bennett's website on chronology: <http://www.tyndalehouse.com/egypt/ptolemies/memphites_fr.htm>. Huß (2001), for example, renumbers the late Ptolemies, which can be very confusing.

6 Mooren (2001).

7 In the late nineties some of the data of DYABOLA <https://en.wikipedia.org/wiki/Projekt_Dyabola> were distributed on CD and DVD (as communicated to us by an anonymous reviewer).

of stable identifiers became obvious. When the PP was integrated into Trismegistos People, a new, purely numeric, arbitrary and stable numbering system for attestations and individuals was introduced.[8] Kallikrates son of Boiskos, an elite official known under various numbers referring to different aspects of his persona as 'PP III 05164+add.', 'PP IV 10086', 'PP VI 14607', 'PP E2399' etc., is now identified as TM Per 2137 (or http://www.trismegistos.org/person/2137). This same man, however, also features in the *Lexicon of Greek Personal Names* (LGPN) as Kallikrates no. 130 in volume 1 (1-Καλλικράτης-130). The connected persistent identifier in the online version there is V1-45988 (or http://www.lgpn.ox.ac.uk/id/V1-45988).

Here a similar problem emerges in a slightly different form: the prosopographical method by nature normally focuses on specific groups and subsets of the population. But since most individuals have multi-faceted personalities, there is bound to be overlap between various projects. For one example, PIR², which collects the names and offices of Roman elites of the first three centuries C.E., has an entry for a certain Aelia Pithia (PIR² A 306); LGPN, which includes only people with names attested in Greek, but from a much wider period, lists a Πειθιάς (LGPN 1-Πειθιάς-10). Both cite the same inscriptions and sources as references for her name, and indeed the LGPN cites PIR² among them.[9]

Many people will therefore be known under various numbers, and in the worst case the digital prosopographies remain isolated silos of information. It would greatly facilitate communication and data exchange in a digital context if a single number could identify all individuals of the ancient world,. An approach, which has shown much potential in achieving this aim is known as Linked Open Data (LOD). First proposed by Tim Berners-Lee, the inventor of a suite of technologies that underpin the World Wide Web, Linked Open Data Approaches similarly combine two digital techniques to make the connection and integration of independent and heterogeneous digital resources such as prosopographies possible.[10]

Any project that attempts to organize large amounts of discrete records is faced with the problem of naming those records in an easily referenceable way. Epigraphic corpora, for example, typically assign sequential numbers to inscriptions. PIR, as we have already seen, uses the first letter of the name in question plus sequential numbers. LOD systems do precisely the same thing, except that instead of numbers, they use Uniform Resource Identifiers (URIs). URIs are a system of global, Internet based identifiers, which allow any given record, or atomic concept to be uniquely referenced. A number of URI schemes can be used, but popular practice is to use the kind of HTTP Uniform Resource Locators (URLs), with which Web users are familiar as a means of specifying web pages. The advantage of the system is obvious – anyone in control of the Web domain forming the basis of the URI can associate it with disambiguating information, which clarifies the nature of the concept to which the URI refers. When the web address is resolved – a process known as 'dereferencing' – a human or machine user can immediately get that information. This is a radical departure not only from paper-based identification systems, but also from the so-called 'siloing' effect of using privately assigned identifiers to records in a personal database. A fundamental premise of LOD is that mutual use of the same URI implies the common referencing of the same concept.

The second technology, known as Resource Description Framework, or RDF, uses URIs as a basic vocabulary for constructing simple assertions about the concepts they represent. These

---

8   Trismegistos People: <http://www.trismegistos.org/ref/index.php>. See Depauw/Van Beek (2009).

9   As Aelia Pithia/Πειθιάς is an unambiguous case, SNAP has already merged these records as http://data.snapdrgn.net/person/673754/.

10 Berners-Lee (2006).

statements, named triples due their subject-predicate-object structure, are composed predominantly of URIs, but may also end with literal values such as strings of text or numbers. So we have dereferenceable names for all of the important entities in our LOD system and we also have a way of linking those entities together via semantic links. Just as with natural language, the use of a common set of URI terms across a series of statements creates a complex Web of assertions, but with the additional computational benefit of being formally describable as a network graph. These twin ideas, of a common digital vocabulary and the easy combination of separate datasets, gives rise to its alternative designation as a 'Semantic Web'.

It must nevertheless be emphasised that the potential for digital interconnectivity cannot solely reside in technological solutions. The establishment of cross-navigable networks of related information requires an ecology of resources – some offering controlled vocabularies, others hosting content, or linking it together, and further services which can search, analyse or visualise these complex graphs in informative ways. This in turn raises issues not only of trust and trustworthiness, but also of how to foster open and decentralised ecosystems that sum to more than their parts, without creating single points of dependency or failure.

For places in the Classical world, the Pelagios Commons project has shown the usefulness and viability of collecting location references across projects, based on various gazetteers of toponyms, working in the context of an ecosystem of ancient world Linked Data projects.[11] Standards for Networking Ancient Prosopographies (SNAP)[12] takes on the challenge of creating a similar resource for persons and person-like entities, again adopting a linked open data approach. The main objective is to create a virtual authority list, based on an aggregation of many digital prosopographies, person-lists, and even library catalogues, to which digital projects can link to identify and disambiguate person-references in their sources.

In this paper we will focus on three areas which illustrate the difficulties in codifying person data using the SNAP model to facilitate a linked open data approach: record matching when projects disagree on the entity resolution with respect to the same textual co-references (case study 1); identifying possible matches between entity records where we do not have a shared textual reference (case study 2); and mythological, fictional and other pseudo-historical and non-historical entities whose definition and conceptual identification within the source texts themselves can be fluid (case study 3). These three examples represent three different ways in which computational methods, as embodied in the SNAP model, interplay with prosopographical research. From the solvable (case study 1) to the unsolved, and possibly unsolvable (case study 3), we show how this lightweight model can further research impact. Not only does it bring research outputs together, it also opens up channels of collaboration and academic debate around person and person-like entities, because it reflects both ambiguity and scholarly certainty.

2. Case study 1: people who don't match between prosopographies

Different prosopographies may disagree about the identity of a person for a variety of reasons. Therefore SNAP has to implement mechanisms both to merge entities identified in multiple resources, and to account for the discrepancies between them. In the SNAP system, every entity

---

11 See Isaksen et al. (2014); cf. the various papers in Elliott et al. (2014) for an overview of related projects.

12 <https://snapdrgn.net/>. See Lawrence/Bodard (2015).

identified as a person in a source prosopography has a new SNAP Person id generated for it. By adding further information to that identifier, SNAP does not discard or alter information, but keeps track of the source of each item.

The new SNAP Person points back at the identifier of the entity it was derived from. If two records are determined to be about the same individual, a merge operation may be performed. This results in the creation of a third SNAP Person id, a "Merged Resource" as well as a Person. It indicates that it replaces the original two entries, cites the person or process that performed the merge, and provides a reason for it, e.g. because both records have the same name and cite the same inscription.[13] Note that the original source data is unchanged by these operations, it just gains additional associations, which go beyond the assertions of the sources.

But what happens when, for example, a SNAP source prosopography asserts that its single record represents two records in another prosopography? There are at least two examples from LGPN where records cite more than one entry in the *Prosopographia Imperii Romani*: 5a-Βάσσος-54 (Π. Ῥουτίλιος Βάσσος Ἰουνιανός)[14] cites PIR² R 243 and 244, and 5a-Πρόκλος-34 (Ἰούλ. Πρόκλος)[15] cites PIR² I 492 and 493. In the first case (Rutilius Bassus) PIR admits the possibility that both are the same but also mentions a second solution: Rutilius Bassus R 244 being the son of Rutilius Bassus R 243.

In an email to the authors, Matthäus Heil points out that the second case is less clear. LGPN follows the commentary of the relevant inscription (I.Ephesos 1103), which states that both occurrences of Iulius Proculus in the inscription refer to the exact same person.[16] But for Iulius Proculus (PIR² I 493) the commentary cites Alföldy, who argued that this person was probably a suffect consul between AD 145 and 160.[17] The Historia Augusta, *Commodus* 7,7 states that the other Iulius Proculus (PIR² 492) was killed by the emperor, so after AD 180 (or better: ca. AD 189/190, as the context of HA shows). If I 492 is the same person as I 493, he must have been a very old man when killed. This is not impossible, but because Commodus' other victims mentioned in *HA* 7,6 ff. are as far as we can tell about a generation younger, it may be doubted.

SNAP does not provide a technological solution to resolve either of these ambiguities. They must be solved (or not) in the same way as all such problems: via the accumulation of evidence and scholarly argument. What SNAP does give us is a way to represent the different possibilities. A Merged Resource may be created that combines LGPN 5a-Πρόκλος-34, PIR² R 243, and 244, while the second case might be represented by a Merged Resource combining just LGPN 5a-Πρόκλος-34 with PIR² I 493, and a SNAP identifier representing PIR² I 492. These multiple Merged Resources and entries may result from scholarly disagreement, with one scholar possibly arguing in favour of the unification of all three and another against the identification. In the case above, however, the difference is not one of scholarly opinion (both would no doubt agree on the uncertainty concerning the two Iulii Proculi), but one of editorial practice: one database combines the uncertainly different figures under a single name, while the other divides them into two potential people. Competing, mutually exclusive SNAP Persons could be created, each pointing to the other in such a way that it is clear that accepting one means excluding the other. The SNAP system does not, therefore, attempt to reconcile

---

13   See Bodard et al. (2014), s.v. "Scenario 3. Establishing alignment between prosopographies."

14   http://www.lgpn.ox.ac.uk/id/V5a-38858

15   http://www.lgpn.ox.ac.uk/id/V5a-38861

16   Matthäus Heil, pers. comm.

17   Alföldy (1977) 168–69.

contradictions in its sources in the absence of further scholarly investigation. Its goal is to represent the state of scholarship on a given individual in a way that can be easily queried and referenced by researchers, potentially leading to tools for further research that may contribute to the resolution of such questions.

A further issue therefore arises from this recording of scholarly uncertainty, disagreement, qualification and other complexity: the research tools that navigate, query and perform reasoning upon the linked data about historical persons need to take account of the limitations to the extrapolation based on such statements of identity and co-reference. Where two person-records have been unambiguously merged into a new resource, all available data can be combined without problem. A simple example is that of Ailia Pithia above, and one prosopography might record her family relationships in great detail, while the other gave higher precision for her dating and religious titulature. Would any such combination of data be sound in the case of Iulius Proculus, or would there be too great a danger of automated reasoning leading to misleading or impossible assumptions being codified in the data? At the very least, the uncertainty and contingency of the relationship would need to be inherited by all extrapolated data.

3. Case study 2: Associated Information and Disambiguation

As we expand the possibilities of mapping overlap between datasets even in the absence of direct co-references in the source prosopographies, contextualising identifiers become key to highlighting – although not proving – possible matches. This is not a problem unique to classical prosopographical data: research that requires entity extraction and mapping across modern social networks addresses comparable questions. Anonymization and de-anonymization of person-data also enjoy increasing attention because of the greater prominence of ethical questions.[18] The difficulties of mapping entity records even when we have co-reference points in the source text may be similar. But apart from the lack of privacy-related ethical issues for the Classical period, a key difference between dealing with ancient and modern data is how the factors supporting the cross-referencing are regarded.

Researchers dealing with the more modern data often have both significantly more datapoints to work with and the advantage of having the entities largely pre-defined as distinct records. The integration of disparate datasets to identify and extract missing data can be seen as an exemplar of the potential of incomplete person-data, data that is 'dirty by design'. The key question for us at this stage is not the computational algorithm with which to create the mappings, but what information is needed to support those processes within the constraints that ancient world data brings.

The SNAP model identifies the following facets of information as being pivotal for the automatic identification of duplicate entities: Name; Titles/occupations/epithets; Associated dates; Associated places. In addition the model records relationships between entities to allow additional reasoning, e.g. if we know Person A gave birth to Person B and the Person C was Person B's maternal uncle, we can deduce that Person A and Person C were siblings. In the following section we will consider some of the questions related to these categories, as an example of a lightweight mapping approach, and how they feature in a cross-project problem space.

---

18   Examples of works in this area include Sweeney (2002), Aggarwal (2005), de Montjoye et al. (2013).

Names

Despite their great existential relevance, names – when we have them – are recognised as not being unique identifiers. The addition of epithets and other descriptive features helps modern disambiguation much as it must have done in antiquity: Pliny the Elder is a different person to Pliny the Younger, Apollonios the poet from Rhodes is not the same individual as Apollonios the philosopher from Tyana. Even when they are acting as quasi-identifiers – e.g. 'Pliny the Elder' – how the name is recorded can detrimentally affect record mapping. The issue is not only in defining what we classify as being part of a 'name' (e.g. Is any epithet included and under what conditions?), but how that information is then recorded. This is partly dependent on the type of onomastic data available, but also on the choices made by a given project about the processing, normalising and storage of the data: e.g. is the name stored as one string or broken into components? Is the name given in the language and script of the text? Is it given as written in the source or normalised? etc.

Titles

The datasets that contribute person-data to the SNAP graph offer little or no consistency in the taxonomy of terms for titles or relationships, in part due to differences in the sources. 'Titles' can be broadly defined as more or less standardized (sets of) words providing information about someone's social position, function, geographical context or even genealogical ties. They often lie at the heart of a prosopography, especially those focusing on the holders of a given position such as Emperor or Consul. They are often equally crucial in wider datasets due to their ability to tie people together. Yet they can also be confusing, as only the pragmatic context allows the determination of their true meaning: addressing someone as your father in a letter is in most cases and periods just a polite phrase, but in a contract this really points to a genealogical connection.[19] Moreover, some titles can be used in a broad sense or with a more restricted meaning: in Ptolemaic Egypt 'royal scribe' (*basilikos grammateus*) can point to the main official in a nome`s administration, but it can also just refer to people that are part of the administration in general.[20] Further, it is not unusual for a person to take or be given multiple titles or epithets over the course of their life. As a result they may be referred to by one title in one source, but named and associated with another title elsewhere. The recording of variant titles both within and across projects will have the further benefit of creating a de facto thesaurus of titulature, which will improve discovery across the larger network of prosopographical authorities, and potentially reveal new links between data. Even when an exact match is not possible, the association of an entity with a specific group of like-titled entities can surface possible matches. While these type of 'fuzzy' matches should not be implemented automatically, the possibility of identifying clusters of entities around given titles or roles opens avenues for the academic to focus their research on potentially overlapping records.

In considering titles a useful, although not necessarily defining, disambiguator, we return to the previously mentioned issue of cross-project consistency. From a computational perspective the ideal would be to have a defined taxonomy of all titles and epithets, with strict rules as to how each would be recorded and used. This would not only place an intolerable burden on ongoing projects, however, but would also make it extremely difficult to include information from published completed projects that are merely being maintained. In this we clearly see the conflict between the desire to impose order and the need to both acknowledge the practicalities of the situation – simpler is often more useable if not 'better' – and the expertise within a project as it relates to their sources. As projects increasingly think outwards and consider existing taxonomies during their development stage, we may see a trend towards consensus. This

---

19   Dickey (2004).

20   Clarysse (1978).

process, however, relies on the sharing of data and the understanding of the taxonomies and normalisations used in the creation of that data. In this respect, the SNAP model facilitates the sharing of data, in the form of a given project's normalised terms. As more mappings are made, SNAP also provides a platform through which the folksonomy of titles, epithets and other such descriptors can be developed and explored. At the same time SNAP promotes awareness of project choices within the larger community context.

Associated Places and Dates
The final facets that SNAP utilises are associated date and associated place. These markers may overlap with the epithets discussed above and are used to encode any significant time or place related to the entity. In many cases, especially with regard to date, the link is with particular occurrences such as birth or death. The association is intentionally undefined, however, to allow projects to select the most appropriate date range or locations based on the available data, which may well be minimal. Zenon son of Zenon would be associated with Aphrodisias because that is where the inscription in his honour (IAph 13.152) was dedicated,[21] while Publius Cornelius Scipio Africanus might be associated with Africa having gained the epithet "Africanus" in 201 BCE following his defeat of Hannibal at Zama.[22] As well as leaving the connection between any temporal or geographical information open, the granularity of the data is not predetermined by the model either. M. Aurelius [·· ? ··]os (IAph 12.215) might be linked to a specific site, such as his home city Aphrodisias; he might also be associated with Nicomedia or Ancyra where he both won races and held citizenship; or with Hadrianea, Heraclea on the Pontus, Chalcedon, Nicaea and Philadelphia where he won races.[23] The associated date is always given as a date range, but depending on the source database could exactly define the entity's lifespan (known birth and death dates), reflect the date of a grave monument (and therefore presumably death), or provide an estimation of the age of the archaeological context of the source, when no other information is available. For example, the previously mentioned M. Aurelius [·· ? ··]os (IAph 12.215) might have associated date ranges of 200–250 (from LGPN) or 211–233 (from *Inscriptions of Aphrodisias*), based on the same evidence: the contests that he won and Roman citizenship implied by his tria nomina. This flexibility sets the SNAP model apart from similar but more structured models such as CIDOC-CRM[24]. While these are widely used in cultural heritage and allow the description of similar information, they are predicated by the difficulties and inexactness of much of the data being collected. Creating a more structured system would add an additional barrier on projects in sharing the data and bring an illusion of precision, which would not only be unwarranted but potentially misleading. The lightweight approach makes it easier for data to be shared and reflects its reality.

Stronger Together
None of the facets listed above offer a full solution to the problem of mapping between entity records but the combination of quasi-identifiers allows for greater reduction in potential matches, even when the specificity of the data is low. The deliberate lack of control emphasises the vital truth that any alignment discovered, especially through automated methods, only reveals the potentiality of alignment. We know that the datasets that we are working with are incomplete and, even if the data was complete, there is redundancy in the human population and similarity of information[25] disguises the accurate separation of entities.

---

21   Reynolds et al. (2007) 13.152.

22   Broughton (1951) 320–21.

23   Reynolds et al. (2007) 12.215.

24   Doerr (2003); Crofts et al. (2006).

25   Sweeney (2002)'s ‚k-anonymity' in modern data anonymisation terms.

4. Case study 3: gods and cultic epithets

SNAP includes prosopographies and name lists of literary and mythological figures, including divine and semi-divine heroes, and even gods, within its scope. This introduces a new dimension of problems in disambiguation and co-reference. Mythological persons usually lack the crucial disambiguating factor of date, and more critically there is no expectation that different sources will be consistent in their attestation of titles, occupations, geography and relationships – to the extent that in some cases different poetic versions of a named figure might arguably not be merged into a single coherent individual at all.

Recording person-data of this type in machine-actionable formats such as RDF brings in sophistication and spurious exactitude.[26] The triples of which RDF consists, connect two entities through formalised statements, systematized in so-called ontologies (e.g. OWL).[27] To deal with the ambiguous identities of mythological figures, we clearly need to be able to express relationships between two person-records that are more granular than "these information resources are unambiguously about the same person-record" (snap:MergedResource), and certainly more expressive than "these entities should be considered to be functionally identical" (owl:sameAs). The human mind is able to cope with two versions of Odysseus' death and recognise both as in some sense referring to the same Odysseus, without considering one or both of them wrong. The difficulty, and we would argue impossibility, would be to express this fluidity and ambiguity in an OWL ontology.[28]

The problem is compounded in the case of gods in local cult and practice. Modern scholars are often not in complete agreement about how best to express the relationship between a literary or panhellenic deity on the one hand, and its many local or specialist variants, each with differing attributes, epithets or even names. Evidence for divine epithets exists in several forms: in poetry (notably the Panhellenic epics of Homer and Hesiod), the names, characteristics, spheres of influence and attributes of gods and goddesses are highly formulaic and clearly very traditional; in the names of particular cult sites, often influenced by local practice or history, temples and shrines are often named for deities whose name-plus-epithet combination seem to differentiate them from other incarnations of the "same" deity; in certain ritual contexts, for example oaths, sentences or curses in legal practice, deities are invoked with specific epithets (for which in some cases there is no cultic or literary evidence).

The relationship between these epithets, functions, contexts, locations and the deities they denote are sometimes hard to disentangle. An archaeologist or historian of Greek religion will often talk as if "Poseidon Soter of Sunion" and "Poseidon Isthmios" are two separate entities, as they are two separate cults; a reader of Homer might be surprised to learn they are not both temples of the Earthshaker, brother of Zeus, patron of Troy they are so familiar with. Both positions are of course correct, in their contexts, and no doubt the ancients had some way of resolving these apparent contradictions in their religious world-view. With reference to Artemis, Ivana Petrovic expresses one interpretation: "To a degree, it was possible to merge the Homeric goddess with the local Athenian Artemis, and to adapt the picture of the goddess to the cultic reality."[29] Is this a description of one goddess, or two?

---

26   Resources Description Framework: <https://en.wikipedia.org/wiki/Resource_Description_Framework>.

27   For ontologies, see <https://en.wikipedia.org/wiki/Ontology_engineering>; for OWL, see <https://en.wikipedia./wiki/Web_Ontology_Language>.

28   For an ontology describing conflicting narratives with reference to fictional characters, see Lawrence et al. (2010).

29   Petrovic (2010) 221.

Jon Mikalson, with reference to three aspects of Poseidon (*Soter, Hippios, Asphaleios*), does not attempt to hide the twenty-first century reader's difficulty in resolving ancient religious thought: "To us they might appear as three separate gods … but the Greeks, for reasons about which we can only speculate, brought all three together under the name Poseidon."[30] The scholarly agonising over the identity or non-identity of god-cult combinations is by no means close to resolution, and is not restricted to understanding ancient thought. The aetiology of the association of several aspects or natures under a single god's name is another concern, as Ken Dowden points out with reference to Zeus: "Some have thought the Meilichios functions are so separate from others that they originally belonged to a separate god."[31]

Zeus alone is attested in myth or cult with dozens of individual epithets or aspects.[32] The Homeric or Panhellenic Zeus has certain concerns, powers and family relationships, and would have been recognisable to all Greeks by most of these attributes: father of the gods, bringer of justice, wielder of lightning bolts, defender of social mores. In our historical sources, Zeus *Hikesios* (of suppliants) was invoked in legal oath-swearing contexts,[33] and occasionally in local cult;[34] Zeus *Xenios* (of guests/hosts or foreigners) could be invoked to punish transgressors.[35] To add to the confusion, a Zeus *Chthonios* (of the underworld) is sometimes invoked in epic, and in magical or necromantic ritual, but this usually seems to be a reference to Hades.[36]

A prosopography of Greek cult deities might therefore list several dozen Zeuses, each with a separate, unique identifier. At some level these would all (or mostly) have some relationship to a Panhellenic, literary, or "Platonic ideal" Zeus, as indeed they did in Greek thought, but that relationship is not one of clear, unambiguous identity – nor indeed of disputed or uncertain identity. Rather it is a different nature of question than whether Diogenes mentioned in a late third-century tax return, and Diogenes who appears on an early fourth-century tombstone, are the same man or not, and should be expressed with a different RDF property.

It is already difficult, as discussed in our first case study above, to extrapolate relationships and other indirect information based on the unclear identification of two person references: if person A is an *Archiereus*, and person B is probably the same person, then should the record for person B also return the title *Archiereus*? With the even more tortured classes of relationship and identity between divinities, and the greater difficulty in understanding and assigning titles and epithets, this sort of reasoning across the linked data needs to be filled with extremely careful caveats and unambiguous citation of references, if indeed it has any utility at all.

Do relationships and characteristics associated with the record for Zeus also apply to the record for Jupiter, who is almost universally recognized as the same deity under a different name, in a different language? Only with extreme caution, any scholar would recognize. Clearly even more caution and qualification is needed to avoid absurd conclusions with cross-cultural deity identification: the Romans in particular were prone to linking their gods with those of neighbouring peoples, both in cult naming and in colonial propaganda. Although Zeus/Jupiter seems

---

30   Mikalson (2010) 32.

31   Dowden (2006) 66.

32   Burkert (1985) 130.

33   Cassella d'Amare (2005).

34   E.g. IG XII 4, 1225.

35   See Dee (1994) s.v. B.23 (Ξένιος/Ξείνιος).

36   E.g. Hom. Il. 9.457; Hes. Op. 465; PGM XXIII.3.

Bodard, Cayless, Depauw, Isaksen, Lawrence, Rahtz: Standards for Networking Ancient Person-data:
Digital approaches to problems in prosopographical space

37

DCO 3,2 (2017)

to be derived from the same Proto-Indo-European divine name *Dyeus Piter* as the Germanic Teíws or the Norse Týr,[37] the Romans later linked Jupiter to Thunor/Thor, no doubt due to the association of both with thunderstorms and the sky.[38] An unsophisticated algorithm working on the basis of this identification might therefore conclude that the same figure had a father named Wotan/Odin, and a son named Mercury, based on the family relationships of the two well-known mythologies. A further step in the algorithm would then conclude that Mercury is the same individual as Wotan/Odin, as later authors identified them;[39] Odin is therefore not only the father of Thor, but his son, and so his own grandfather. This has been a reductio ad absurdum, of course, but it illustrates the danger of crudely applying a prosopographical data structure designed for historical figures to person-like figures of different types. Similarly, when transferring relationships between databases across disciplines, time periods, places and cultures, one should be cautious.

The first and to some extent second case study therefore presented problems we can see solutions to, or at least ways in which digital encoding and linking data between projects may aid historians in addressing prosopographical questions. With this third case study our main concern is to capture person-data and relationships in such a way as not to mislead or misrepresent the complex issues in ancient religious scholarship. This may in turn lead to networks or visualizations that help to express or communicate some of the complex issues involved in the identities of divinities, but we do not expect digitization will solve problems that are as old as antiquity itself.[40]

5. Conclusions

The three case studies above present problems of different kinds, and which SNAP will address in different ways. The first is a problem that has a partial solution, the recording of ambiguous or qualified identity between person-records, but for which we do not attempt to solve the ambiguity itself beyond providing structured data for historians to work with. The second is a more significant problem, that of inconsistent and fluid terminology for titulature and occupations, the solution to which should at least partly come in the form of taxonomies and ontologies, but the difficulty in implementing such solutions will be considerable. The third is the most tricky issue, that of representing identities and relationships between figures that are barely understood or agreed upon by scholarship, and where the pressing need is to record these problems clearly.

The methodology we have presented offers a way to bring together existing datasets, and which can be used by new projects to share and interchange scholarly information. Further it creates a platform for further research that will take account of and track changes in our understanding of the sources with the addition of new data, and widers contexts for the analysis of existing data. The model, while lightweight to ensure ease of use and compatibility,

---

37   Burkert (1985) 125–6.

38   Brown (2000) 57.

39   Tac. Ger. 9; on which see further Birley (1999) 107.

40   For an impressive visualisation of the genealogy of ancient Greek mythology by D. Macek, see <http://www.myth-gen.eu/> (information courtesy anonymous reviewer).

creates potential for reasoning across a wide network, to supplement the analytical exploration of sources. Through this analysis and sharing of information we encourage reflection on the inconsistencies in data and in our practice, and the unavoidable contradictions that arise in both of these areas, in the hope that scholars will thereby reach new or improved understanding.

We neither claim, nor present, the SNAP model as a solution to the historical problems discussed here. These are not problems to which there are purely programmatic solutions, and trying to impose them would both be fruitless, and lead to loss of the necessary complexity in the data. Rather, we have shown how such a model can support existing research methodologies and break down the barriers between datasets.

The SNAP model represents but a piece in a larger puzzle, which includes Pelagios, LAWD and Linked Pasts,[41] and is designed to support research, but also relies on engagement with and from the traditional scholarly community. Classicists and historians need to be involved in this research, not only to ensure that the assumptions behind it are sound, but also so that the questions being asked are those that serve historical research into people and identities. Scholars who produce editions of ancient texts also need to be engaged in this work: the unique identifiers in the virtual person authority list are an essential part of the apparatus of reference and disambiguation in digital (and even print) texts, and perhaps even more importantly, the use of SNAP identifiers in the annotation of text editions and other databases will create a massive citation network that will lead to further improvement in the understanding of – and offer new research questions on – people of the ancient world.

We invite historians, including those, whose natural interests do not necessarily include Linked Open Data and information science, to engage with this project as providers of data, as users of data and research tools, and as a sanity check to ground the informatic work in historical needs and scholarship. Prosopographers, who even at their most traditional are familiar with structuring and normalizing data, are the best ambassadors to the rest of the classical, archaeological and historical community of the value of potential of the work we are describing in this paper.

---

41   Pelagios and Linked Pasts: <http://commons.pelagios.org/>; LAWD: <http://lawd.info/>.

6. References

Aggarwal (2005): Charu C. Aggarwal, "On k-anonymity and the curse of dimensionality", in: VLDB '05: Proceedings of the 31st international conference on Very large data bases, Trondheim, 901–909. Available: <http://www.vldb2005.org/program/paper/fri/p901-aggarwal.pdf>.

Alföldy (1977): Géza Alföldy, Konsulat und Senatorenstand unter den Antoninen, Bonn.

Berners-Lee (2006): Tim Berners-Lee, "Linked Data", Design Issues. Available: <https://www.w3.org/DesignIssues/LinkedData.html>.

Birley (1999): Anthony R. Birley, Tacitus, Agricola and Germany, Oxford.

Bodard et al. (2014): Gabriel Bodard, Hugh Cayless, Mark Depauw, Leif Isaksen, K. Faith Lawrence and Sebastian Rahtz, SNAP:DRGN Cookbook. Available: <http://snapdrgn.net/cookbook>.

Broughton (1951): T. Robert S. Broughton, The Magistrates of the Roman Republic, New York.

Brown (2000): John Pairman Brown, Sacred Institutions with Roman Counterparts, Berlin.

Burkert (1985): Walter Burkert, Greek Religion: Archaic and Classical, translated by John Raffan, Cambridge MA.

Cassella d'Amare (2005): Paola Cassella d'Amare, "La denominazione di Zeus Ἱκέσιος con particolare riferimento alla tragedia", in: Nicole Belayche et al. (edd.), Nommer les Dieux: Théonymes, épithètes, épiclèses dans l'Antiquité, Rennes, 121–128.

Clarysse (1978): Willy Clarysse, "Notes on Some Graeco-Demotic Surety Contracts", Enchoria 8.2, 5 – 8.

Crofts et al. (2006): Nick Crofts, Martin Doerr, Tony Gill, Stephen Stead and Matthew Stiff, Definition of the CIDOC Conceptual Reference Model, Paris. Available: <http://www.cidoc-crm.org/docs/cidoc_crm_version_4.2.1.pdf>.

Dee (1994): James H. Dee, The epithetic phrases for the Homeric gods: a repertory of the descriptive expressions for the divinities of the Iliad and the Odyssey, New York and London.

Depauw/Van Beek (2009): Mark Depauw and Bart Van Beek, "People in Greek Documentary Papyri: First Results of a Research Project", Journal of Juristic Papyrology 39, 31–47.

Devijver (1989 – 1992): Hubert Devijver, The Equestrian Officers of the Roman Imperial Army, Stuttgart.

Devijver (1973 – 2001): Hubert Devijver, Prosopographia Militiarum Equestrium quae fuerunt ab Augusto ad Gallienum, Leuven.

Dickey (2004): Eleanor Dickey, "Literal and extended use of kinship terms in documentary papyri", Mnemosyne 57, 131–176.

Bodard, Cayless, Depauw, Isaksen, Lawrence, Rahtz: Standards for Networking Ancient Person-data: Digital approaches to problems in prosopographical space

40

DCO 3,2 (2017)

Doerr (2003): Martin Doerr, "The CIDOC CRM – An Ontological Approach to Semantic Interoperability of Metadata", AI Magazine 24.3. Available: <http://www.cidoc-crm.org/docs/ontological_approach.pdf>.

Dowden (2006): Ken Dowden, Zeus, London and New York.

Elliott et al. (2014): Thomas Elliott, Sebastian Heath and John Muccigrosso, Current Practice in Linked Open Data for the Ancient World. ISAW Papers 7, New York. Available: <http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/>.

Groag et al. (1933–2016): Edmund Groag, Arthur Stein, Leivia Petersen, Klaus Wachtel, Matthäus Heil, Johannes Heinrichs, Marietta Horster, Andreas Krieckhaus, Anika Strobach and Werner Eck, Prosopographia Imperii Romani saec. I. II. III, Berlin.

Huß (2001): Werner Huß, Ägypten in hellenistischer Zeit 332–30 v. Chr., München.

Isaksen et al. (2014): Leif Isaksen, Elton Barker, Rainer Simon and Pau de Soto, "Pelagios and the Emerging Graph of Ancient World Data", in: WebSci'14 Proceedings of the ACM Conference on Web Science, 22–26 June 2014, Bloomington IN, 197–201. Available: <http://oro.open.ac.uk/43658/1/2014_Isaksen_Barker_etal_Pelagios_WebSci.pdf>.

Janiszewski et al. (2014): Paweł Janiszewski, Krystyna Stebnicka, Elżbieta Szabat, Dorota Dzierzbicka, Prosopography of Greek Rhetors and Sophists of the Roman Empire, Oxford.

Keats-Rohan (2007): Katharine S.B. Keats-Rohan (ed.), Prosopography: Approaches and Applications. A Handbook, Oxford.

Lawrence et al. (2010): K. Faith Lawrence, Michael O. Jewell and Paul Risse, "OntoMedia: Telling Stories to Your Computer", in: Proceedings of the First International AMICUS Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts, Szeged. Available: <http://ilk.uvt.nl/amicus/WS01/faith_lawrence_ea.pdf>.

Lawrence/Bodard (2015): K. Faith Lawrence and Gabriel Bodard (2015). "Prosopography is Greek for Facebook: The SNAP:DRGN Project." Proceedings of the ACM Web Science Conference Web Sci 15. Available: <http://dx.doi.org/10.1145/2786451.2786496>.

Mikalson (2010); Jon D. Mikalson, Ancient Greek Religion, 2nd edition, Chichester.

De Montjoye et al. (2013): Yves-Alexandre de Montjoye, César A. Hidalgo, Michel Verleysen and Vincent D. Blondel, "Unique in the Crowd: The privacy bounds of human mobility", Scientific Reports 3, article number: 1376. Available: <http://www.nature.com/articles/srep01376>.

Mooren (2001): Leon Mooren, "The automatization of the Prosopographia Ptolemaica", in: Isabella Andorlini et al. (edd.), Atti del XXII Congresso Internazionale di Papirologia, Firenze, 23–29 agosto 1998, Firenze, 995–1008.

Petrovic (2010): Ivana Petrovic, "Transforming Artemis: from the Goddess of the Outdoors to City Goddess", in: Jan N. Bremmer and Andrew Erskine (edd.), The Gods of Ancient Greece: Identities and Transformations, Edinburgh, 209–227.

Bodard, Cayless, Depauw, Isaksen, Lawrence, Rahtz: Standards for Networking Ancient Person-data: Digital approaches to problems in prosopographical space

41

DCO 3,2 (2017)

Reynolds et al. (2007): Joyce Reynolds, Charlotte Roueché and Gabriel Bodard, Inscriptions of Aphrodisias. Available: <http://insaph.kcl.ac.uk/iaph2007>.

Sweeney (2002): Latanya Sweeney, "k-anonymity: a model for protecting privacy", International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10 (5), 557–570. Available: <http://dataprivacylab.org/dataprivacy/projects/kanonymity/kanonymity.pdf>.

Verboven et al. (2007): Koen Verboven, Myriam Carlier and Jan Dumolyn (2007), "A short manual to the art of prosopography", in: Katharine Keats-Rohan (ed.), Prosopography: Approaches and Applications. A Handbook, Oxford, 35–70.

Bodard, Cayless, Depauw, Isaksen, Lawrence, Rahtz: Standards for Networking Ancient Person-data: Digital approaches to problems in prosopographical space

42

DCO 3,2 (2017)

Authors[42]

Dr. Gabriel Bodard
University of London
Institute of Classical Studies/ School of Advanced Study
Senate House
Malet Street
London WC1E 7HU

Email: gabriel.bodard@sas.ac.uk

Hugh Cayless, Ph.D.
Software Development & Integration Services
Duke Collaboratory for Classics Computing
411 Chapel Drive
Durham, NC 27708

Email: hugh.cayless@duke.edu

Prof. Mark Depauw
KU Leuven
Ancient History
Blijde Inkomststraat 21 – box 3307
3000 Leuven

Email: mark.depauw@kuleuven.be

Dr. Leif Isaksen
Lancaster University
Department of History
Bowland College
LA1 4YT
Lancaster

Email: l.isaksen@lancaster.ac.uk

Dr. K. Faith Lawrence
Ghent Centre for Digital Humanities and Archaeology Department of Ghent University (Belgium)
King's College London
Digital Humanities
Room 218, 26 – 29 Drury Lane
London
WC2B 5RL

Email: faith.lawrence@kcl.ac.uk

Sebastian Rahtz[†]
(1955 – 2016) Chief Data Architect, Academic IT Group, Oxford University Computer Services

---

Romans 1by1 v.1.1
New developments in the study of Roman population

Rada Varga

Abstract: The proposed paper documents *Romans 1by1*, a population database working on Roman-era inscriptions. The database architecture is built for accommodating all categories of people attested epigraphically. Besides the 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. 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.

1. Introduction

This paper documents *Romans 1by1*, a population database working with Roman-era inscriptions. The first general presentation of the database and its metadata were published[1] in the proceedings of the EAGLE 2016[2] international conference on *Digital and Traditional Epigraphy in Context*. But as this is a very dynamic tool, its configuration has considerably changed since August 2015 (when the article was sent for publishing). Besides the database architecture, we will currently present the first practical employments of *Romans 1by1*: the HISCO adaptation and encoding for epigraphically attested professions and some first phase network studies.

2. Database architecture

Romans 1by1 is a relational database, built with MySQL and following the best practice models for population databases.[3] We began with a metadata schema adapted for registering members of the middle classes coming from the Latin language provinces of the Roman Empire, but expanded it in the fall of 2015 to accommodate all individuals epigraphically attested (at least) in Roman Dacia, Moesia Inferior and Moesia Superior.[4] Thus the epigraphic categories diversified, the attributes required for registering life details (*cursus honorum*) became more complex

---

1   Version 1.0.

2   Varga (in print).

3   Mandemakers, Dillon (2004).

4   For the sources we use the main online inscription databases (p://www.manfredclauss.de/; http://edh-www.adw.uni-heidelberg.de/home; http://inscriptions.packhum.org/) and the published epigraphic corpora: CIL, AE and regional compendia.

and the linguistic scope extended due to the Greek inscriptions of Moesia Inferior – bringing forth new epigraphic habits and patterns as well. At[5] the day of writing (December 2016) the metadata (see Fig. 1) is structured into four major tables (Inscriptions, Bibliography, Personal Data and Personal relationships – Fig. 2), summing up to more than 100 separate attributes and over 8.000 individuals registered.[6]

Figure 1: Diagram of the metadata

Figure 2: Database tables structur

The first table to be filled in is the file of the source – *Inscriptions* (Fig. 3). To begin with, each inscription gets an identification code, formed of 5 digits and an acronym of the province's name (D for Dacia, MS for Moesia Superior, DAL for Dalmatia, etc.) – so we have, for example 00001MS. The form for the *Inscriptions* category contains free text fields (*Relevant expressions, Stylistic details, Atypical features, Observations, Place of discovery, Place of provenience, Ancient name provenience, Timestamp/Timeframe and External links*), as well as fields with controlled value lists (*Type of inscription, Language, Material*).[7] At this point, linking our fields to the already existing online resources is very important. Currently, we are working on linking the toponyms to Pleiades[8] and/or Trismegistos.[9] Although we are aware that some of the data (*Timestamp/Timeframe*) could have benefitted from a standardized entry form,[10] we opted in this case for more flexible possibilities of expression and adaptation. Making this field operational and easy to work with is a very important desideratum, as temporal data is paramount for all cross-sectional studies we would undertake. The *Inscription bibliography* section was conceived in a way so that extracting complete or selective bibliographical lists would be possible. Thus, a normalization table includes all bibliographical titles referred to and being cited; with the help of a value list, one can choose one *Bibliography abbreviation* for which the full reference, detailed information and comments are shown. Of course, all data are linked to the *Inscription code*, selected as well from a value list.

---

7   For these, we are mainly using the Eagle vocabularies: http://www.eagle-network.eu/resources/vocabularies.

8   https://pleiades.stoa.org/.

9   http://www.trismegistos.org/.

10   Standardization will be undertaken when the data will be made open access.

Figure 3: The Inscriptions table

The core of the database is a table used for recording data about individual persons (labelled *Personal data*– Fig. 4), around which the entire network of relations needed to ensure proper information recording is built. Each new entry represents a singular epigraphic attestation of an individual, and a unique ID is generated, which will help to link the character within 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 case of one person being attested by multiple epigraphic sources, each attestation will be represented by a new entry, to which a new unique ID will be assigned and which will be doubled during linkage procedures by a common ID for all instances of the same person.

Expectedly, this metadata is the most volatile one, being in a continuous process of modification and enlargement, although some attributes remained unchanged since the initial implementation of the database[11] (*Praenomen, Nomen, Cognomen/Personal name, Father/Master name, Agnomen, Signum, Ethnicity* for each name category). As a principle, we are trying to remain faithful to the source and to record during the first phase only the minimum of deduced information but nonetheless we are operating with some conventional onomastic standardizations (e.g. for AUR, we have recorded Aurelius from the beginning). As we considered them important in the analysis of (self-assumed) identities and epigraphic habits, we opted for a detailed registration of geographic origin and geo-political appurtenance details; thus, we have *Natione, Ethnicity, Origo, Domus* and *Local citizenship* as separate attributes. As already stated,[12] some information will be recorded deductively: such are the cases of *Gender* and *Juridical status* (though the servile one often is literally recorded). The rest of the fields accommodate supplementary information, if needed: *Occupation, Deities, Age* (at death), *Details of life/death* and *Observations*. For *Collegium/Association* and *Religious position* we opted for check-boxes

---

11   And are documented as such: Varga (in print).

12   Varga (in print).

which open a series of fields. For the associative forms we have three free text boxes: *Type of association, Position within the association, Activities within the association* which allow for certain flexibility and the possibility to accommodate information and realities from both Roman and Greek inscriptions. In the case of *Religious position*, the situation is a bit more complex: when checking this option, a drop down list of *Sacerdotal office* opens whose values are (at the moment) *Augur*, *Flamen* and *Pontifex* and a series of further check-boxes: *Coloniae/ Municipii sacerdos*, *Military sacerdos*, *Imperial priest*, *Divinity priest*, *Laurentium Lavinatium*. All of them, with the exception of the last one, open a *Details* text field for description. We decided to register *Laurentium Lavinatium* separately as it cannot be associated with other religious positions and it has been recorded on several inscriptions from Ulpia Traiana Sarmizegetusa (Dacia).

Figure 4: The Personal data table

The second half of the *Personal information* table is dedicated to the political and social status of the elites. This part of the metadata consists of a series of check-boxes each opening different categories of specialized attributes (Fig. 5). *Ordo senatorius* and *Ordo equester* each open two text fields, concerning *Details* and *Cursus honorum*. Though a standardization of the various steps of the *cursus honorum* was thought to be an option at some point, due to the great diversity of possible existing functions and positions we decided for this descriptive approach.

Figure 5: The tables covering the political/social status of the elites

In a following phase we will standardize these data entries as much as possible.

The next two check-boxes *Provincial Governor* and *Procurator* both open attributes for *Details*, but also two further options marking if we are dealing with a Governor of *Praetorian* or *Consular Rank*, respectively a *Praesidial* or a *Financial Procurator*. The *Local magistrate* check-box again contains a details field. While so far we haven't generally faced major problems with modelling these attributes, the *Decurionate* option raised some issues connected to the possibility of multiple decurionates. Our first idea was to give the possibility to choose multiple options in the *City/Town* table with each option being described with a separate *Details* field. But this would have caused problems with creating the .csv files exported from the database, and later with the import in statistical software (R) or even Excel spreadsheets. Trying to avoid these disturbances, we opted instead for another check-box (*Multiple decurionate*) with a field in which the city names can be entered as free text. Even though this solution will impose restrictions for some types of analyses, we believe that the final export file will not require too much additional systematization work in the case of a research focused on decurions. On the other hand, for the external user this will not constitute an inconvenience as one will have all the available search options necessary for finding the relevant information: the possibility of choosing *Multiple decurionate*, of searching for the registered city names and – eventually – of using the general search field. The next information items cover *Imperial slave*, *Imperial freedman* and the *Military personnel*, all equipped with *Details* attributes. Only the

last category needs additional options: two text fields for details and career and two drop-down menus for rank and unit. The final field refers to *Decorations* that we considered necessary to being recorded separately.

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 (Fig. 6).
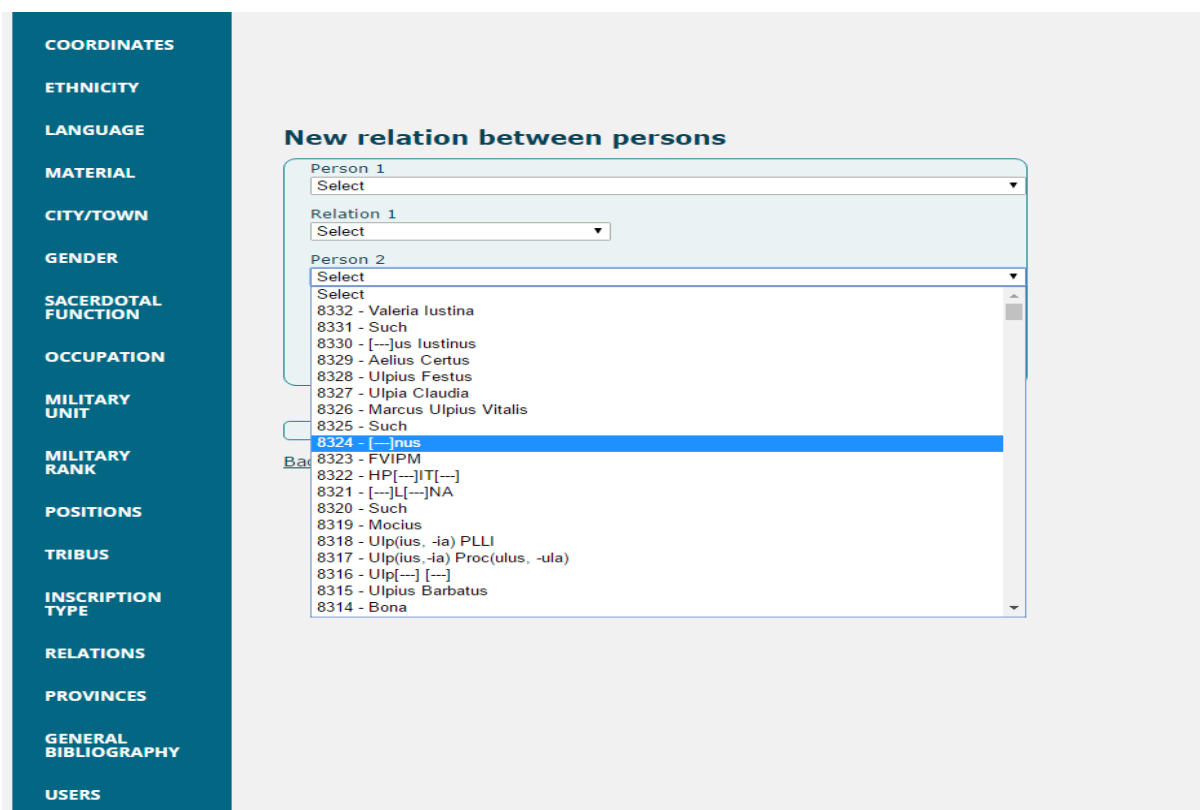
Figure 6: The Relationships table

The relationship values have been encoded (see table 1) and we have tried to adjust the concepts to the SnapDrgn ontology.[13] We have assigned 1- figures for 1st degree family, 2- for 2nd degree family, 3- for more distant and/or not blood kin, 4- for the *familia Romana* (slaves, clients, alumni, etc.) and 5- for unspecified or unreadable relationships. For males we use odd numbers and for females even numbers. Encoding them from the start was very useful for analyses on dedication habits, gender-related patterns and expressions of family connections.

101	HusbandOf
102	WifeOf
103	FatherOf
104	MotherOf
105	SonOf
106	DaughterOf
201	BrotherOf
202	SisterOf
203	GrandfatherOf

---

13   https://snapdrgn.net/ontology.

204	GrandmotherOf
205	GrandsonOf
206	GrandaughterOf
301	UncleOf
302	AuntOf
303	NieceOf
304	NephewOf
305	Cousin (M) Of
306	Cousin (F) Of
307	Son-in-lawOf
308	Daughter-in-lawOf
309	Father-in-lawOf
310	Mother-in-lawOf
311	Brother-in-lawOf
312	Sister-in-lawOf
401	Heir (M) Of
402	Heir (F) Of
403	Testamentary executor (M) Of
404	Testamentary executor (F) Of
405	Slave (M) Of
406	Slave (F) Of
407	FreedmanOf
408	FreedwomanOf
409	Patron (M) Of
410	Patron (F) Of
411	AlumnusOf
412	AlumnaOf
413	Co-worker/associate (M) Of
414	Co-worker/associate (F) Of
415	AmicusOf
416	AmicaOf
417	Partner (M)Of
418	Partner (F) Of
501	Unspecified relationship (M) Of
502	Unspecified relationship (F) Of
503	Unreadable relationship (M) Of
504	Unreadable relationship (F) Of

Very important for us was the search interface of the database (see in Fig. 7 the search options for personal data). Built with Ruby-on-rails, it was designed to respond to most manifold and varying search options. Every component of the database has search filters for every particular field, as well as a general search. For administrators and data entry operators there is also an *Author* search field but it is not accessible for external guest users.

Figure 7: The search options for personal data

3. Encoding of occupational titles

The codifications are an essential part for the whole project not only for the individual linkage procedures, but also for the analyzing process. Once the database will comprise enough data, properly recorded and with all codifications undertaken, statistical software can be used in order to identify trends and run comparisons over large scale geographical and administrative units. These results might lead to a better understanding of the social history of the Roman Empire.

At this point, as part of an already finished project,[14] we have encoded the epigraphically attested professions from the European Latin language provinces of the Empire. For this, we have used

---

14   *Carving a Professional Identity. The Occupational Epigraphy of the Roman Latin West*, Fritz Thyssen Stipend – http://romans1by1.com/doc/Carving.pdf.

and adapted HISCO[15] – *Historical International Standard Classification of Occupations*. The HISCO codes were later on classified in HISCLASS,[16] HISCAM[17] and SOCPO[18] – systems that put profession into relation with social status. Although efficient for the periods they were built for, they are completely irrelevant for the Roman era (servile status is ignored, it cannot be adjusted to the fact that the socio-economical high upper classes are basically excluded from any professional/occupational category, etc.). Thus we have decided not to use any of the social encoding systems.

Online and in print,[19] HISCO has been thoroughly documented and therefore we will not describe the details here in length. Basically, it offers a standard codification system for historical registered occupations – albeit only from Early Modernity up to nowadays. The system is based on the HISCO-tree, which contains 9 major professional groups:

> 0/1 – Professional, technical and related workers;
> 2 – Administrative and managerial workers;
> 3 – Clerical and related workers;
> 4 – Sales workers;
> 5 – Service workers;
> 6 – Agricultural, animal husbandry and forestry workers, fishermen and hunters;
> 7-8-9 – Production and related workers, transport equipment operators and labourers.

Each major group contains minor groups (01 – Physical scientists and related workers, 02 – Architects, engineers and related workers) unit groups (011 – Chemists, 012 – Physicists), and micro groups (01110 – Chemist general, 01120 – Organic chemist). Each occupational title from the unit group files is associated with a description and a link to equivalent occupational titles, in various languages. Of course, these are modern languages, directing the search towards databases dealing with occupations and/or demographic databases for the times indicated above.

Adapting the system for the occupational titles of the Roman period was rather painstaking, as certain *mutatis mutandis* processes had to take place. After normalizing the professions attested in inscriptions, we ended up with approximately 250 different professions, which is an unexpectedly high number. The number more or less corresponds to the one extracted from pre- and early-industrial era church registers for a given geographical region. The general conclusion of this encoding process, which will be presented and discussed in detail in a separate publication, is that most of the attested occupational titles belong to the HISCO major group number 4 (sales workers). Besides the large variety of types of sellers attested, this group also raised a serious problem and we were faced with the necessity of adapting the encoding system: for the Latin speaking part of the Roman empire it is often hard to make a clear-cut distinction (as HISCO does) between producer and seller. For example, the *ampullarius* most probably did not only produce flasks but also sold them, just as the *sagarus*, at least sometimes, also produced the cloaks he sold, etc. These kinds of situations are quite frequent and a decision is to be made regarding the encoding, as HISCO does not have codes for any type of producer-seller, let alone for all the multiple sub-specializations attested in inscriptions. The options were either adapting the codes and creating new ones, or assigning general and/or two codes. We have decided for the second option, as introducing new codes would have made the system more opaque for outside users. So, sometimes we assigned two codes: 4-32.00 (commercial

---

15   http://historyofwork.iisg.nl/.

16   Van Leeuwen, Maas (2011).

17   Lamber, Zijdeman, van Leeuwen (2013).

18   Van de Putte, Miles (2015).

19   Van Leeuwen, Maas (2002).

traveller code) or 4-90.90 (general salesman code), plus a specialized producer code. Coming back to one of the mentioned examples, for *ampullarius* we took the code 8-91.90 (Other Glass Formers, Cutters, Grinders and Finishers) plus 4-90.90 (Other Sales Workers – as we don't know anything about his status as a seller).

Other problems faced when encoding in HISCO were connected to the inner structure of the system itself. Thus, the medical specializations do not have separate codes – which, in our opinion, would be very useful.[20] Equally, the system provides no separate codes for itinerary artists[21] or for amateur teachers who occur in Greek inscriptions. In all these cases adjustments and integration into more general classes of codes were necessary.

In spite of the raised problems and the doubts faced during the encoding process, the codifications were completed (Fig. 8) and proved to be a useful step for analyzing the data. In the last section of this article a few examples which demonstrate the benefits will be discussed.
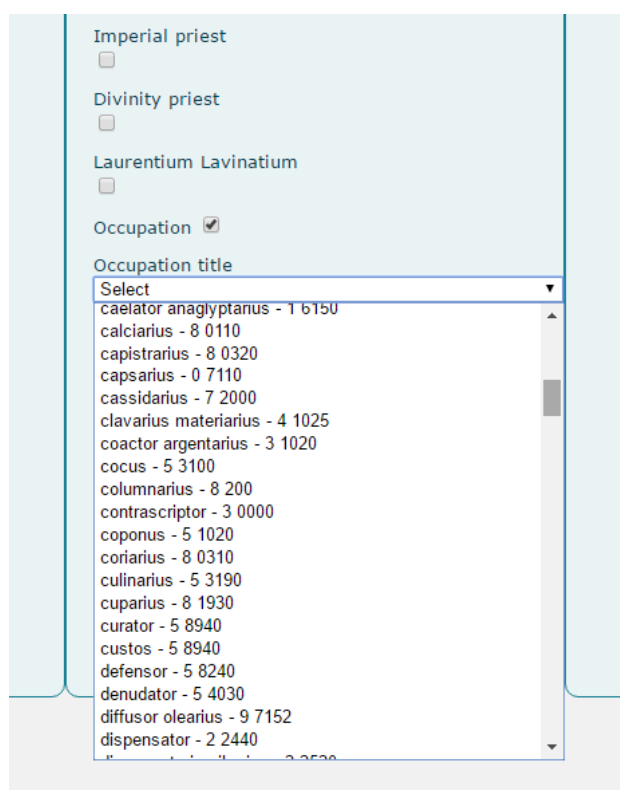

Figure 8: The occupational titles

4. Micro-historical reconstructions

One of prosopography's main stakes and most appealing features is the possibility to reconstruct history at a micro level: history of individuals, families, small groups. Of course, the *Romans1by1* database facilitates this kind of research enormously, as it helps to identify connections and possible similarities.

---

20 One of the databases linked to HISCO, *TRA preliminary version 1803–1970*, encodes ophthalmologists with 0-75.20, but the description of this code is: *Optometrist – Examines eyes and prescribes spectacles or treatment not involving the use of medicines, surgery or drugs, to conserve or improve vision*. Thus, the solution is one of compromise – http://historyof-work.iisg.nl/detail_hiswi.php?know_id=47007&lang=.

21 We have *pictor pelegrinus* attested.

For this purpose, we will present the example of Aurelius Aquila,[22] a *negotians* from Dacia, relocated in Salona and whom we identified on two different inscriptions, along with family members and business connections. He is attested on two inscriptions,[23] which previously were not related to each other. The arguments which support an identity of character are not only based on Aurelius Aquila's mere name. In the first inscription we have Aquila, decurio of Potaissa and *negotiator* from Dacia, erecting a funerary monument for his wife, Valeria Ursina, daughter of Titus, and for himself. The two spouses lived together without a single fight, the text says, for seven years. Even if the *decurio* had in mind to be buried next to/ together with his wife, he lived for many years longer, if we take into consideration the second monument. Here, Aquila buries a son,[24] Titus Aurelius Apollonius (who died at the age of 33 years and bears the name of his former father-in-law), together with two friends: Aurelius Flavus, *negotians natione Surus* and Aurelius Lucianus. The former is also present, along with an Aurelius Alexander, on a dedication for Iupiter Optimus Maximus Dolichenus from Apulum.[25] Expectedly, these two are the only epigraphic testimonies from the Roman Empire which attest Aurelius Flavus, *negotiator Surus*. In Fig. 9, one can see the ties identified through the research summed up above.

This small-scale prosopographical reconstruction of a network from the (most probable) beginning of the 3rd century AD, offers details on the great mobility of traders in the Roman Empire as well as on how strong and time-enduring business connections could be.
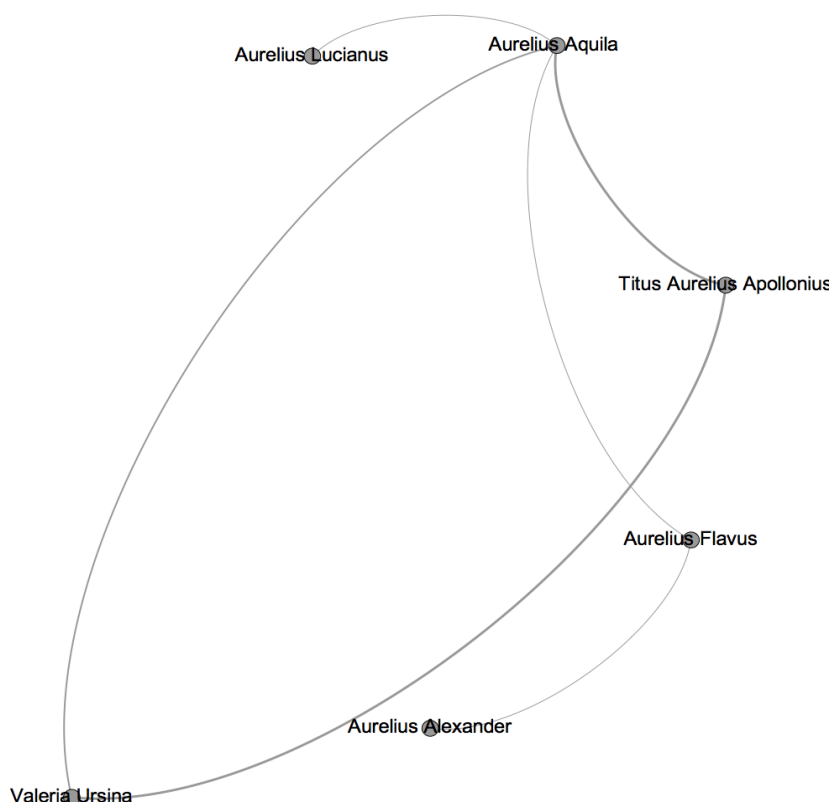
Figure 9: Aurelius Aquila's personal network

---

22   A detailed article on this is under press: Varga (2016).

23   CIL III 2086 (http://edh-www.adw.uni-heidelberg.de/edh/inschrift/HD058504; EDCS-ID: EDCS-27700136;) and CIL III 2006 (http://edh-www.adw.uni-heidelberg.de/edh/inschrift/HD054337; EDCS-ID: EDCS-27601572).

24   There is a duality of roles implied by the inscription (son/brother) leading to different opinions between epigraphists and historians who mention this text. The only explanation we have found for this duality of roles refers to a semantic overlapping of terms denominating both family connections and business relationships. As well, it could denote a religious relation – and maybe this hypothesis is more plausible, as initiation degrees could be as important as blood-ties. Another possibility, given the name from the wife's family, is that Apollonius was Aquila's brother-in-law, a bigger age gap justifying the paternal feelings as well.

25   CIL III 7761 = IDR III/5, 218; http://edh-www.adw.uni-heidelberg.de/edh/inschrift/HD038327; http://www.ubi-erat-lupa. org/imagelink/index.php?Nr=17510.

5. Network visualizations

Although in the preceding section we have mentioned analyzing a network of connections, the real network analyses based on our database imply large samples and characters connected by common features not necessarily by real-life relationships. Trends, sometimes visible in case studies, can be proved or highlighted better by statistical quantification and by employing social networks analyses (SNA)[26] means – connecting people based on common traits, obtaining and finally visualizing various networks. Over the last few years these methods have become increasingly popular for historical research in general[27] and antiquity studies in particular. For the visualizations we have used the tool Gephi.[28]

Below, we will present some case studies we have worked on so far. The first one (Fig. 10) is connected to the occupational studies described above, presenting the occupational clusters from Dacia (blue), Moesia Superior (pink) and Moesia Inferior (yellow). The predominance of Moesia Inferior is due to the Greek epigraphs and the different Greek epigraphic habit, which implies inscriptions more detailed and narrative than the average Roman provincial inscriptions.

The predominating clusters are HISCO 0/1 (lower right corner) and HISCO 4 (middle left), as expected, but the visualization makes this statement very easily presentable and beyond doubts.
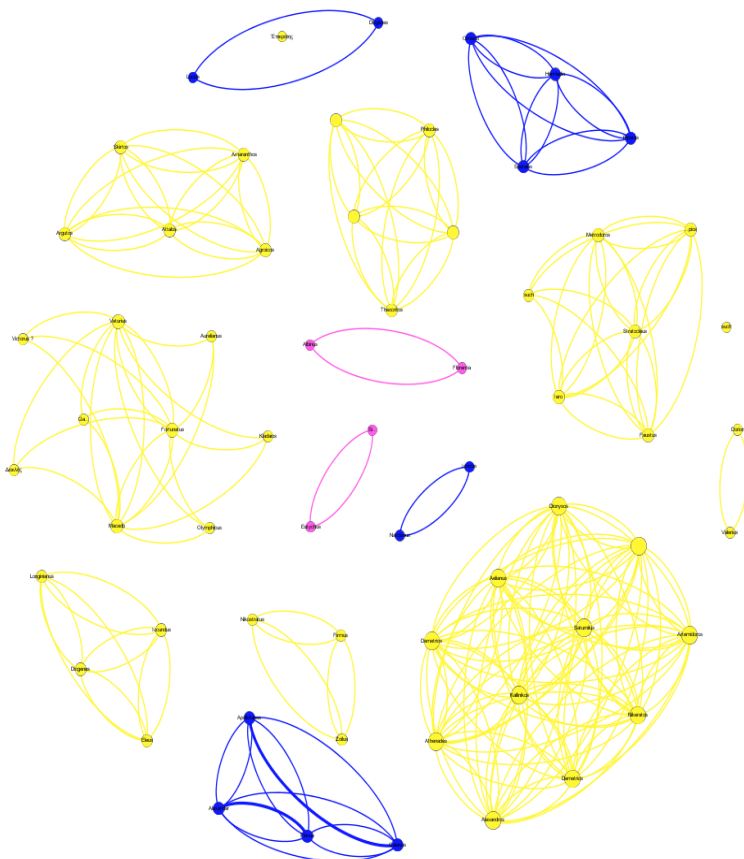
**Figure 10: The occupational networks of Dacia and Moesia**

26 Barabási (2002).

27 http://historicalnetworkresearch.org/resources/bibliography/provides a ratherextensivebibliographical list.

28 https://gephi.org/.

The other example represents the people attested on military funerary monuments from Roman Dacia (Fig. 11). The military personnel is represented by red nodes and the connected civilians by blue nodes. The edges' thickness is determined by the degrees from the relationships ontology (1–5, with 1 being the thickest). On this visualization, one can see small-sized networks, usually connected through family ties. Also interesting, some examples of so-called military families are visible – cases where two brothers, or father and son, have both opted for a military career.

The networks presented and shortly analyzed serve as example of what one can do with the help of quantification and visualization, based on the information extracted from the population database. We want to underline the scientific benefits of employing this technical tool and the informational profits of such enterprises.
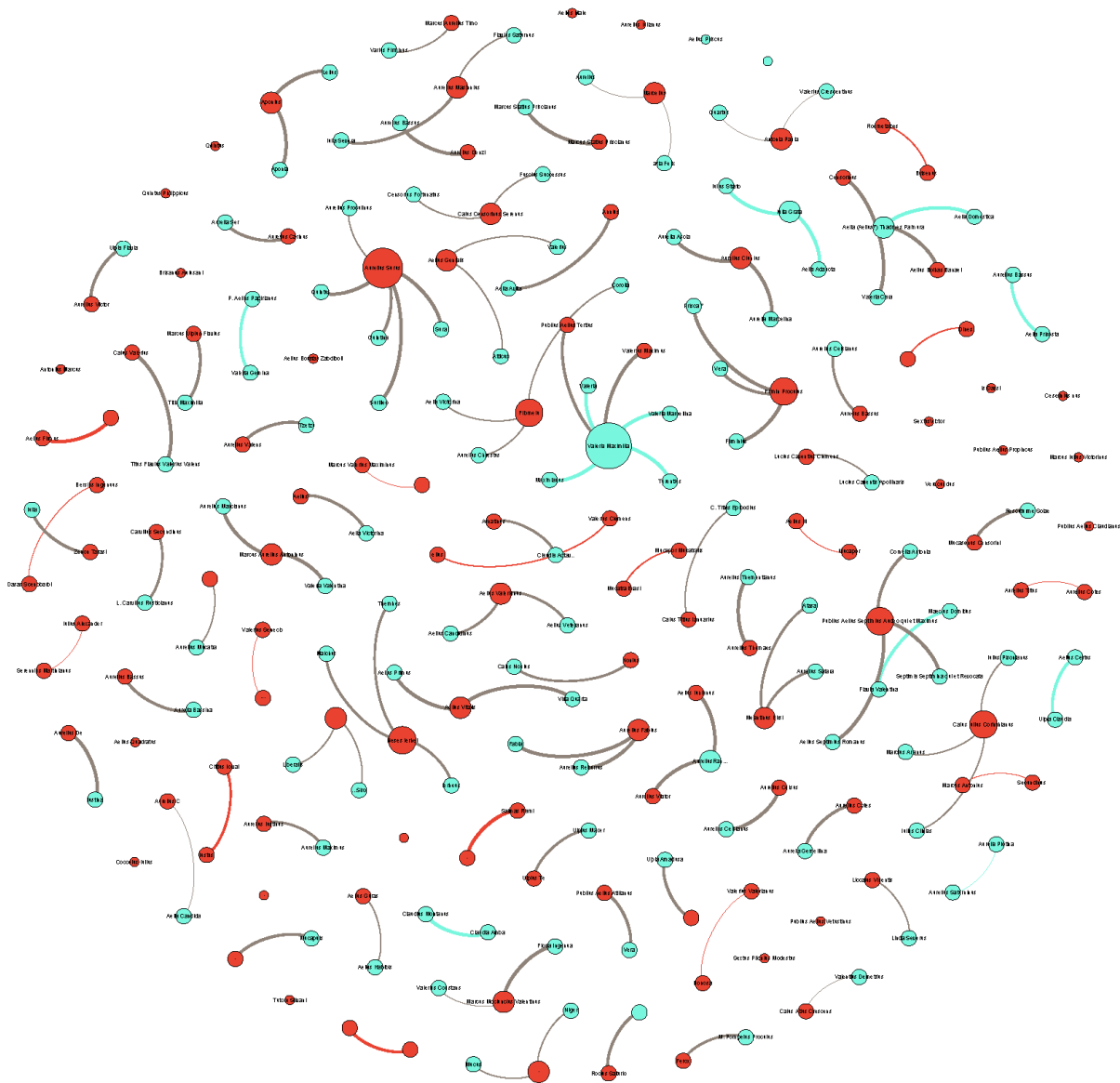
Figure 11: Links of the people attested on military funerary monuments

6. Conclusions

This article documents the *Romans 1by1* database and presents some of the first scientific results of our work. Being a population database built exclusively for epigraphic sources, it can shed light on many aspects of life from the Roman provincial world. The research on professions revealed an amazing variety of occupational titles; the encoding of these occupational titles with the HISCO-codes was a first step to better understand the identitary values of professional status and their epigraphic presentations. The prosopographical and network analyses shed light on various aspects, trends and particularities of the Roman society. Expanding and making *Romans 1by1* more complete will be a scientific gain and a step forward in knowing and understanding the classical world.

7. List of abbreviations

AE: L'Année Épigraphique, Paris.

CIL: Corpus Inscriptionum Latinarum, Berlin.

IDR: Inscriptiones Daciae Romanae, Bucureşti – Paris.

8. Bibliography

Barabási (2002): A.-L. Barabási, Linked: The Science of Networks. Cambridge (Mass.).

Lamber, Zijdeman, van Leeuwen (2013): P. S. Lamber, R. L. Zijdeman, M. H.D. van Leeuwen, "The Construction of HISCAM: a Stratification Scale Based on Social Interactions for Historical Comparative Research", Historical Methods, 46, 2, 77–89.

Van Leeuwen, Maas (2002): M. H.D. van Leeuwen, I. Maas, A. Miles, HISCO: Historical International Standard Classification of Occupations. Leuven.

Van Leeuwen, Maas (2011): M. H.D. van Leeuwen, I. Maas, *HISCLASS*. A Historical International Social Class Scheme, Leuven.

van de Putte, Miles (2015): B. van de Putte, A. Miles, "A Social Classification Scheme for Historical Occupational data", Historical Methods, 38, 2, 61–92.

Varga (2016): R. Varga, "Aurelius Aquila, negotiator ex provincia Dacia. A prosopographic reconstruction", in R. Ardevan, E. Beu-Dachin (eds.), Mensa rotunda epigraphica Napocensis, Cluj-Napoca, 27–34.

Varga (in print): R. Varga, "Romans 1by1. Documenting a population database for the Roman world", in S. Orlandi, R. Santucci, P. M. Liuzzo, F. Mambrini (eds.), Digital and Traditional Epigraphy in Context, Roma, in print.

Author[29]

Rada Varga
Babeş-Bolyai University, Cluj-Napoca
Str. Napoca, nr. 11, Cluj-Napoca

Email: radavarga@gmail.com

DIGITAL CLASSICS ONLINE

The life cycle of the *Book of the Dead* as a Digital Humanities resource

Ulrike Henny, Jonathan Blumtritt, Marcel Schaeben & Patrick Sahle

Abstract: This contribution tracks and analyzes the life cycle of the Book of the Dead as a digital project and a rather complex research resource. It gives an account of how the digital archive "Das altägyptische Totenbuch – Ein digitales Textzeugenarchiv" was constructed in the context of the digitization efforts of the Academy for Science of North Rhine-Westphalia. From the beginning, the design of the archive has factored in a life of the digital archive beyond its funding period and has sighted to create a sustainable information resource. The main issues to be discussed here are what experiences have been made with sustainability, use and reuse of the Book of the Dead archive since the official end of the project in December 2012, with a focus on conceptual, technical and organizational aspects. The lessons learned can be of interest for future undertakings in the creation of XML and web-based digital platforms in Digital Classics and beyond. In a nutshell, they are: (a) the importance of wary technological choices in an initial phase cannot be underestimated, (b) the application and presentation layers of a digital resource, if present, are an essential part of it, (c) a certain degree of commitment from the research community and funding bodies alike is indispensable for maintaining a web-based complex Digital Humanities resource.

1. The Project

The Old Egyptian Book of the Dead (BoD) is a collection of c. 200 magic spells which can be found in varying selections and combinations on different objects such as papyrus rolls, linen bandages used in mummification, palls, coffins, temple or tomb walls and other grave goods. The spells were intended as to assist the journey of a deceased person through the underworld, at the judgement of the dead and in the transition to the afterlife. In the paradigmatic case, the small texts form a *book* on papyrus, with the single spells written in hieroglyphs or hieratic script and/or illustrated by so called vignettes. Such a book was composed for and dedicated to a particular person and showed a more or less canonical structure and order of texts through the long period from the New Kingdom to the Roman times (c. 1550 BC to c. 300 AD).
As one of the central sources for the study of Egyptian religion and funerary culture, the BoD has long been an important topic in Egyptology. In the early 1990s, a team led by Ursula Rößler-Köhler at the University of Bonn, Germany, started a research project (BDP) to collect and document all known witnesses.[1] Initially funded by the German Research Council, it became a long-term project of the Nordrhein-Westfälische Akademie der Wissenschaften / North Rhine-Westphalian Academy of Sciences (AWK) lasting until 2012. Some years before the foreseeable end of funding, it became clear that the only reasonable way to publish the

---

[1]   The results of the project were principally published in the two series HAT and SAT, for editions of the BoD manuscripts and for research studies on the subject, respectively Rößler-Köhler (1995) and Rößler-Köhler (1998). The original project website is available at https://www.totenbuch-projekt.uni-bonn.de/.

collected knowledge and the outcomes of research would be an online database. Printing such an extensive metadata and image archive in its entirety would simply have been impossible. For this reason, the Academy granted additional funds for two years in order to combine the ongoing research in the subject with a Digital Humanities (DH) track.[2]

The primary goal of this collaboration was to secure the results of the research conducted over a period of nearly 20 years by the Egyptologists involved in the project and to make them publicly available in a topical and highly useful manner. But, as it is often the case, such practical goals have had more far-reaching effects, and it remains to be determined whether they lead to a more fundamental change in the setup and methodology of research. But at the outset, the assignment was clear: to transform a given archive of knowledge on the BoD into a sustainable digital resource that is comprehensive, rich in information, easy to use and stable far beyond the lifespan of the funded research project. The original archive contained a bibliographic database intended to be comprehensive as well as descriptions of and documentation for c. 3000 objects, comprising in many cases photographic or microfilm images of these objects.

2. (Re)Birth: the Book of the Dead as a digital resource

2.1 Data model, formats and conversion

Data about the BoD had already been gathered in digital form within the long-term research project at the University of Bonn. General descriptive information had been stored in a FileMaker database and bibliographic data in a Citavi instance. However with respect to the characteristics of the objects to be described, the choices of data model and software had been made rather arbitrarily, without being rooted in a formal education in database modeling. The resulting structure and usage of the databases did not follow the underlying relational model in a strict way, but gradually accrued from the daily needs of the project.

The first step towards the creation of a digital resource in a narrower sense was therefore to remodel the data. The goal was to preserve as much as possible of the existing data structure. At the same time, it was decided to use XML as the underlying data model of the Book of the Dead digital archive (BDA), and this entailed a significant change from a relational to an hierarchical data structure using markup. Most notably, the objects carrying the BoD spells were reorganised from flat, single database entries – where each part of an object was described separately – into historical full objects (*Gesamtobjekte*) which could consist of several actual partial objects (*Teilobjekte*). In this way, redundancy of information common to fragments of the same original object such as dating or place of origin could be avoided. Moreover, the original full object became the basic concept of the new data model, independent of its material fate and the question whether it is now preserved as a complete, fragmented or incomplete carrier of textual and visual material. In the process of reorganisation, the BoD object records were aligned with data sets of the Trismegistos database.[3] In some cases, this resulted in joining objects previously considered as separate entities.

---

2   At that time, the Bonn Egyptologist team comprising Rita Lucarelli, Florence Albert, Annik Wüthrich, and Felicitas Weber was led by Marcus Müller while the DH team at the Cologne Center for eHumanities (CCeH), University of Cologne, consisted of Patrick Sahle, Ulrike Henny, Jonathan Blumtritt and Franz Fischer. Recently, Marcel Schaeben has worked on an update of the database underlying the digital archive which is hosted at http://www.totenbuch.awk.nrw.de/. The BoD as a DH project has previously been presented on the following occasions: Sahle/Henny (2012a), Sahle/Henny (2012b), Sahle/Henny (2013a), Sahle/Henny (2013b), Henny (2013), and Legowski (2015).

3   Cf. Depauw/Geldorf (2014) on Trismegistos in general. Trismegistos provides unique and stable identifiers for texts from the ancient world.

In addition to the remodeling of the basic entities, fields describing properties of the objects were developed and further differentiated. Where possible, narrative descriptions that did not provide any additional value were resolved into a more explicit structure of XML elements with attributes and simple text values as content. Where appropriate, controlled lists of values were created to avoid inconsistencies as, for example, variation in spelling. The controlled lists were added to a separate XML file which functioned as a central knowledge base. The following figure gives a basic, non-comprehensive overview of the data model for a BoD object:
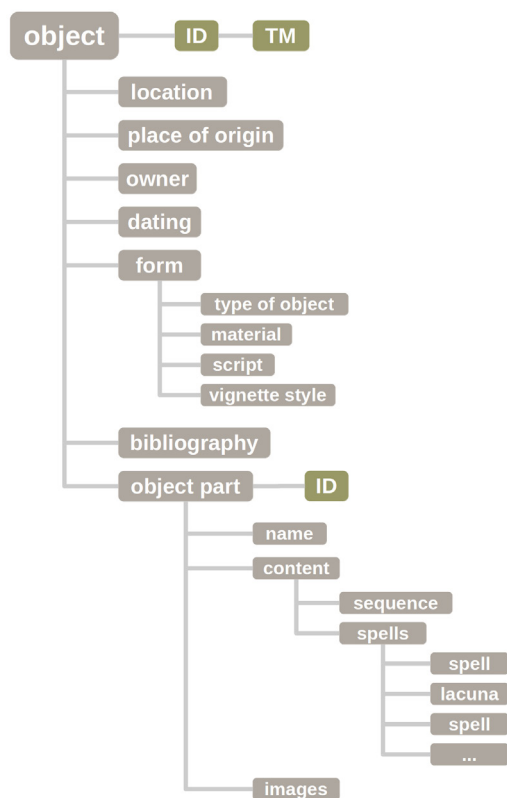
Figure 1: Basic structure of the data model for a BoD object.

Refining the data model proved to be especially fruitful for the description of the object's contents. In the original practice, the sequence of spells found on an object was written down as a single string, e.g. "1: rto: Tb $15 - 22 - 25 - 26 - 1$, vso: ? 2: rto + vso: Tb 1 4: Tb 102 8: rto: Tb $26 - 27$; vso: Tb $2 - 3 - 5 - 6$."[4] In the new model, the string was resolved to a sequence of spell and lacuna elements, each spell element containing just one spell or vignette name and the lacuna elements left empty. Because of the complexity of the original spell sequence string, it has been kept alongside the dissolved content description to make the remodeling transparent at this critical point. Making the structure of the books explicit, i.e., to mark up the components of the sequence of spells and vignettes, was the precondition for a computational and analytic approach to the material.

Similarly, in case of the spells a canonical list was added to the central knowledge base and mapped to the instances of spells witnessed on the objects. Likewise, other levels of description such as places of origin, current locations and institutions, epochs and periods for the dating of objects, etc. were controlled. Besides the control and regulation of values, the knowledge base had the function of being the anchor to which external information (encyclopedic knowledge) could be attached (e.g. geographical coordinates to place names or date specifications to names of historical epochs). Notably, translations of the spells stemming from the academy project

---

4   Cf. Müller et al. (2012b).

Altägyptisches Wörterbuch and available via the platform Thesaurus Linguae Aegyptiae[5] have been added to the knowledge base. In this way, the knowledge base can thus be exploited for networking (interconnection of data in the project itself, linking of project data to external data, integration of external information into the project, support of external tools and services) and as a basis for layers of presentation and analysis.

The third main component of the BDA data model, besides the objects and the knowledge base, is the images of the text bearing objects. Egyptologists in the project had acquired existing images both from museums, and in part, the images were obtained by photographing and scanning reproductions found in secondary literature. In total images were obtained from 2,245 of the 2,992 objects. It was a goal to keep the administration of images simple, so the image metadata was attached directly to the object descriptions as can be seen in Figure 1. Each image was assigned to one of the user groups "guest", "community", "project" or "admin", in order to allow for a detailed access control, as some parts of the graphical material are accessible only for registered users who have accepted the terms and conditions. Of the 21,684 images, 94 are open to the public, 20,701 are only accessible to registered users and 889 only to project members.

It was decided to use a custom, local XML schema[6] but to offer the data in two standardised export formats: OAI Dublin Core[7] and EpiDoc[8]. The local XML dialect permitted the expression of characteristics and relations specific to the objects of study without the need to compromise and possibly lose information coming from the original databases. It also meant that the potential of the new data model to capture information internal and external to the project as desired was not limited, especially with respect to spell sequences and object-part relations.[9]

The process of converting the data so as to conform to the new model was automated as much as possible. XML exports of the FileMaker and Citavi databases were generated and processed further with XSLT. The images were batch-processed with IrfanView to produce derivative formats suitable for a web presentation. Much of the refinement of descriptive categories and values was done in a semi-automated, iterative process, where data reports stating which part of the data could not be converted automatically were created by the DH team and dealt with by the Egyptologists.[10]

---

5   Cf. BBAW (2014), http://aaew.bbaw.de/tla/index.html. Also inside the project itself, the knowledge base has been used to add information which goes beyond the single object descriptions, for example an additional index of motifs which classifies, lists and depicts types of motifs that occur on vignettes, as well as links to the objects that witness them. Cf. Müller et al. (2012c).

6   Cf. Müller et al. (2012d).

7   A metadata format suggested to be used when implementing the Open Archives Initiative Protocol, cf. Lagoze et al. (2005).

8   A subset of the Text Encoding Initiative's standard TEI, used for the encoding of ancient documents, cf. Elliott et al. (2011–2014).

9   The direct use of EpiDoc was considered but not pursued because the archive consists of complex object descriptions and not of texts and editions thereof. On the one hand, EpiDoc would not have been used to full capacity regarding the possibilities for encoding texts; on the other hand, it would have been necessary to extend the TEI header to suit the extensive metadata.

10   A special task consisted in the conversion of characters. In the original project, a traditional non-Unicode font had been used for transliterations, which in the object descriptions appeared for example in names of the book's owners and relatives. In the course of data conversion, the transliterated characters were converted to Unicode characters.

2.2 Technical framework

Ultimately, the basic choices made for the technical framework of the digital archive where a consequence of the decision to use XML as a data model. Because the goal was to build a rich and interactive digital platform for the web, with HTML as the markup language, it was advisable to use technologies which would facilitate as much as possible the interplay between the underlying data and the presentational format. The X-technologies were an obvious choice because the path from XML to XHTML is straightforward with the help of standards like XPath, XSLT and XQuery[11] which have been devised precisely for the retrieval and transformation of XML data without the need to convert the data back and forth between different kinds of formats. In addition, these standards are easy, text-based, human- as well as machine-readable and, as such, offer a promising solution to the problem of long-term accessibility and usability of the data.

As they are, the basic XML standards can be used directly to construct web pages. They might not suffice, however, for the building of an application that is dynamic inasmuch as the content is not simply delivered "as is" but as a response to how users engage with the platform. Second, it was a goal to create a system that could be updated by members of the community even after the end of the project, because it was clear that the launch of the digital archive, planned for February 2012, would predate the overall end of the long-term research project by only nine months. In other words, the system had to support changes and updates of the data which could be made not only by non-technicians but also by non-project members. It became necessary to turn to a database management system with support for user administration, simultaneous access by multiple users, updates and versioning of the data. With their ability to ensure and facilitate the model-related and technical integrity of the platform, native XML databases were considered an intelligent option. At the time of decision, two of the Open Source candidates were BaseX and eXistdb.[12] It was decided to use the latter because there were no striking arguments in favor of one or the other and because colleagues from other academies and the University of Cologne (particularly the Monasterium project[13]) had some experience with it. According to Siegel and Retter, eXist is, amongst other things, a "NoSQL document database for XML and binary (including text)", a "web server for consuming and serving documents", a "web application platform", and a "document creation and capture platform (XForms)".[14] It is thus possible to store XML files as well as other types of data in the database: eXist is not just a database but includes its own web server and various programming interfaces.[15] XQuery scripts residing in the database can be executed directly via requests from the browser. This means that it is possible to construct *one* package including the data and the application that can be stored in the database, run with the eXist server and can (at least in theory) be exported and migrated as a whole if necessary. Figure 2 provides an outline of how eXist was used in the technical framework of the BDA:

---

11   XPath, XSLT and XQuery stand for "XML Path Language", "Extensible Stylesheet Language Transformation" and "XML Query Language". See W3C (2007), https://www.w3.org/standards/xml/ for further information.

12   See Siegel/Retter (2014) for a comprehensive introduction and account of eXist and Grün (2011) on baseX.

13   Monasterium.net is a collaborative and virtual archive which, at least to date, is built on eXist, see ICARUS (2017), http://monasterium.net/mom. Krah (2009) and Heinz (2010) introduce the archive, but do not broach the issue of technological choices.

14   Siegel/Retter (2014), p. 3.
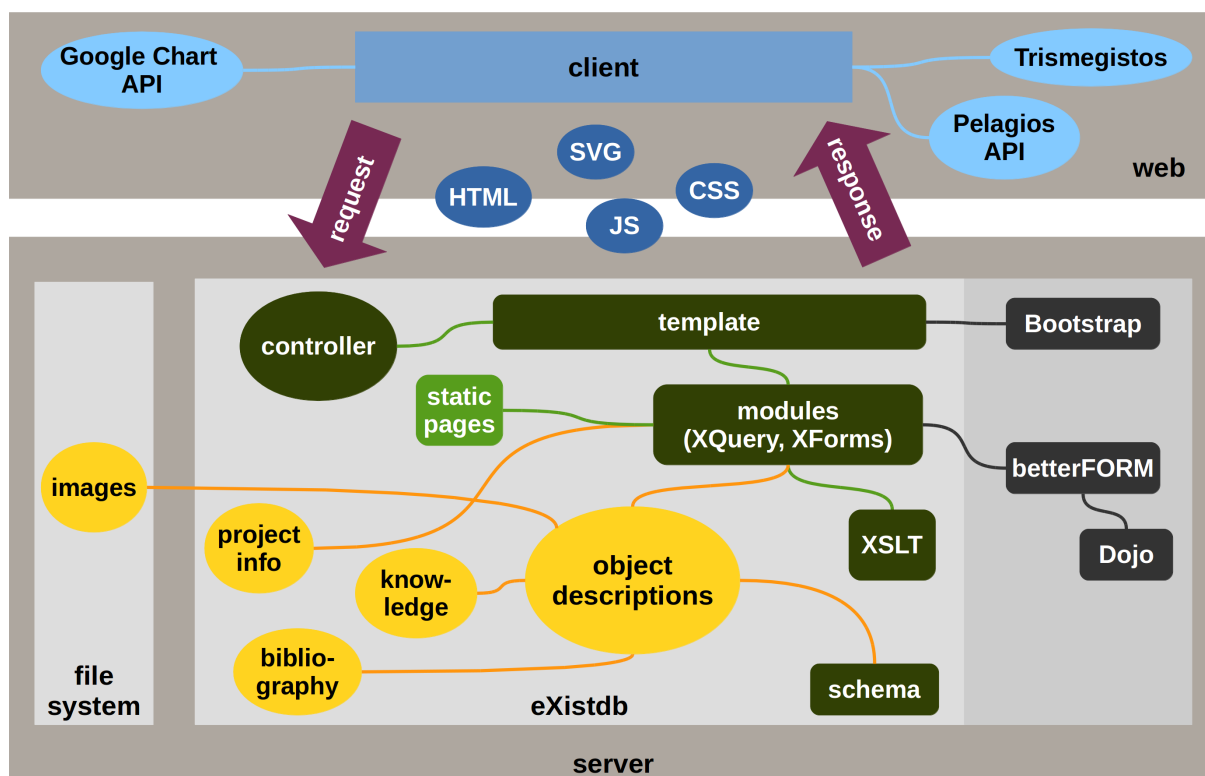
15   E.g. HTTP REST and WebDAV.

Figure 2: Outline of the technical framework behind the BDA.

As can be seen in the figure, the digital environment of the BDA resides completely in the X-world: underlying XML data, an eXist database as the backbone and XQuery, XSLT and XForms as processing methods for creating the user interface. An ingoing request is handled by the eXist controller that rewrites URLs and passes queries on to the central HTML template managing the website layout. From there, calls go to XQuery and XForms modules and functions to lookup data and produce results that are passed on to the client in response. The XML data is organised into different collections that are interrelated: objects, knowledge, bibliography and project information. Information that is more static and rarely updated is organised in pre-generated static pages. JavaScripts and CSS stylesheets used for design and functionalities of the web application are included via the template and through the eXist-internal XForms implementation BetterFORM. All in all, the only kind of data in the BDA that does not reside inside eXist are the image files; these are organised in a separate file system and linked to the object descriptions. Beyond the usage of web application building blocks which are directly included in eXist, the Google Chart API[16] and the Pelagios API[17] were used to support visualisations, linkage to project-external data and the use of third-party widgets in the web presentation. The W3C standard SVG[18] was used to produce visualisations that went beyond the capabilities of the Google Chart API.

In the BDA, forms play a special role in supporting user activities. The platform was designed to include a search form called *browsing* that is always visible on every page and looks up object descriptions immediately, whenever the user changes the search criteria. In addition, a special search form is provided for owners of the books and their relatives. Moreover, forms are needed to support both the addition of new data to the archive, and changes to existing data, in particular object descriptions, bibliography entries, user registration and administration data. All of the forms are complex. In case of the search forms, they are complex because they allow

---

16   Cf. Google (2017), https://developers.google.com/chart/.

17   See Pelagios (2017a), https://github.com/pelagios/peripleo. "Peripleo" is a new name for the former Pelagios API.

18   Scalable Vector Graphics, an XML vocabulary for vector graphics.

for the combination of several kinds of search criteria, e.g., a specific place of origin with certain spells and a free search term. The search criteria might be hierarchically organised and in some cases (spells and vignettes, for example) the number of search fields is not predetermined. Regarding the object description editing forms, the complexity of nearly the entire underlying data model had to be taken into account. Where values were controlled in the knowledge base – such as the canonical names of spells – mappings were necessary to connect the *clean* list shown to the user with the *untidy* object description entries. Finally, the interrelations between the different types of data (object descriptions, knowledge, bibliography, user data, images) had to be taken into account in the forms.

eXist comes with two XForms implementations: XSLTForms and betterFORM.[19] For the BDA, betterFORM was chosen to be used for the development of the aforementioned search and input forms because, at the time, its implementation of the XForms standard seemed more complete than that of XSLTForms and it was more widely used. The choice of betterFORM had some consequences: internally, betterFORM uses the JavaScript library Dojo Toolkit.[20] At the time of development, it was not foreseeable that the dependencies between specific versions of eXist, betterFORM and Dojo could be an issue in the future.[21] Second, the BDA data model pushed the capabilities of betterFORM (and XForms) to its limits with, for example, the high number of nested form controls and the necessary handling of special Unicode signs in the inputs. For these reasons, at certain points it became necessary to extend betterFORM via JavaScript.[22] Scripting was also used to extend a standard image viewer provided by Dojo (dojox.image.Lightbox[23]) according to the needs of the BDA.

2.3 Features and usability

While planning the structure, functionalities and design of the BDA web presentation, the primary challenge was to supply experts in the field with quick and straightforward access to the information that he or she assumed to be part of the archive. At the same time, it was challenging to open up the complex and multifaceted material to interested lay users. As a first step in responding to these requirements, we had to clarify which structures, items and contents actually had to constitute and represent the resource on the surface level. In other words, we needed to answer the question (yet again!): what is the resource? First and foremost, the archive was deemed to be about text witnesses that needed to be made accessible through their properties and content. In the BDA, the text witnesses are manifest in the form of object

---

19   XForms is a W3C standard for web forms which are embedded into other markup languages. An important feature of XForms is that it distinguishes between a data model, instances of data, form controls and "bindings" between the different components, and thus follows a model-view-controller design, cf. W3C (2009), https://www.w3.org/TR/xforms/ and Dubinko (2003). Regarding eXist's XForms implementations, XSLTForms is a client-side XForms processor while betterFORM is a server-side implementation which processes the XForms on the server and sends the resulting HTML and JavaScript of the forms to the browser. User interactions with the forms are then managed with the help of Ajax calls to the server. Cf. Siegel/Retter (2014), 256–271.

20   Cf. Dojo (2017a), https://dojotoolkit.org/.

21   The first version of the BDA ran on eXist 1.4.1, with the betterFORM limeGreen PreRelease and Dojo 1.6.1.

22   This was even prepared and supported by betterFORM: "Isn't it one of the most mentioned arguments that XForm [sic] makes writing JavaScript superfluous? The authors of the XForms spec. Have been more cautious by stating '(...) *reduces the need for scripting*'. This clearly says that scripting might and most likely will be needed to build full applications. XForms is not the answer to all application needs." Turner (2010).

23   Cf. Dojo (2017b), https://dojotoolkit.org/reference-guide/1.

descriptions holding manifold information, in corresponding images, in extracted information such as indexes, visualisations and motifs and in the form of contexts, e.g., bibliographies. The text witnesses are the anchor to which all access points in the presentation are bound.

A faceted browsing allows users to find objects on the basis of multiple search criteria that can be employed individually or in conjunction with one another (see Figure 3 for example).

Experts who want to examine certain spells and know the spells' names or identification numbers can directly search for the texts and vignettes. In the example, the search is conducted for objects which bear the spell no. 5 and 11/49 in text or vignette form, combined with the criterion "date of origin" set to the epoch "New Kingdom": this yields just a single hit. Thus, experts can filter the objects according to very specific criteria while the lay user can search for objects on the basis of more general properties, e.g., type of object or place of origin.

Figure 3: Faceted browsing (example).

In addition to the faceted browsing, other kinds of entry points to the archive have been created. An overview of the 238 different spells registered in the corpus is given in the website menu and conveys an impression of how the corpus is actually composed. Several indexes allow for direct retrieval of specific information: a bibliography, an alphabetical as well as semantically grouped index of motifs, an index of owners (the deceased persons for whom the

books were created) and their relatives with its own elaborate search filter,[24] an index of institutions by country and name, and, finally, an index of all the objects that are contained in the archive by name. Figure 4 gives an abridged example of an object description as it is presented in the digital archive:[25]

TM 135212, Holzbrett, Abusir	← Zurück zur Trefferliste
Objektgruppe:	Holzbrett
Material:	Holz
Material Detail:	mit Stuck
Standort:	Abusir (Ägypten) , Grabungsmagazin der Tschechischen Mission
Herkunft:	Abusir
Datierung:	Spätzeit, 26.-30. Dynastie
Besitzer:	Nkȝw , Ḥˁ- Nkȝw , it- nṯr sm ḥw- wȝḏ ḥˁ- (rȝ) Ranke: I, 213.16, Geschlecht: männlich
Verwandte:	Irti- rw (Mutter)
Schrift:	Hieratisch
Vignettenstil:	keine Angaben
Gesamtmaße:	1: H. 9 x 41 cm; 2: H. 11 x 34 cm; 4: H. 7 x 19 cm; 8: H. 9 x 43 cm
Name:	Holzbretter 1-9
Bibliografie:	Janák, J.; Landgráfová, R., Nekau's Book of the Dead Reopened, in: Backes, B.; Müller-Roth, M.; Stöhr, S. (Hg.), Ausgestattet mit den Schriften des Thot. Festschrift für Irmtraut Munro zu ihrem 65. Geburtstag, Wiesbaden 2009, S. 83-87.

Holzbretter 1-9

Inhalt:	1: rto: Tb 15 - 22 - 25 - 26 - 1, vso: ? 2: rto + vso: Tb 1 4: Tb 102 8: rto: Tb 26 - 27; vso: Tb 2 - 3 - 5 - 6

Abbildungen:

Zu diesem Objekt ist weiteres Bildmaterial vorhanden, das rechtlichen Beschränkungen unterliegt. Bitte registrieren Sie sich bzw. melden Sie sich mit Ihrem Benutzernamen an, um Zugang zu diesen Bildern zu erhalten.

Zitationshinweis
Bitte zitieren Sie die Objektbeschreibung als *Totenbuchprojekt Bonn, TM 135212,* *<totenbuch.awk.nrw.de/objekt/tm135212>.* Diese Adresse wird dauerhaft unterstützt.

**Figure 4: Description of object TM 135212, wooden board, Abusir (detail).**

Another feature of the archive's web presentation are the so-called *overviews*, which are simple visualisations and statistics of the data, including maps. On the one hand, they provide orientation for those consulting the material; on the other hand, they represent yet another access point for the object descriptions. Overviews have been created for the objects and their tradition, the images, locations and places of origin, dating, types of objects and material, owners and relatives, scripts, vignette styles, spells and groups of spells. Moreover, they initiate analytical approaches to the data, especially when the sequences of spells and vignettes found on the books are analyzed and when the visualisations go beyond the possibilities that standardised types of charts offer. Figure 5 depicts an analytical visualisation of the neighbourship of spells in the books compared to the canonical order of spells established by the field.[26]

---

24 Cf. Müller, inter alia (2012e). The following information is given for owners and relatives: name, short name, title, status as owner or relative, degree of kinship for relatives, a bibliographic reference to rank, gender, objects that witness the name, epoch and type of object.

25 See Müller, inter alia (2012b).

26 See Müller, inter alia (2012f) for details.
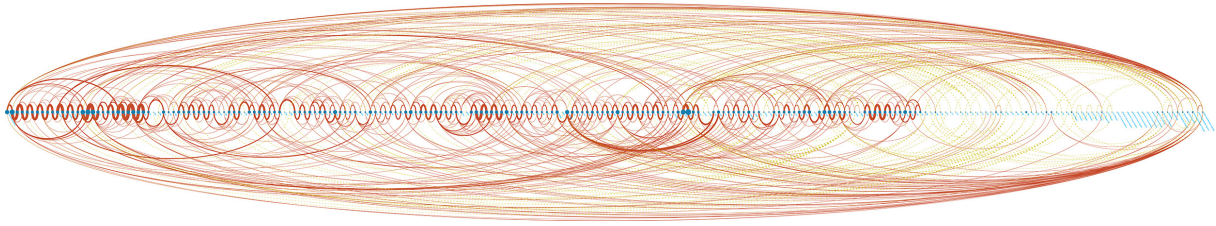
Figure 5: Visualisation of the neighbourship of spells on the books (preview).

As the *active* phase of the project's Digital Humanities track was relatively short[27] and the end of the overall long-term project drew near, it was clear that the question of sustainability of the resource and especially the features of the web application could not be postponed. Two central components of the digital archive in this regard are persistent URLs and APIs[28]. An example of such a stable URI for a BoD object description is http://totenbuch.awk.nrw.de/objekt/tm135212. Even though the application runs on a server of the University of Cologne at the moment, it was decided to use the academy's basic address awk.nrw.de to ensure long-term accessibility. The usage of Trismegistos numbers in the URL assures that the possible variability of object names does not put at risk that the addresses remain stable and citable. The URL as a whole is mapped onto the technical system behind it but is not directly entangled with it. Several kinds of APIs have been implemented for the digital archive: OAI-PMH,[29] REST,[30] Pelagios (place name) annotations in RDF and an unAPI for the retrieval of bibliographic references in various formats.[31]

3. Afterlife of a project: continuous curation of a resource

3.1 Launch, usage, and community

The digital resource was introduced to the public at the closing convention of the BDP in February 2012. It has acquired a stable user base: as of May 2016, over 600 users have registered. The statistics consistently show over 30 unique daily visitors with occasional peaks around meetings and conferences. While these appear to be very modest numbers compared with what we are accustomed to in times of mass social media, it is in fact a very respectable achievement in the context of humanities projects dedicated to a highly specialised topic and geared towards the very specific usage patterns exhibited by its users.[32]

---

27   The digital reworking of the resource started in January 2011 and ended in December 2012.

28   API stands for Application Programming Interface. Here it means functionalities of software that are opened up to external programs through a documented interface with callable functions that deliver results in standard formats.

29   The OAI interface of the BDA builds on work from the Monasterium project, namely André Streicher, and has been adapted to the needs of the BDA by Ulrike Henny.

30   REST stands for Representational State Transfer and is a programming paradigm for distributed systems, especially Web services and APIs.

31   See Müller, inter alia (2012g) for a detailed documentation of the available APIs.

32   The analysis relies on the open source software Piwik for user statistics. Piwik analytics for the BDA has been in place since February 2013. All statements on long-term developments and trends refer to the timeframe from February 2013 to the present day. All information that can be exploited to identify a user, namely the IP-address, is processed and stored solely in an anonymised form. Cf. Piwik (2017), https://piwik.org/.

The most visited pages are, of course, the core contents of the application, i.e., the object descriptions followed by the overview pages for single spells. In absolute numbers "Papyrus Turin 1791"[33] and "Papyrus BM EA 10477"[34] attracted the most public attention over the entire timeframe. This is not surprising since they are without any doubt among the most prominent manifestations of the BoD. Of the numerous registers and other aids on the website the bibliography and the motif registers seem to be of particular interest.

While users from German-speaking countries constitute the largest single user group, requests from outside surpass these, despite the fact that the website is available only in German. Apparently, the language barrier is no serious obstacle for a trained Egyptologist, as the information is presented in a simple and structured way and crucial technical terms are either familiar to the international community even in their German form, or they translate well to English. Language does present a challenge when it comes to certain features implemented in the search function, user registration or in other places where instructions or documentation is needed. Of course, we would have liked to have an English version of the website, but resource restrictions led us to concentrate on content and features rather than on the translation of the user interface and introductory or help texts.

Users who stay for at least ten minutes spend an average of 37 minutes on the BDA. Those who fall into that category typically focus on a selected number of specific object descriptions for a prolonged time and examine the attached pictures. These power users make up roughly a quarter of the website's users.

The BDA's impact can also be traced through citations across both born-digital and traditional print publications. Object descriptions are cited in various essays and monographs around the world.[35] The digital archive supports citation and referencing through clear, persistent, and human readable web addresses not only for object descriptions (as described above) but also for other parts of the application.

Beyond monitoring statistics and impact there is real interaction with users through an e-mail address published on the website. Recipients of that address include technical staff as well as subject researchers, i.e., members of the former research project, who were primarily responsible for the edition of the scholarly content. This arrangement allows for flexible reaction to inquiries with both typical support questions as well as questions regarding the content. Traditional user support mainly concerns issues that arise during registration.

We receive community feedback on the contents of the BDA from people representing different disciplines including Egyptology, Archaeology, theology, Restoration Sciences, LAM (library, archive, museum), along with hobby researchers and spiritually inclined lay users. This diversity most likely reflects the composition of the *power user* group. Input from the community can range from pointing out minor technical bugs or content related errors, to on-going debates in the identification of spells on objects, additions to the bibliography, and updates or corrections submitted by the archive hosting a particular object.

From a methodological point of view, it is clear that these kinds of changes and additions are almost exclusively editorial and should be authorised by the scholarly editors of the BoD. Although the editors are in principle available through the support mailing list and remain responsive to inquiries, the project shares the same fate as many before it: research groups dissolve after the end of the funding period and their members are gradually absorbed by their new line of work in academia or elsewhere, which leaves them less and less time for content curation.

---

33   Cf. Müller, inter alia (2012h).

34   Cf. Müller, inter alia (2012i).

35   The correct measurement of citation impact itself can be subject to scientific research. We let ourselves content with a general positive impression. See, for example, Google Scholar to find references to the BDA web application: https://scholar.google.de/scholar?q=totenbuch.awk.nrw.de, accessed on June 28, 2016.

3.2 Maintenance, update and relaunch

In the case presented here, the funding body, the AWK, took on responsibility for keeping the results of the research project available to the public through its digital presentation. While the academy owns the data and the editorial responsibility remains with the scholarly editors, maintenance and technical support was delegated to the CCeH at the University of Cologne, which was also responsible for the programming of the application during the project phase. Financially and on an institutional level, this arrangement was backed by a cooperation agreement between the academy and the University of Cologne.

In this instance, continual maintenance challenges can be divided into at least four fields of activity. First of all, there are indispensable tasks on the level of technical support. Neglecting them would compromise the functionality and ultimately render the application unavailable.[36] The second area is comprised of the fixing of minor bugs and the addition of new features. Although the absence of these features and persistence of bugs may be an annoyance, strictly speaking they do not break the application. We distinguish a third category of code curation that has an eye on prospective developments. This may include updates and subsequent restructuring of code. We prefer to think of such development as an investment. Again, while omitting such development may not necessarily break the application, it may generate even higher and unpredictable expenses in the future, increase risk for hacking, which may, in turn, lead to the abandonment of the BDA. Finally, as a fourth area, there are, of course, the tasks generated by the editors and their requests for content updates or questions and additional information supplied by the users.

With respect to the continuous service of the website, since the launch in February 2012, beyond the basic maintenance tasks, we have encountered a number of incidents that threatened the functionality of the resource. In January 2014, an unannounced update of the Java version on the servers prevented the database instance from starting up. In August 2014, a widget developed by an external partner[37] partly stopped working because it was relying on an API that was discontinued. Hacking attempts exploiting common vulnerabilities to spam or disable sites happen every day. On two occasions, these attempts were partly successful so that the application had to be restored and additional measures had to be taken to strengthen it against particular attacks. These kinds of incidents cannot be anticipated and the extent of the work required to fix them is hard to calculate, but they require a prompt intervention. Altogether, these efforts did not take up more than two person-days per year. This may sound like a relatively insignificant amount of time, but it should be remembered that responding to these incidents requires in-depth knowledge of the software; the need to keep this expert knowledge alive and available – even when needed for an isolated incident – is one of the major challenges confronting digital humanities centers.

In the second field of activity, bug-fixing, most of the work was done in the first year. Thereafter the amount of time spent on bug-fixing was reduced to a very low level. Fixing of errors is now being fitted into the normal schedule according to its priority. The team keeps a wish list of features posted by the users. Implementation of these features, though, has a relatively low priority while there is no dedicated funding.

---

36   On the first level the basic infrastructure consisting of networking, DNS, server virtualisation, storage management etc. has to be in place. These services can be provided by external hosting agencies or academic computing centers. They are a prerequisite and of course generate overhead, but they are well-established services and can be priced in reliably and thus are not tackled here.

37   The Pelagios Place Widget, developed by the Pelagios community, cf. Pelagios (2017b), https://github.com/pelagios/pelagios-widgets. At the time of writing, the latest commit was registered on August 14, 2014.

Regarding the third category, investments that deal with sustainable and long-term perspective of the project, plans for adjusting the code to recent developments were made as early as spring 2013. The team at the CCeH discussed an update to eXist 1.4.3 to solve security issues[38] and to improve the performance of the application. After consulting with Wolfgang Meier, the chief developer of eXist, this initiative was suspended because it was clear that technical intricacies would require a considerable amount of work and it seemed advisable to put that effort into the next major version. An anew hack in 2016 revived the plan to update the database and relaunch the application.

In principle, it is possible to export an entire application from one eXist version and transfer it to another eXist instance.[39] In case of the BDA, the devil was in the details. After transferring the core of the application[40] to the new eXist version, dependencies between project code and eXist-internal third-party components (mainly BetterFORM and Dojo) became apparent and had to be fixed.

At the time of initial development, BetterFORM was a solid choice as a user interface framework that tightly integrates with XML technologies and eXist in particular. A version of the BetterFORM XForms processor is bundled with the eXist database, so initially it was an advantage not to have to include additional external dependencies such as an HTML/Javascript framework. The main difficulty with the migration of the forms was that the newer version of BetterFORM in eXist 2.2 came with a newer version of Dojo. Unfortunately, there had been fundamental changes in the Dojo API as well as minor but decisive and undocumented changes in the XForms syntax expected by BetterFORM and in the eXist specific XQuery implementation.

All in all, the update of some parts of the XQuery code, the adaptation of CSS styles and the rewriting of most of the Dojo based JavaScript code were necessary. It became obvious that the scripting that had be done to extend BetterFORM where it did not suffice for handling the complex structure of the object descriptions was problematic in the moment of the migration. The same can be said for the extension of the Dojo-based image viewer and the general usage of the Dojo API to extend the features and usability of the presentation.

In the light of the problems faced, it should not be forgotten that much of the migration was completely unproblematic. The bundle of XML data, XQuery and XSLT scripts, static HTML pages and SVG graphics and additional binary resources could be transferred directly to the new database. It was not necessary to alter the data, their structure and interrelations. The migration of the user management system was straightforward, as well.

Following the classification laid out above, the personnel resources needed to keep the BDA alive thus amount to the following:

1. Continuous server and system maintenance: ca. 16h per year[41]
2. Minor bug-fixing and additional features: 60h in the last four years
3. Update and relaunch four years after the initial launch (including planning): c. 100h once
4. Communication with users (login problems etc.) and curation of the data (corrections, additions): c. 10h per year

---

38   The website had been hacked in November 2013 and an outdated version of the database represented a weak point in this regard.

39   From eXist-db 2.0 onwards, this is supported by a dedicated package repository (cf. eXistdb 2017, http://exist-db.org/exist/apps/doc/repo.xml), but export and import routines existed even before.

40   Cf. section 2.2 and especially Figure 2 for an account of the technical framework.

41   Not including the hosting infrastructure which is managed externally.

4. Conclusion and outlook

In conclusion, the BDA shows that it is possible to transform a *traditional* long-term research project into a *digital* project of a high quality level in a relatively short amount of time and with limited financial resources – if the interdisciplinary cooperation between researchers and digital humanists works well. In this case, two years of 100% FTE of a graduate were sufficient for the initial development of the digital archive.

Regarding the stability and sustainability of the data model, technical framework and user interface, the X-technologies and eXist as a database management system have proven to be good choices until now. However, there is also a rising curve of complexity from data through application[42] to user interface when it comes to maintenance, update and relaunch. In the case of the BDA, the local XML model was and is reasonable because it made the modeling easy and the use of a local dialect can be justified when exports of the data are offered in standard formats. As regards the application and user interface, a digital resource for the BoD, and probably for other humanities research subjects as well, can hardly be created with standard software because of the particularity and complexity of the material. For the BDA, this became especially visible when the detailed and hierarchically organised object descriptions had to be mapped to and represented through user interface functionalities such as faceted browsing as well as search and input forms. A modular and flexible system that can be adapted to the requirements of the project is needed, which, at the same time, does not trouble future developments. In principle, eXist is such a flexible system but seems to be more future-proof in its core functionalities (the handling of XML data, XQueries and basic apps) than in specialised features that are in part integrated into the database as third-party libraries and interdependent with the core database in their versions.

As regards the user interface, its structure, browse and search features and its design have all stood the test of time. Most users seem to get by with it and find what they are looking for. One of the greatest rewards has been to witness continual and consistent usage and community feedback. *Power users* do not solely consume but also contribute to the information in the database. Although modest in numbers these usage patterns are of great quality as they reflect a qualified scientific use and reveal that the BDA is, as intended, an established research tool for a specialised community, supporting the production of theses, essays and other scholarly publications. In the light of the interest that the international Egyptologist community shows, it is thus reasonable and important not just to archive the underlying data but to keep the database, presentation and user interface alive and running. The web-based platform ensures that the digital resource is used because it furthers the visibility of the contents and keeps the threshold low to engage with them. A technically advanced Egyptologist might well make use of the APIs and work directly with the XML data to carry out specific formal analyses. Most users, however, will be happy to have a user interface that is as comfortable to use as possible. The key to ensuring long-term availability of digital resources is to take on responsibility and to organise continuous maintenance and curation. There is a risk inherent to this type of commitment inasmuch as there is no safe way to calculate the required workload and expenses in advance.[43] Researchers and funders alike push towards complex digital representations of research data and call for sustainability at the same time, while there are no well-practiced and institutionalised solutions in place that would guarantee long-term availability of digital systems. Fortunately, this topic has come into focus recently and is vigorously debated

---

42   Understood as the *middleware* of code and function modules providing the basic functionality of the digital archive.

43   There are few published experience reports and even fewer numbers to build on. Future developments of technology and prices are hard to predict, especially in the long term.

especially in the context of humanities.[44] Our experience is that the maintenance of a digital resource such as the one presented here is reasonably manageable. It must be said, however, that the BDA is backed by a Digital Humanities center with a lifespan longer than a single project and the corresponding personnel, and that the curation of the contents relies on the voluntary commitment of the subject researchers. The continuous use and development of the platform and ongoing interest in the subject may as well lead to new initiatives and follow-up projects which take the digital resource into a new phase and present the data in a new context. As the current user interface would be replaced by another one, maintenance could be reduced to archiving the underlying research data. At present, however, such initiatives are not in the offing. Likewise, should it become clear that the costs for maintaining the current web application surpass the benefit, the web application would be abandoned. As long as this is not the case, keeping the online BDA alive as long as possible is worthwhile.

---

44 In 2016, the second edition of the conference "Forschungsdaten in den Geisteswissenschaften" (Research Data in the Humanities) in Hamburg was announced, dedicated solely to the question of sustainability of digital research applications (cf. Universität Hamburg 2016, https://www.gwiss.uni-hamburg.de/gwin/ueber-uns/forge2016.html). Also, the 2017 edition of the Digital Humanities Conference in German-speaking countries will address digital sustainability and stresses the unresolved issue of long-term availability of digital platforms, editions and databases (cf. DHd 2017, http://www.dhd2017.ch/calls).

5. List of Abbreviations

API	Application Programming Interface
AWK	Nordrhein-Westfälische Akademie der Wissenschaften
BDA	Book of the Dead digital archive
BDP	Book of the Dead Research Project
BoD	Old Egyptian Book of the Dead
CCeH	Cologne Center for eHumanities
CSS	Cascading Style Sheets
DH	Digital Humanities
DNS	Domain Name System
FTE	Full-time equivalent
HAT	Handschriften des Altägyptischen Totenbuchs
HTML	Hypertext Markup Language
HTTP	Hyper Text Transfer Protocol
LAM	Library, Archive, Museum
OAI	Open Archives Initiative
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
RDF	Resource Description Framework
REST	Representational State Transfer
SAT	Studien zum Altägyptischen Totenbuch
SVG	Scalable Vector Graphics
TEI	Text Encoding Initiative
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
WebDAV	Web Distributed Authoring and Versioning
XHTML	Extensible Hypertext Markup Language
XML	Extensible Markup Language
XPath	XML Path Language
XQuery	XML Query Language
XSLT	Extensible Stylesheet Language Transformation

6. Bibliography

BBAW (2014): Berlin-Brandenburgische Akademie der Wissenschaften, Altägyptisches Wörterbuch, "Thesaurus Linguae Aegyptiae (TLA)", http://aaew.bbaw.de/tla/.

Depauw/Geldorf (2014): Depauw and Gheldof, "Trismegistos. An Interdisciplinary Platform for Ancient World Texts and Related Information", in: Bolikowski, Casarosa, Goodale, Houssos, Manghi, and Schirrwagen (edd.), Theory and Practice of Digital Libraries – TPDL 2013 Selected Workshops, 416: 40–52. Communications in Computer and Information Science, Cham.

DHd (2016): Digital Humanities im deutschsprachigen Raum (DHd), "Call for Papers", (DHd 2017). Digitale Nachhaltigkeit, http://www.dhd2017.ch/calls.

Dojo (2017a): The Dojo Foundation, "Dojo Toolkit 1.11", https://dojotoolkit.org/ (March 2nd, 2017).

Dojo (2017b): The Dojo Foundation, "dojox.image.Lightbox", https://dojotoolkit.org/reference-guide/1.10/dojox/image/Lightbox.html (March 2nd, 2017).

Dubinko (2003): Dubinko, "XForms Essentials", Sebastopol, http://xformsinstitute.com/essentials/browse/.

Elliott et al. (2011–2014): Elliott, Bodard, Mylonas, Stoyanova, Tupman, and Vanderbilt, "EpiDoc Guidelines: Ancient Documents in TEI XML (Version 8)" The Stoa Consortium, http://www.stoa.org/epidoc/gl/latest/.

eXistdb (2017): eXist-db Project, "Package Repository", http://exist-db.org/exist/apps/doc/repo.xml (March 2nd, 2017).

Google (2017): Google Developers, "Google Charts. Interactive Charts for Browsers and Mobile Devices", https://developers.google.com/chart/.

Grün (2011): Grün, "Storing and Querying Large XML Instances", Konstanz, http://files.basex.org/publications/Gruen [2010], Storing and Querying Large XML Instances.pdf.

Heinz (2010): Heinz, "Monasterium.net – Auf dem Weg zu einem europäischen Urkundenportal", in: Regionale Urkundenbücher. Die Vorträge der 12. Tagung der Commission Internationale de Diplomatique, 14:139–45 (NÖLA. Mitteilungen Aus Dem Niederösterreichischen Landesarchiv), St. Pölten.

Henny (2013): Henny, "Aufbau eines Digitalen Textzeugenarchivs zum Altägyptischen Totenbuch" (Best Practice: Digitale Korpora. Workshop der Berlin-Brandenburgischen Akademie der Wissenschaften und der Union der deutschen Akademien der Wissenschaften, Berlin, August 10, 2013), https://prezi.com/pthztgusy4g8/aufbau-eines-digitalen-textzeugenarchivs-zum-altagyptischen-totenbuch/.

ICARUS (2017): ICARUS – International Centre for Archival Research, "Monasterium.net", http://monasterium.net/mom/home (March 2nd, 2017).

Krah (2009): Krah, "Monasterium.net - Das Virtuelle Urkundenarchiv Europas. Möglichkeiten der Bereitstellung und Erschließung von Urkundenbeständen", in: Generaldirektion der Staatlichen Archive Bayerns (ed.), Archivalische Zeitschrift 91, no. Sonderdruck, 221–46.

Lagoze et al (2005): Lagoze, Van de Sompel, Nelson, and Warner, "Implementation Guidelines for the Open Archives Initiative Protocol for Metadata Harvesting. Guidelines for Repository Implementers" Open Archives Initiative, http://www.openarchives.org/OAI/2.0/guidelines-repository.htm.

Legowski (2015): Legowski, "The Project Is Completed! What Now? The Ancient Egyptian Book of the Dead – A Digital Textzeugenarchiv" (Altertumswissenschaften in a Digital Age: Egyptology, Papyrology and Beyond, Leipzig, May 11, 2015), https://prezi.com/z2hcb_m_y815/the-project-is-completed-what-now/.

Müller et al. (2012a): Müller et al., "Das Altägyptische Totenbuch. Ein Digitales Textzeugenarchiv", http://www.totenbuch.awk.nrw.de/.

Müller et al. (2012b): Müller et al., "Totenbuchprojekt Bonn, TM 135212", in: Das Altägyptische Totenbuch. Ein Digitales Textzeugenarchiv, http://totenbuch.awk.nrw.de/objekt/tm135212.

Müller et al. (2012c): Müller et al., "Register Motive (Nach Gruppen)", in: Das Altägyptische Totenbuch. Ein Digitales Textzeugenarchiv, http://totenbuch.awk.nrw.de/register/motive-gruppen.

Müller et al. (2012d): Müller et al., "Totenbuch-Schema.", in: Das Altägyptische Totenbuch. Ein Digitales Textzeugenarchiv, http://totenbuch.awk.nrw.de/schema/totenbuch.xsd.

Müller et al. (2012e): Müller et al., "Register der Besitzer", in: Das Altägyptische Totenbuch. Ein Digitales Textzeugenarchiv, http://totenbuch.awk.nrw.de/register/besitzer.

Müller et al. (2012f): Müller et al., "Sprüche. Benachbarte Sprüche", in: Das Altägyptische Totenbuch. Ein Digitales Textzeugenarchiv, http://totenbuch.awk.nrw.de/uebersicht/sprueche#Nachbarschaft.

Müller et al. (2012g): Müller et al., "Dokumentation. Die Außenwelt, Schnittstellen und Metadaten", in: Das Altägyptische Totenbuch. Ein Digitales Textzeugenarchiv, 2012ff, http://totenbuch.awk.nrw.de/projekt/dokumentation#SchnittstellenMetadaten.

Müller et al. (2012h): Müller et al., "Totenbuchprojekt Bonn, TM 57201", in: Das Altägyptische Totenbuch. Ein Digitales Textzeugenarchiv, http://totenbuch.awk.nrw.de/objekt/tm57201.

Müller et al. (2012i): Müller et al., "Totenbuchprojekt Bonn, TM 134299", in: Das Altägyptische Totenbuch. Ein Digitales Textzeugenarchiv, http://totenbuch.awk.nrw.de/objekt/tm134299.

Pelagios (2017a): PELAGIOS Project, "Peripleo. A Search Engine for the Pelagios Universe, with a Comprehensive JSON API", https://github.com/pelagios/peripleo (March 2nd, 2017).

Pelagios (2017b): PELAGIOS Project, "Pelagios-Widgets. Embeddable JavaScript Widgets to Access Pelagios Data", https://github.com/pelagios/pelagios-widgets (March 2nd, 2017).

Piwik (2017): Piwik.org, "PIWIK. Open Analytics Platform." https://piwik.org/ (March 2nd, 2017).

Rößler-Köhler (1995): Rößler-Köhler (ed.), Handschriften des Altägyptischen Totenbuchs (HAT), Wiesbaden.

Rößler-Köhler (1998): Rößler-Köhler (ed.), Studien zum Altägyptischen Totenbuch (SAT), Wiesbaden.

Sahle/Henny (2012a): Sahle and Henny, "The New Book of the Dead Database" (Third International Colloquium for Book of the Dead Studies, Bonn, February 28, 2012), http://web.archive.org/web/20130422052335/https://www.totenbuch-projekt.uni-bonn.de/kolloquium-tb.

Sahle/Henny (2012b): Sahle and Henny, "Visualising the Book of the Dead" (unConference DHD, Hamburg, July 17, 2012), https://prezi.com/s_3muupnarfs/visualising-the-book-of-the-dead/.

Sahle/Henny (2013a): Sahle and Henny, "Aspekte Digitaler Forschung Am Beispiel Des Totenbuch-Projekts", ("Nicht für das Leben lernen wir, sondern für den Tod". Abschlussveranstaltung des Akademienprojektes "Edition des altägyptischen Totenbuches vom Neuen Reich bis zur Römerzeit," Düsseldorf, February 19, 2013), http://www.awk.nrw.de/veranstaltungen/veranstaltungsrueckblick/2013/totenbuch-projekt-abschlussveranstaltung.html.

Sahle/Henny (2013b): Sahle and Henny, "Egyptology Meets Digital Humanities: The Book of the Dead" (Digital Classicist Berlin Seminars, Berlin, August 1, 2013), http://de.digitalclassicist.org/berlin/2013/01/07/Sahle-Henny.

Siegel/Retter (2014): Siegel and Retter, "eXist. A NoSQL Document Database and Application Platform", Sebastopol.

Turner (2010): Turner, "JavaScript in betterFORM." betterFORM, https://betterform.wordpress.com/2010/06/16/javascript-in-betterform/ (June 16, 2010).

Universität Hamburg (2016): Universität Hamburg, Fakultät für Geisteswissenschaften, "Forschungsdaten in den Geisteswissenschaften (FORGE 2016) — Jenseits der Daten", https://www.gwiss.uni-hamburg.de/gwin/ueber-uns/forge2016.html (March 2nd, 2017).

W3C (2009): World Wide Web Consortium (W3C), "XForms 1.1. W3C Recommendation 20 October 2009", https://www.w3.org/TR/xforms/ (March 2nd, 2017).

W3C (2017): World Wide Web Consortium (W3C), "XML Technology", https://www.w3.org/standards/xml/ (March 2nd, 2017).

7. List of Images

Figure 1: Basic structure of the data model for a BoD object; by Ulrike Henny.

Figure 2: Outline of the technical framework behind the BDA; by Ulrike Henny.

Figure 3: Faceted browsing (example); screenshot from website (2017).

Figure 4: Description of object TM 135212, wooden board, Abusir (detail); screenshot from website (2017, http://totenbuch.awk.nrw.de/objekt/tm135212).

Figure 5: Visualisation of spell neighbourship on the books (preview); screenshot from website (2017, http://totenbuch.awk.nrw.de/static/overviews/benachbarte-sprueche-absolut-gerade.svg).

Authors[45]

Ulrike Henny
Universität Würzburg
Institut für Deutsche Philologie/ Lehrstuhl für Computerphilologie
Am Hubland
97074 Würzburg

Email: ulrike.henny@uni-wuerzburg.de

Jonathan Blumtritt
Universität Köln
Data Center for the Humanities (DCH)
Albertus-Magnus-Platz
50923 Köln

Email: jonathan.blumtritt@uni-koeln.de

Marcel Schaeben
Universität Köln
Cologne Center for eHumanities (CCeH)
Albertus-Magnus-Platz
50923 Köln

Email: m.schaeben@uni-koeln.de

Prof. Dr. Patrick Sahle
Universität Köln
Cologne Center for eHumanities (CCeH)
Albertus-Magnus-Platz
50923 Köln

Email: sahle@uni-koeln.de

Illuminating the Herculaneum Papyri: Testing new imaging techniques on unrolled carbonised manuscript fragments

Kathryn E. Piquette

Abstract: This article describes the results of advanced imaging pilot testing on unrolled Herculaneum Papyri undertaken in 2014 as part of a collaboration between the Universität zu Köln and the Biblioteca Nazionale di Napoli.[1] Building on results obtained by a Brigham Young University team in 1999–2002 using near infrared (NIR) photography, the aim was to test the potential for improving legibility and documentation of surface structures by combining the non-destructive techniques of NIR and Reflectance Transformation Imaging (RTI). The results show that, in addition to achieving super high-resolution images with enhanced contrast between the black ink and carbonised papyrus, NIR-RTI enables more accurate recording of surface morphology and enables the user to clarify the relationships between the ink constrate and the papyrus substrate. The article also identifies areas for improving image acquisition and usability.

1. Introduction

This article concerns advanced imaging undertaken in 2014 on unrolled Herculaneum Papyrus fragments in order to assess the potential of integrating near infrared (NIR) photography with Reflectance Transformation Imaging (RTI) for improved legibility and understanding of the scrolls' physical structure. In what follows, I briefly introduce the dataset and summarise previous imaging work and some of the persistent challenges that motivated the pilot testing. The imaging hardware, software, capture technique and processing methods are then described,

---

followed by presentation and assessment of selected results. Proposals for developing the imaging technique and user tools will be offered, as well as observations on integrating new imaging techniques into scholarly workflows. At this juncture, I should note that I come to this work not as a Herculaneum papyrologist nor as a formally trained imaging scientist, but as an Egyptologist whose research revolves around archaeological approaches to ancient script[2] with the aid of advanced digital imaging technologies. For several years now I have been working with dome-based and Highlight RTI techniques in museum and field contexts,[3] but it is thanks to this pilot project and the support of numerous colleagues that I have had the tremendous opportunity to delve into the world of the Herculaneum papyri and spectral imaging.

2. The Herculaneum Papyri and Previous Imaging

The survival of Herculaneum papyri is a direct result of a catastrophic eruption of Mount Vesuvius in 79 CE, which buried the coastal city of Herculaneum and neighbouring Pompeii. Pyroclastic flows of superheated gas, steam and mud overwhelmed these cities, causing great destruction and loss of life. At the same time, the resultant rapid rise in temperature to 300–320 degrees Celsius acted to carbonise and preserve various organic materials, including a portion of a large library of papyrus scrolls.[4] The site of Herculaneum lay buried until the mid-1700s when antiquarian activity led to the discovery of some 1000 charred scrolls in the so-called '*Villa dei Papiri*'. An unknown number were destroyed before the identity of these crumpled, blackened cylindrical objects was recognised, with many more being partially or completely destroyed during attempts to unroll and read them. Certain documentation methods, such as copying the exposed inscribed surface and then scraping it away to reveal the layer underneath, resulted in yet further loss.

Of those 'successfully' opened and sufficiently legible, the majority contain Greek writings including treatises from the Epicurean philosopher Philodemus of Gadara, while a smaller number (c.80) preserve Latin poetry and prose. The Herculaneum papyri are thought to be part of an ancient private library that may have belonged to Lucius Calpurnius Piso Caesoninus, a Roman politician and father-in-law of Julius Caesar.[5] Hundreds of unrolled fragments and a number of rolled scrolls are now stored in the Biblioteca Nazionale di Napoli, the former consisting of about 6000 fragments mounted on numerous trays or "*cornici*", with between 2–30 fragments per *cornice*. The papyrus fragments I worked on were mounted on animal gut, in turn, glued to card (*cartoncino*) affixed with thumbtacks to a thin wooden board measuring about 50cm × 30cm and stored in a metal tray with a glass lid (Figure 1). The wooden board mount could be removed from the metal tray for photography, but further dismounting was not possible.

---

2   E.g. Piquette (forthcoming); Piquette (2008); Piquette and Whitehouse (2013).

3   E.g. Earl et al. (2011); and Piquette (2016), respectively.

4   Janko (2002), 25.

5   Gigante and Obbink (2002), 49.

Figure 1: PHerc. 1506, Cornice 01 in storage box (Kathryn E. Piquette, courtesy Biblioteca Nazionale di Napoli).

The carbonised papyri with their black ink writings are notoriously difficult to read due in great part to their crumpled shape. In contrast to the usually planar surfaces of parchment and paper,[6] the shape of papyrus pith is inherently uneven at the micro-level, and even more so for the Herculaneum Papyri, which are uniquely complex in their meso- and macro-geometry (Figure 2). During pyrolysis – the process of carbonisation – water present in the papyrus substrate evaporates, causing buckling, twisting and shrinkage (estimated to be up to 30% of original scroll bulk).[7] Moreover, these effects are not uniform across a given scroll or the 'corpus'. Found in different locations within the villa, the scrolls would have been susceptible to different environmental conditions, as would sections of papyrus within the scroll itself, depending on whether closer to the inside or outside and depending on the presence of an umbilicus – a central baton around which the papyrus was wound. The saturation of the black colour of different fragments from the same or different scrolls therefore varies, and contrasts to different degrees with the ink. Likewise, the types of pigment, binder, or drying agent used in the ink and the techniques of its application will influence how it now contrasts with the papyrus substrate under visible light or other wavelengths.[8]

---

6   But cf. Pal et al. (2013).

7   Janko (2002), 30.

8   Tack et al. (2016).

Figure 2: PHerc. 78, Cornice 08 detail showing complex shape of papyrus (Sarah Hendriks, courtesy Biblioteca Nazionale di Napoli).

A variety of methods to increase contrast between the papyrus substrate and ink constrate have been attempted over the centuries. Recent efforts have been directed toward developing a non-destructive means of accessing the interior of scrolls that are still rolled and these have achieved remarkable results. X-ray computed tomography has proved successful for discerning the layers (and umbilicus) within the roll[9] while enhanced X-ray phase contrast tomography is bringing us closer to recording the ink as it sits on the papyrus within a scroll.[10] Otherwise, efforts have focussed on the unrolled fragments in Naples and it is on this material that our pilot testing was undertaken. The particular work that informs our own began over 15 years ago, as part of a two-phase spectral imaging project undertaken by a team from Brigham Young University (BYU). Commencing with tests in the autumn of 1999, the BYU team, supervised by Steve Booras, proceeded for over a year and a half to capture more than 30,000 photographs[11] of 1600 unrolled fragmentary papyrus scrolls.[12]

The equipment for the photography included a Kodak MegaPlus digital camera with standard Nikon 50mm and 100mm macro lenses, and produced images with a 300–600 pixels per inch (ppi) resolution. A motorised filter wheel with eight narrow bandpass filters was used. Illumination was applied using a German-made Dedo variable power lighting system that

---

9   Seals et al. (2011).

10   Mocella et al. (2015); Seales et al. (2016).

11   Information about the papyri can be searched here, although access to digital images is currently restricted: http://guides.lib.byu.edu/c.php?g=216482&p=1429231.

12   Booras and Seely (1999); Chabries et al. (2003), 368–371.

ensured the ambient temperature around the papyrus did not exceed 85° F.[13] Booras and Chabries report that attempts to increase contrast through near infrared radiation resulted in different responses from different carbonised papyri[14] but for most Herculaneum papyri, the optimal contrast response to NIR was between 850–1000 nanometres (nm).[15]

The clarity of the NIR images exceeded all expectations and launched a major leap forward in Herculaneum Papyri research, occupying papyrologists for well over a decade with enhanced and newly visible written content. Despite the major contribution of these visualisations, as with any new technique or evidence, new questions and issues arise. Certain technical drawbacks have been noted in publication[16] and also flagged up during my own discussions with Herculaneum papyrologists:

1. NIR photographs appear quite flat
2. Focus too soft in some areas
3. Fixed light position results in:
    a. Incomplete documentation of surface reflectance properties
    b. Difficulty discerning papyrus layers (*sottoposti* and *sovrapposti*)
    c. Difficulty discerning black ink from black colour of holes in papyrus
    d. Surface details obscured by self-shadowing
4. Overhead perspective and undulating surface results in distortion of letter shape
5. Some registration problems
6. Absence of scale bars

While gains in contrast and image resolution have vastly expanded access to the written content, conventional digital photography unavoidably produces fixed-light visualisations regardless of the light wavelength applied. Several of the shortcomings listed above are the direct result of this fixed relationship between the light source and complex geometry of the surface (3a–d), as well as the fixed angle between the camera and subject (3a, 4). Other issues relate to hardware, and imaging decisions or possible oversights (1, 5, 6).

Advances in hardware, software, capture technique and image processing, especially over the past decade, are now able to provide solutions to these issues. The computational photographic technique, Reflectance Transformation Imaging (below), has proven extremely helpful in addressing similar visualisation problems for, *inter alia*, writing on metal tablets from the Classical world, e.g. thin sheets of lead or other metals inscribed with a stylus, rolled or folded and ritually deposited.[17] Upon rediscovery, they are often unrolled for reading, their corroded and undulating surfaces presenting significant reading and recording challenges.[18] In late 2013, I had begun work applying RTI to lead tablets from the ancient Near East inscribed with Greek magical texts for the *Magica Levantina* project based at Universität zu Köln and led by Robert Daniel. During review of preliminary RTI results with colleagues, Herculaneum papyrologist Jürgen Hammerstaedt suggested the technique be trialled on the Herculaneum Papyri fragments in Naples with the aim of assessing whether the spectral imaging method employed by Brigham Young University could be fruitfully augmented to address some of the shortcomings described above.

---

13   Booras and Chabries (2001), 3.

14   The reasons for this, no doubt contingent upon the variable micro and macro conditions of pyrolysis (above), have yet to be fully explored and present a promising area for spectral research.

15   Booras and Chabries (2001), 3; cf. Janko (2002), 28.

16   Macfarlane et al. (2007); Janko (2002), 29–30.

17   E.g. Gager (1992).

18   E.g. Piquette et al. (2011).

With the kind cooperation of the Biblioteca Nazionale di Napoli, two phases of pilot testing were planned and undertaken in February[19] and June 2014. The aim of Test 1 was to assess generally the effectiveness of Reflectance Transformation Imaging (RTI) for better capturing papyrus surface shape. Depending on the outcome, Test 2 would then combine the Highlight RTI capture technique with NIR photography for full comparison with the previous NIR photography.

3. Test 1: Visible Highlight Reflectance Transformation Imaging

For Test 1 undertaken on 6–7 February 2014, I used the manual highlight procedure for applying RTI. RTI is a method of structured light photography where a series of exposures are made, each with illumination applied from a different location and angle in a hemispherical configuration. This can be accomplished with a lighting dome[20] or illumination arch,[21] but for our purposes Highlight RTI (H-RTI)[22] presented a relatively affordable option with greater flexibility and portability for low-impact use in a library context.

H-RTI requires a tripod-mounted camera set up in a stable environment on a solid surface. Stability between both the camera and the subject is essential so that the shots register precisely, pixel on pixel. Only the light source moves. As set out in *Reflectance Transformation Imaging: Guide to highlight image capture* v2.0,[23] the photographer moves the flash around the subject, iteratively applying a full hemisphere of illumination. Systematic application of the light is aided with the use of a string for measuring distance and ensuring the centre of the light is aligned with the centre of the subject. The placement of one, or preferably two, reflective spheres in shot enables the position of the light source to be calculated for each exposure during processing. The shots are fitted together using the Polynomial Texture Mapping (PTM)[24] or the Hemispherical Harmonics (HSH)[25] mathematical algorithms. The resultant image file can be viewed in an RTI viewer, allowing the user to virtually relight the surface and apply enhancements to disclose features of potential significance that are not apparent during first-hand inspection or in conventional fixed-light photographs.

The H-RTI imaging setup for Test 1 was composed of Canon EOS 5D Mark III digital SLR (22.3MP) with a Canon EF 100mm f/2.8L IS USM Macro lens (with a B&W UV filter for lens protection). The illumination source was a Canon Speedlite 600EX-RT flashgun[26] controlled and synced with the shutter using a Canon-compatible Calumet Pro Series 4 Channel Wireless Trigger Kit (including transmitter and receiver) and an extra receiver to allow single operator functionality. The camera was mounted on Gitzo carbon fibre tripod on a Gitzo ball head on a geared column. The camera was positioned vertically on the reverse end of the column on a bottom plate, facing down to bring the camera close to the papyrus. This is also safer

---

19   Test 1 was undertaken with Highlight RTI system I built during a Marie Curie COFUND fellowship at the Dahlem Research School, Freie Universität Berlin, in cooperation with Excellence Cluster TOPOI.

20   E.g. Earl et al. (2011).

21   E.g. Webb and Wachowiak (2011).

22   Mudge et al. (2010); (2006).

23   Cultural Heritage Imaging (2010).

24   See Malzbender et al. (2001).

25   Gautron et al. (2004).

26   As a typical xenon flashtube, this has a fairly linear output in the range of visible light (400nm–700nm), but also has a powerful output in the near infrared region above 800nm. Dyer et al. (2013), 45, figs. 2–3.

for the fragile subject since the centre of gravity is between the tripod legs and reduces the likelihood of the tripod tipping under the weight of the macro lens – as compared with the camera mounted on the centre column oriented at 90°, although tripod weights aid stability and are advisable for all setups. This setup was installed on two large desks (Figure 3). Ideally a single surface should be used but vibration tests confirmed stability of the floor and overall setup. The camera was tethered to a laptop on a separate surface to ensuring that hard drive or fan spin did not compromise stability.

Figure 3: Test 1 workspace comprised of two large desks pushed together in upper room of library with stable floor. Mariacristina Fimiani, CISPE collaborator, looks on (Kathryn E. Piquette).

A selection of fragments exemplifying the issues described above was prepared by Gianluca Del Mastro and a subset were imaged (Table 1). Based on conversations with potential users regarding comfortable onscreen viewing of written characters measuring approximately 3mm–4mm at up to 100% magnification, a preference for resolution was established at 200 pixels per centimetre (ppc).[27] After taking account of the average width of a column of text and need to fit in a scale and two RTI spheres in shot, in most cases a 10cm–12cm area of papyrus was imaged at 100ppc–130ppc. Many fragments included in the test were larger than 10cm–12cm, necessitating consultation of two separate RTI files to view a single column of text (devising capture procedures that enable the stitching of RTI datasets remains an area for further work[28]).

---

27 Cf. Twycross (2008), 27–29.

28 Lindsay MacDonald (2015), 104–105 has developed a method for manual stitching data from two different RTI capture sequences post capture.

Test 1: Visible H-RTI
PHerc. 395, Cornice 04, Pezzo 01, right
PHerc. 862, Cornice 01, Pezzo 01, right
PHerc. 862, Cornice 01, Pezzo 02, right
PHerc. 1485, Cornice 10, Pezzo 01, lower left
PHerc. 1506, Cornice 32, Pezzo 01, lower left
PJoannowsky, Cornice 03, left

Table 1: Herculaneum papyri imaged for Test 1 using visible H-RTI.

Capture entailed taking about 48 images as well as a shot with the same camera settings of the X-Rite ColorChecker Passport for controlling colour, white balance and exposure. Cultural Heritage Imaging (hereafter CHI)[29] processing guidelines were followed. Adobe Lightroom 5 was used for pre-processing, including white balance and exposure adjustment and conversion of the RAW files to DNGs (with RAW embedded) and JPG. The latter were used with RTIBuilder 2.0.2[30] for processing and fitting into RTI user files. Both PTM and HSH formats were produced. The latter provided the best visualisation given that it calculates the surface normals more accurately and deals with self-shadowing more successfully than the PTM algorithm.[31] In the RTI viewer, the settings that produced the best visualisations were the specular enhancement rendering mode and the normals visualisation (see examples below).

The results demonstrated the success of the H-RTI technique, especially when processed with the HSH algorithm (Figure 4). However, the problem of contrast between the black ink and blackened papyrus remained largely unaddressed, necessitating a second test involving the integration of an NIR light source in the H-RTI capture procedure.

---

29   Cultural Heritage Imaging (2011).

30   Barbosa (2011).

31   Happa et al. (2010), 161–162.
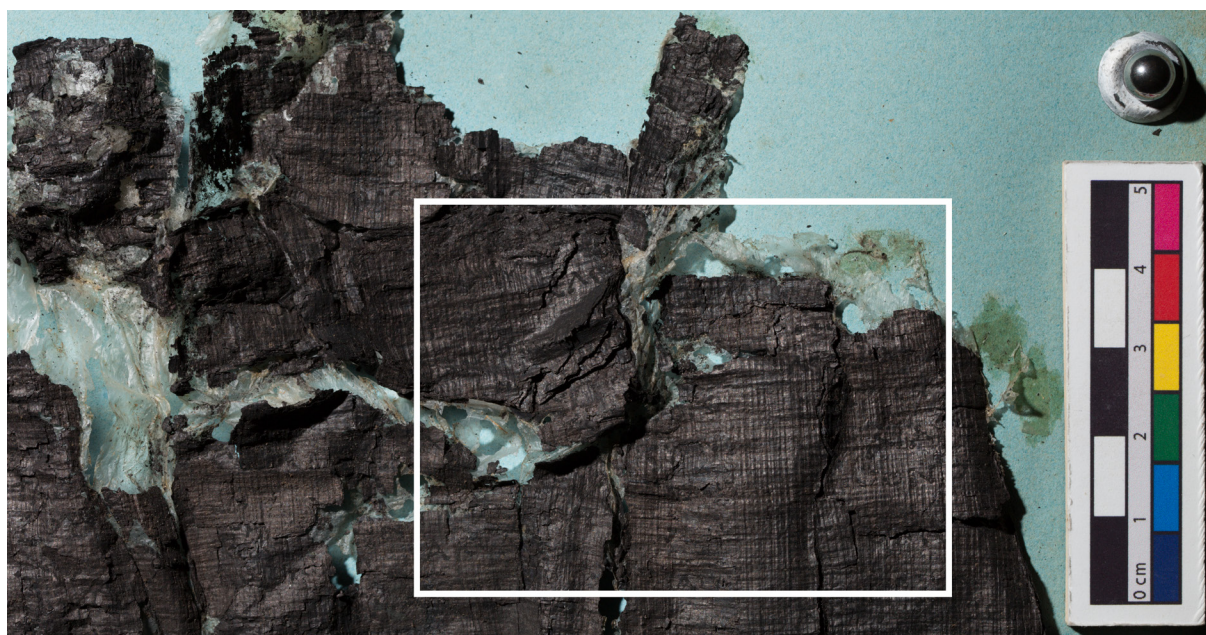
Figure 4: PHerc. 862, Cornice 01, Pezzo 02, detail of right side. a. As shot for the visible H-RTI Test 1. b. Detail of BYU spectral image (righam Young University, courtesy Biblioteca Nazionale di Napoli). c. Detail of RTI image (HSH), with the Specular Enhancement rendering mode applied (a. and c. Kathryn E. Piquette, courtesy Biblioteca Nazionale di Napoli).

4. Test 2: Near Infrared Highlight Reflectance Tranformation Imaging

Test 2 took place between 11–19 June 2014, the aim of which was to combine the Highlight RTI capture technique with near infrared photography and compare the results with the previous BYU NIR photography. With a modest project budget and extra-institutional collaboration, I was able to assemble an NIR H-RTI setup. The hardware used for Test 2 was composed of a UV-VIS-NIR modified[32] Nikon D800 digital SLR (36MP) with a CoastalOpt UV-VIS-NIR

---

32   Modification performed by Foto Gregor in Cologne, Germany.

APO, 60mm Macro lens with F74-F-45 F-mount,[33] and using a IR 950nm Delamax longpass/lowpass optical filter[34] (allows near infrared light with a wavelength greater than 950nm to pass but excludes shorter wavelengths). A near infrared BW® 48 LED Illuminator[35] with custom-made aluminium foil light throw provided a continuous light source (Figure 5). The camera was mounted on a Manfrotto carbon fibre tripod on a Gitzo ball head with the column reversed so the camera faced downward (Figure 6). Table 2 lists the fragments imaged during Test 2.

Figure 5: Near infrared BW® 48 LED Iluminator (30° beam spread) employed for Test 2 with customised light throw attached.

Figure 6: The author with the NIR H-RTI setup in the Biblioteca Nazionale di Napoli (Michael McOsker).

---

33 For equipment support, thanks are due to the UCL Centre for Digital Humanities, Multi-Modal Digitisation Suite, and Stuart Laidlaw of the UCL Institute of Archaeology.

34 With 58 > 52 adaptor ring.

35 This item was purchased on Amazon.co.uk and while the details did not state the wavelength of emission, very similar models elsewhere were stated as outputting at 850nm. It was not possible to confirm this with interferometric testing.

Test 2: Near Infrared H-RTI
PHerc. 862, Cornice 01, Pezzo 01, right
PHerc. 862, Cornice 01, Pezzo 01, upper right
PHerc. 862, Cornice 01, Pezzo 01, lower right
PHerc. 994, Cornice 09, Column 22, upper
PHerc. 994, Cornice 09, Column 22, lower
PHerc. 994, Cornice 09, Column 23, upper
PHerc. 994, Cornice 09, Column 23, lower
PHerc. 994, Cornice 09, right piece, upper left
PHerc. 994, Cornice 09, right piece, upper right
PHerc. 994, Cornice 09, right piece, lower left
PHerc. 994, Cornice 09, right piece, lower right
PHerc. 1506, Cornice 01, Pezzo 01
PHerc. 1506, Cornice 01, Pezzo 04
PHerc. 1506, Cornice 01, Pezzo 02, left
PHerc. 1506, Cornice 01, Pezzo 02, right
PHerc. 1506, Cornice 01, Pezzo 01
PHerc. 1506, Cornice 01, Pezzo 03
PHerc. 1506, Cornice 02, Pezzo 05, left
PHerc. 1506, Cornice 02, Pezzo 05, middle
PHerc. 1506, Cornice 02, Pezzo 05, right
PHerc. 1533, Cornice 01, Pezzo 05
PHerc. 1533, Cornice 01, Pezzo 04
PHerc. 1533, Cornice 01, Pezzo 01
PHerc. 1533, Cornice 01, Pezzo 02
PHerc. 1533, Cornice 01, Pezzo 03

Table 2: Herculaneum papyri imaged for Test 2 using NIR H-RTI.

As with any technological innovation, a period of familiarisation and problem solving is par for the course and is subject to a particular budgetary and scheduling framework. A literature survey in preparation for the work revealed only a small number of online resources and journal publications reporting on the integration of IR with RTI.[36] Eleni Kotoula kindly shared aspects of her doctoral research including NIR H-RTI work on the Derveni papyrus. Her advice was immensely valuable for planning this second phase, along with guidance from several other colleagues. Whereas Kotoula[37] was able to employ transmitted NIR for her work on the glazed, card-mounted Derveni papyrus (as well as reflected NIR), the mounting of the Herculaneum fragments precludes transmission.

Incandescent lamps output IR but produce heat that may be damaging to archaeological materials, in contrast to NIR LEDs which do not accelerate thermal degradation.[38] The choice of an LED illumination source followed Kotoula's lead, although the particular LED unit I was able to obtain within the available time and budget was quite low in power with a narrow beam spread. These combined to create several challenges. The beam spread limited the size

---

36   Caine et al. (2011); Gabov (2010); Kotoula (2014).

37   Kotoula (2013).

38   Kotoula (2014).

of the area that could be evenly illuminated while near invisibility made it difficult to ensure accurate light direction.[39] An improvised light throw of aluminium foil improved coverage slightly (Figure 5), but some vignetting did occur. Fortunately, the sections to be imaged were 10cm–12cm in length or width (for resolution reasons noted above) and even illumination could just about be achieved when holding the light at the recommended distance of 3×–4× the diagonal measurement of the target surface.[40] Although the ideal ISO setting for RTI capture is 100, the low power of the illumination source required an increase in both ISO and exposure time. ISO 400 produced a noticeable amount of visual noise, but a usable result was obtained with ISO 200 and an exposure time of 1/4s. It was also possible during processing and viewing, as described below, to enhance image brightness. Low light strength also necessitated setting focus with a more powerful light source. A modelling lamp on a mini studio strobe was placed at a distance from the camera setup and turned on only briefly to set focus (Figure 6, left).[41] Compared with visible RTI acquisition using a speedlight or similar source, NIR capture was somewhat slow and cumbersome. The exposure time required holding the light source steady while maintaining somewhat contorted positions as I progressed around the setup. Efficiency was also impacted by the LED illuminator being powered from the mains, necessitating careful manoeuvring with a trailing power lead.[42] A short safety lanyard attached to the unit and my wrist ensured that if I dropped it, the illuminator would not come into contact with the papyrus. With more time and funds, and as technology improves, a battery-powered IR LED source with higher and adjustable power and wider beam spread will provide more accurate illumination, increased capture speed, and reduce vibration risks.

The reader may rightly ask why a speedlight flash unit was not used. Time constraints in gaining access to funds for purchasing hardware, complex logistics involved in camera modification and its reunification with the rest of the new system left limited time for familiarisation and testing of the full setup before deployment on-site in Naples. Previous flawless performance of Calumet wireless triggers with the Canon setup used for Test 1 led to misplaced (in retrospect) confidence in a repeat performance with the new Nikon configuration with Calumet wireless triggers (Nikon brand triggers could not be afforded within the project budget). It was discovered upon arrival in Naples that the Calumet triggers sold as compatible with the then new Nikon D800 were, in fact, incompatible and syncing the shutter and flash proved impossible. After Test 2 was completed, an afternoon of troubleshooting back in London with Calumet and Nikon specialists resulted in a workaround being devised. By using the Nikon Commander Mode and pop-up flash for IR communication with the Speedlight, correct sync was achieved. So that the pop-up IR flash does not, however, interfere with the RTI IR light source, a Nikon FXA10358 IR Panel can be used to cover the built-in flash and prevent its influence on exposure. It is worth noting that, as was discovered during subsequent use, if this panel sits flat against the pop-up flash, it may interfere with Speedlight communication when in certain positions, but if the panel is tipped away slightly, communication is maintained as the Speedlight is moved around the setup during capture. As tests demonstrated, the small amount of leakage from the pop-up IR flash does not influence the exposure.

---

39 Such NIR LED units often have a built-in light sensor, and in order to ensure constant, full power illumination, the sensor must be covered. Removing the protective glass cover, covering the sensor with Blu-Tack, and replacing the unit cover achieved this.

40 CHI (2013), 18.

41 Because most of the energy emitted by such tungsten or tungsten-halogen bulbs is in the NIR range and produces heat that may degrade papyri (Leach and Tait (2001), 239–240), use of the modelling lamp was kept to a minimum.

42 Use of a battery pack is advisable to avoid trailing cables that might present tripping hazards or bump the setup. However, I was able to keep the lead clear by attaching and trailing it up my arm and over my shoulder with Velcro strips.

Returning to the workflow of Test 2, in order to record the position of the light in each exposure in the series, two reflective ball bearings were placed in shot. These were machined down beforehand at UCL but ensuring their original diameter was preserved. This truncation had the benefit of reducing sphere height and thus minimising the depth of the effective focus range required to accommodate the upper third of the spheres along with the papyrus surface. This is important as an aperture smaller than f/11 should be avoided as this results in a loss of sharpness.[43] The spheres were carefully placed on the *cartoncino* or, to avoid surface contact, affixed to chopsticks suspended above the papyrus at the edge of the framed view, together with a scale. At the end of each capture sequence with camera settings unchanged, an X-Rite ColorChecker Passport was imaged with illumination from 65°.

As above, CHI processing guidelines[44] were followed, although additional steps were required for the NIR data. Because the image sensor of the modified camera is colour, the image will have a reddish cast. The images were therefore converted to greyscale before conversion to DNG and JPG. As mentioned above, the exposure achieved with this particular LED light source was usable but suboptimal. In order to try to brighten up the image while avoiding degradation, I conducted a series of tests where by the JPGs required for creating RTI image files were exported with incrementally higher exposure adjustment (i.e. +0.7, +1.0, +1.5, +2.0, +2.5, +3.0), followed by processing with RTIBuilder. Comparison of surface normals visualisations at 200% magnification in the RTI viewer revealed no visible[45] degradation for adjustments between +0.7 and +2.0, but did begin to appear around +2.5. I therefore processed with a +2.0 exposure adjustment. It is important to note that in all cases when the capture series was fitted using the Polynomial Texture Mapping (PTM) algorithm,[46] the results were too dark to be usable. However, excellent visualisations were produced when the data was processed with the Hemispherical Harmonics (HSH) algorithm and when viewed in the RTI viewer using the Specular Enhancement rendering mode with the following settings: Diffuse color = 0; Specularity = 100; Highlight Size = 1.

5. Assessment of Results

Throughout the pilot testing and especially following the completion of Test 2, user files were reviewed with Herculaneum papyrologists, including Jürgen Hammerstaedt, Gianluca Del Mastro and Richard Janko.[47] Their assessment was that the integration of NIR photography with the H-RTI capture technique presented a marked improvement over previous imaging methods and provided a vital complement to first-hand inspection using microscopes and raking light. However, barriers to efficient use include the large file size and need to consult multiple files for a single papyrus document due to the present inability of RTI captures to be stitched. For exposed layers of unrolled Herculaneum papyri, this technique with the ability to virtually relight the surface addresses most of the aforementioned shortcomings of fixed-light photographs.

---

43   CHI (2013), 26.

44   Cultural Heritage Imaging (2011).

45   Examination was conducted visually on a Retina Display MacBook Pro with 15.4-inch colour display with 2880×1800 native resolution at 220ppi.

46   Malzbender et al. (2001).

47   See also Janko (2016). Note that the uncredited NIR-RTI research results discussed on p. 126 and illustrated in fig. 5 are the work of the present author while in the employ of Universität zu Köln, and the present publication constitutes the primary publication of that work.
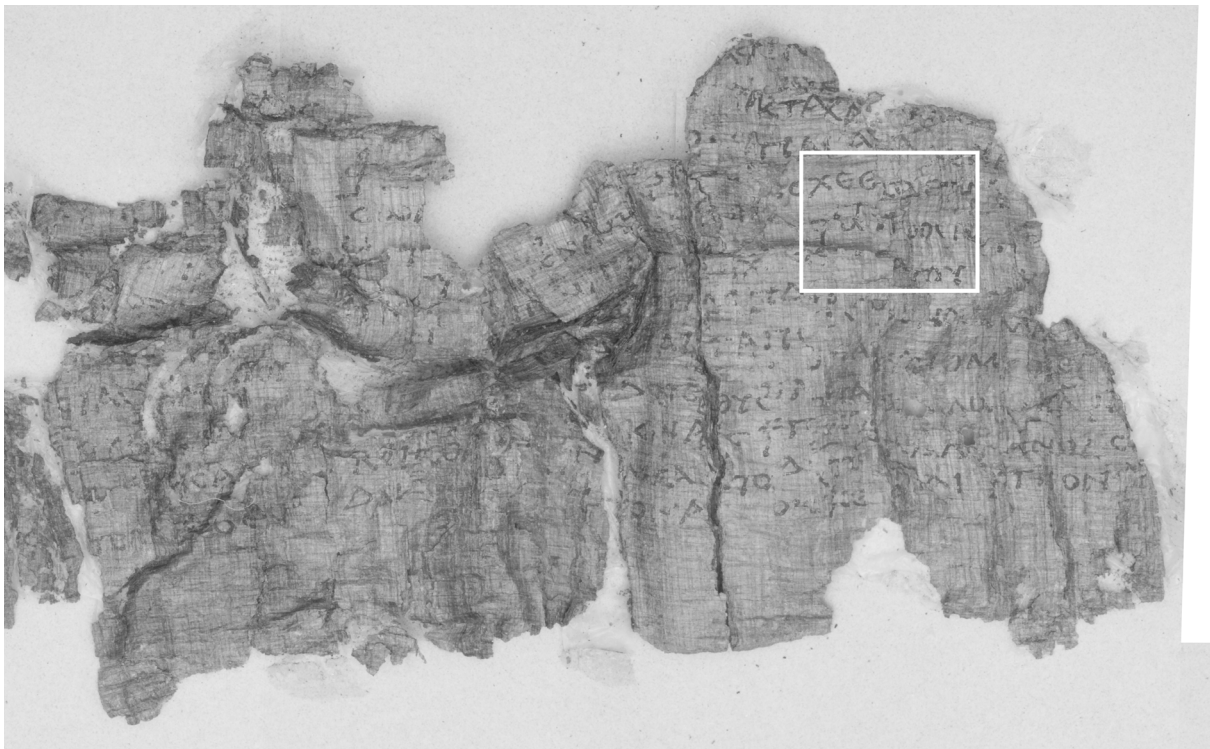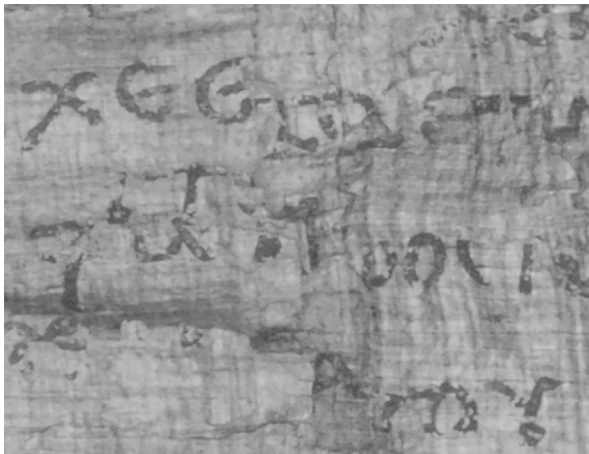
Figure 7: PHerc. 862, Cornice 01, Pezzo 01. a. NIR photograph (950nm) (Brigham Young University, courtesy Biblioteca Nazionale di Napoli). b. Detail. c–d. RTI details, Specular Enhancement rendering mode with light in different positions showing physical structure and relationship between ink and layers (Kathryn E. Piquette, courtesy Biblioteca Nazionale di Napoli).

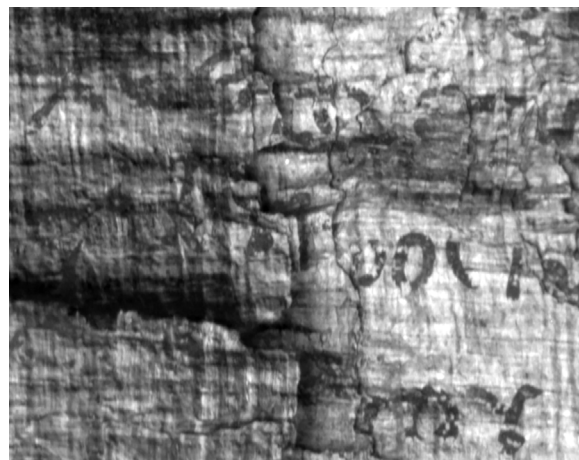Writing previously lost in the dark folds of crumpled fragments can be illuminated. Movement of the light also enables the user to clearly distinguish the edges of dark features such as holes that may be confused with ink or where apparently continuous ink sits on different layers.[48] Likewise, by zooming in and slowly moving the light around an area of interest at low angles, the edges of papyrus layers can be iteratively and systematically traced (Figure 7). The normals visualisation mode offers easier discernment of misplaced fragments, showing in detail the micro-morphology of the pith including the direction and width of the fibres (Figure 8) – data that could aid fibre matching and virtual reconstruction in the future.[49]

a.

b.
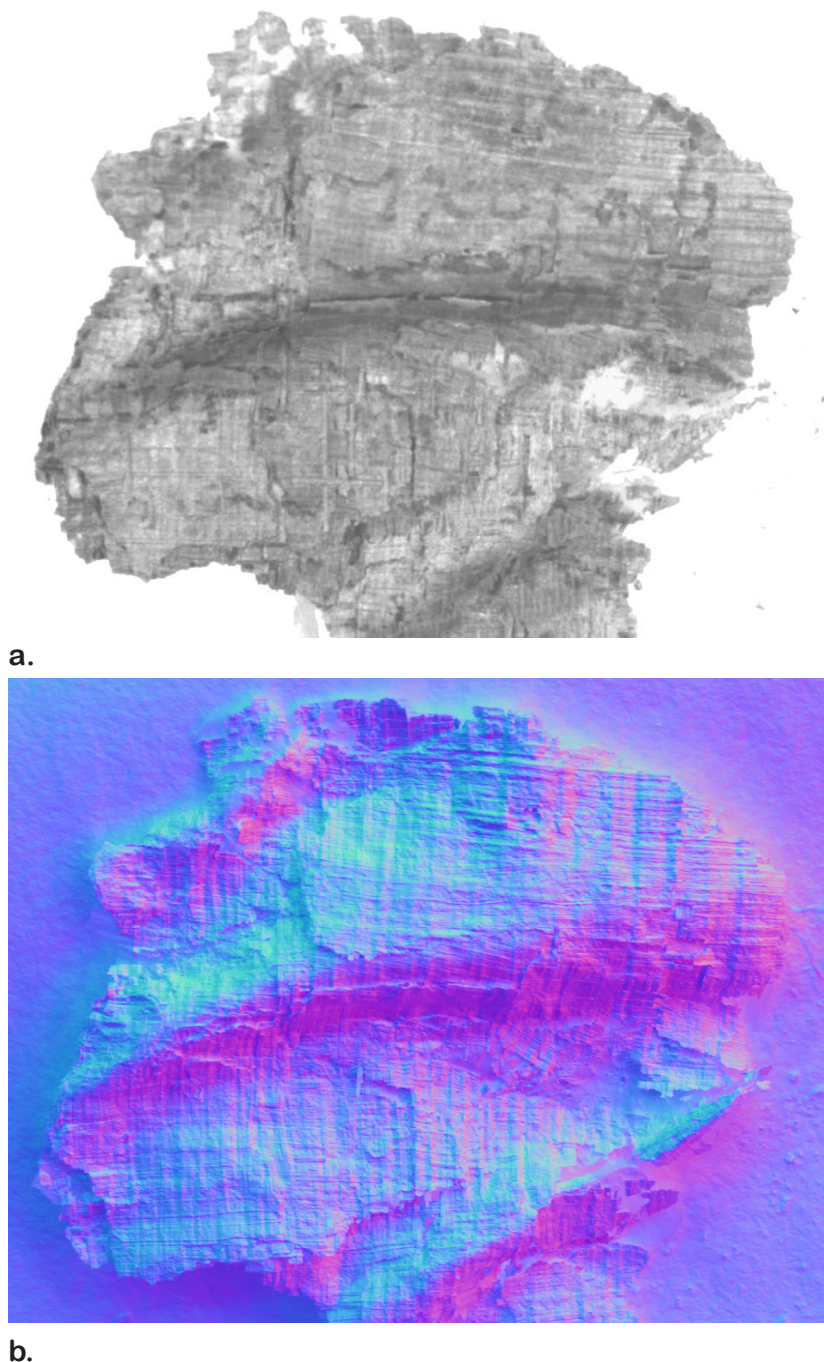
Figure 8: Detail of PHerc. 1506, Cornice 01, Pezzo 01. a. NIR photograph (950nm) (Brigham Young University, courtesy Biblioteca Nazionale di Napoli). b. Normals Visualisation showing the high degree of micro-geometry that can be captured with NIR H-RTI (Kathryn E. Piquette, courtesy Biblioteca Nazionale di Napoli).

---

48   See Janko (2016), 124, fig. 4.

49   See also Stökl Ben Ezra (2014), 95–96; Tov (2015), 267–288.

NIR RTI also enables more comprehensive, systematic and in-depth remote study. Visualisation modes and bookmarking and annotation tools such as those in the Cultural Heritage Imaging RTIViewer[50] play a particularly critical role in supporting processes of "seeing" and "knowing". Observation conditions and viewing experiences that have long been subjective and personal can now be recorded, shared and replicated, making the research endeavour more robust and potentially more collaborative and transparent.[51] Because the RTI images are very high-resolution and visualisations provide excellent contrast and a superb record of surface shape and texture, in-depth desk-based research can be undertaken prior to first-hand inspection in Naples – if not removing this necessity entirely in some cases. By potentially reducing the need and length of time for handling, NIR RTI can perform a valuable preservative function by minimising papyrus document movement and consequent deterioration from exposure to light and other environmental changes, as well as reducing demands on library and conservation staff time. RTI also offers exciting potential for democratising access to the unrolled Herculaneum papyri, given that RTI viewing software is freely available and easy to use.

6. Future Directions

NIR RTI offers a power tool for studying the unrolled Herculaneum fragments, as a complement to both previous documentation and first-hand study. In order to exploit the potential of this technique on the large scale required for the corpus, certain technical challenges need to be met. Highlight RTI can be relatively time-consuming. For imaging large quantities of papyrus fragments/*cornici* with similar visualisation requirements at a set resolution, hardware and software that allow for semi-automated capture and processing will doubtless reap time-cost savings and ensure standardisation in data capture and metadata collection – similar to the way in which many RTI lighting dome systems operate.[52]

A current limitation of RTI is that, depending on camera sensor size and the size of the smallest significant character or other feature to be resolved, up to eight separate capture sequences would be required to image one *cornice*. Each capture sequence is self-contained and at present this subjects the user to the cumbersome procedure of opening multiple RTI files to view the contents of a single *cornice*. A key hardware and software development would be, therefore, to build into the design the ability to automate stitching of multiple RTI sequences.[53]

Another important aspect of future work is the documentation of the full 3-dimensionality of the papyri fragments.[54] A great advantage of RTI is its ability to record colour and 3D shape information, but it does not provide full metric data. Such quantitative information is essential for analysing papyrus layer relationships, correcting surface/sign distortion, and facilitating virtual hypothesising and reconstruction of sections/scrolls. Just as *sovrapposti* and *sottoposti* are fragments no longer in their original positions, the position of fragments mounted on the *cartoncino* cannot be assumed to be correct.[55] It is here that 3D models of the fragments would

---

50   http://culturalheritageimaging.org/What_We_Offer/Downloads/View/ (accessed 25 May 2017)

51   For detailed research into the processes by which users/experts 'read' texts and insight into the supporting role of visua-lisation and other technology, see Terras (2006); see also Tarte (2014) and (2011); Janko (2016) is also relevant here.

52   E.g. Earl et al. (2011).

53   The feasibility of stitching has been confirmed by Tom Malzbender (pers. comm. January 2015), albeit in one dimension only.

54   Cf. Pal et al. (2013).

55   Capasso (1991); Essler (2006); see also Hendriks (2015).

contribute tremendously to virtual reconstruction. A future imaging system should therefore also encompass the acquisition of true 3D data through photogrammetry or a similar technique. Another key to virtual reconstruction is fragment joining. Edges have suffered loss from scroll opening and mounting and further attrition unavoidably occurs due to environmental instability and handling of these extremely friable and delicate objects. Edge refitting presents a particularly difficult challenge and while it can be tackled to some extent by comparing written content, there is potential for computer algorithms to aid general shape matching while simultaneously exploiting papyrus fibre patterns and distribution.[56] More research is also needed to determine what spectral analysis of the substrate, including the papyrus material and adhesives used to join papyrus sheets together (e.g. PHerc. 994, Cornice 09), and constrate (including pigments, binders, and possible drying agents or other ingredients, e.g. lead[57]), can offer. Likewise, it is important to establish the extent to which non-destructive optical imaging can penetrate the upper layer(s) and where interrogation of substrata must await further development of penetrative techniques.[58] Much stands to be gained with a customised UV-VIS-NIR lighting and filter system, and one than allows integrated capture with RTI[59] and photogrammetry. Further, among various digital tools for facilitating scholarly and conservation work is a way to measure the x-y-z coordinates of, and between, points. Functionality that also allows areas of papyrus to be segmented and repositioned virtually, ultimately for scroll reconstruction, will be important. Similar to annotation and view bookmarking tools in the RTIViewer (above), a method of 3D model annotation would also be needed to support the research workflow, for example, for hypothesising scroll reconstructions or content meaning and indicating levels of interpretive certainty.

Thanks to advances in digital photography during the 1990s, great strides have been made in research on the unrolled Herculaneum papyri. With further developments over the past decade and a half, we are poised to take another leap forward. As the results of this pilot imaging project demonstrate, the integration of spectral photography with the RTI capture technique, together with photogrammetry, presents great promise for furthering the work of Herculaneum papyrologists and facilitating greater access to the fragile remains of the only surviving library from the classical world.

---

56 The analytical potential of fibre patterning and distribution would be doubled if the backs of the uppermost layer could be recorded. Research we are currently undertaking at the UCL Centre for Digital Humanities ("Deep Imaging Mummy Cases" (https://www.ucl.ac.uk/dh/projects/deepimaging) led by Melissa Terras and Adam Gibson) is using spectral imaging on multiple layers of (uncarbonised) inscribed papyrus, and tests have shown that turning the papyrus so the upper pith fibres run parallel to the light direction reduces self-shadowing, making the underlying pith fibres, which run perpendicular to the light source, more visible..

57 Brun et al. (2016).

58 E.g. Mocella et al. (2015); Seales et al. (2011).

59 E.g. Hanneken (2016).

7. Reference List

Barbosa, João Carlos Garcia da Cunha (2011): Barbosa, João Carlos Garcia da Cunha, RTIBuilder version 2.0.2. [Software].

Barbosa, João Carlos Garcia da Cunha (2009): Barbosa, João Carlos Garcia da Cunha, RTI-Based Techniques and Tools for Digital Surrogates. MA dissertation. Universidade do Minho. http://repositorium.sdum.uminho.pt/handle/1822/11382 (accessed 04 June 2017)

Booras, Steven W. and Chabries, Douglas M. (2001): Booras, Steven W. and Chabries, Douglas M., "The Herculaneum Scrolls", Imaging Science and Technology Reporter 17(2), 1–4.

Booras, Steven W. and Seely, David R. (1999): Booras, Steven W. and Seely, David R., "Multispectral Imaging of the Herculaneum Papyri", Cronache Ercolanesi 29, 95–100.

Brun, Emmanuel; Cotte, Marine; Wright, Jonathan; Ruat, Marie; Tack, Pieter; Vincze, Laszlo; Ferrero, Claudio; Delattre, Daniel and Mocella, Vito (2016): Brun, Emmanuel; Cotte, Marine; Wright, Jonathan; Ruat, Marie; Tack, Pieter; Vincze, Laszlo; Ferrero, Claudio; Delattre, Daniel and Mocella, Vito, "Revealing Metallic Ink in Herculaneum Papyri", in: Proceedings of the National Academy of Sciences 113, 3751–3754. DOI: 10.1073/pnas.1519958113

Caine, Moshe and Magen, Michael (2011): Caine, Moshe and Magen, Michael, "Pixels and Parchment: The application of RTI and infrared imaging to the Dead Sea Scrolls", in: Stuart Dunn, Jonathan P. Bowen and Kia Ng (eds), Electronic Visualisation and the Arts (EVA 2011), 140–146. PDF: http://ewic.bcs.org/upload/pdf/ewic_ev11_s8paper2.pdf (accessed 04 June 2017)

Capasso, Mario (1991): Capasso, Mario, Manuale di Papirologia Ercolanese, Galatina.

Chabries, Douglas M.; Booras, Steven W. and Bearman, Gregory H. (2003): Chabries, Douglas M.; Booras, Steven W. and Bearman, Gregory H., "Imaging the Past: Recent applications of multispectral imaging technology to deciphering manuscripts", Antiquity 77(296), 359–372. DOI: 10.1017/S0003598X00092346

Cultural Heritage Imaging (2013): Cultural Heritage Imaging, Reflectance Transformation Imaging: Guide to highlight image capture v2.0. http://culturalheritageimaging.org/What_We_Offer/Downloads/RTI_Hlt_Capture_Guide_v2_0.pdf (accessed 04 June 2017)

Cultural Heritage Imaging (2011): Cultural Heritage Imaging, Reflectance Transformation Imaging: Guide to highlight image processing v1.4. http://culturalheritageimaging.org/What_We_Offer/Downloads/rtibuilder/RTI_hlt_Processing_Guide_v14_beta.pdf (accessed 04 June 2017)

Dyer, Joanne; Verri, Giovanni and Cupitt, John (2013): Dyer, Joanne; Verri, Giovanni and Cupitt, John, Multispectral Imaging in Reflectance and Photo-induced Luminescence Modes: A user manual (Version 1.0 October 2013). European CHARISMA Project. The British Museum, London.
http://www.britishmuseum.org/pdf/charisma-multispectral-imaging-manual-2013.pdf (accessed 04 June 2017)

Earl, Graeme; Basford, Philip J.; Bischoff, Alexander S.; Bowman, Alan; Crowther, Charles; Hodgson, Michael; Martinez, Kirk; Isaksen, Leif; Pagi, Hembo; Piquette, Kathryn E. and Kotoula, Eleni (2011): Earl, Graeme; Basford, Philip J.; Bischoff, Alexander S.; Bowman, Alan; Crowther, Charles; Hodgson, Michael; Martinez, Kirk; Isaksen, Leif; Pagi, Hembo; Piquette, Kathryn E. and Kotoula, Eleni, "Reflectance Transformation Imaging Systems for Ancient Documentary Artefacts", in: Jonathan P. Bowen; Stuart Dunn and Kia Ng (eds), EVA London 2011: Electronic Visualisation and the Arts: Proceedings of a conference held in London 6–8 July, London. http://ewic.bcs.org/upload/pdf/ewic_ev11_s8paper3.pdf (accessed 04 June 2017)

Essler, Holger (2006): Essler, Holger, "Papyri als Bilder und Bilder von Papyri", Cronache Ercolanesi 36, 103–143.

Gabov, Alexander (2010): Gabov, Alexander, IRR RTI: Using the Osiris, RTIiCAN's Blog – Research, Art Conservation & Archeology. http://rtiican.wordpress.com/2010/01/10/irr-rti-using-osiris/ (accessed 04 June 2017)

Gager, John G. (1992): Gager, John G. (ed.), Curse Tablets and Binding Spells from the Ancient World, Oxford.

Gautron, P.; Krivanek, J.; Pattanaik, S.; Bouatouch, K. (2004): Gautron, P.; Krivanek, J.; Pattanaik, S.; Bouatouch, K., "A Novel Hemispherical Basis for Accurate and Efficient Rendering", in: Dieter Fellner; Henrik Wann Jensen; Alexander Keller and Mark Ollila (eds), Rendering Techniques 2004: Eurographics Symposium on Rendering, June 21–23, 2004, Norrköping, Sweden, 321–330. http://cgg.mff.cuni.cz/~jaroslav/papers/egsr2004/gautron-egsr2004-hemispherical.pdf (accessed 04 June 2017)

Gigante, Marcello (2002): Gigante, Marcello, Philodemus in Italy: The books from Herculaneum (Body, in Theory: Histories of cultural materialism), translated by Dirk Obbink, Ann Arbor.

Hanneken, Todd (2016): Hanneken, Todd, "New Technology for Imaging Unreadable Manuscripts and Other Artifacts: Integrated Spectral Reflectance Transformation Imaging (Spectral RTI)", in: Claire Clivaz, Paul Dilley and David Hamidović (eds), in collaboration with Apolline Thromas, Ancient Worlds in Digital Culture Digital Biblical Studies. Leiden, 180–195. DOI: 10.1163/9789004325234_010

Happa, Jassim; Mudge, Mark; Debattista, Kurt; Artusi, Allesandro; Gonçalves, Alexandrino and Chalmers, Alan (2010): Happa, Jassim; Mudge, Mark; Debattista, Kurt; Artusi, Allesandro; Gonçalves, Alexandrino and Chalmers, Alan, "Illuminating the Past: State of the art", Virtual Reality 14(3), 155–182. DOI: 10.1007/s10055-010-0154-x

Hendriks, Sarah (2015): Hendriks, Sarah, "Digital Technologies and the Herculaneum Papyri", Digital Classicist London and Institute of Classical Studies Seminar 2015. https://www.youtube.com/watch?v=ao-2FwnPPhI (accessed 04 June 2017)

Janko, Richard (2016): Janko, Richard, "How to Read and Reconstruct a Herculaneum Papyrus", in: Barbara Crostini, Gunilla Iversen and Brian M. Jensen (eds), Ars Edendi Lecture Series, vol. IVI, 117–161. Stockholm. DOI: 10.16993/baj.f

Janko, Richard (2002): Janko, Richard, "The Herculaneum Library: Some recent developments", Estudios Clásicos 121, 25–41.

Karamanou, Ioanna (2016): Karamanou, Ioanna, "The Papyrus from the 'Musician's Tomb' in Daphne (ΜΠ 7449, 8517–8523)", Greek and Roman Musical Studies 4, 51–70. DOI: 10.1163/22129758-12341267

Kotoula, Eleni (2014): Kotoula, Eleni, Papyrus RTI Case Study. https://generic.wordpress.soton.ac.uk/archaeology/2014/05/15/papyrus-rti-case-study/ (accessed 01 October 2017)

Kotoula, Eleni (2013): Kotoula, Eleni, Transmitted RTI. http://generic.wordpress.soton.ac.uk/archaeology/2013/02/14/transmitted-rti/ (accessed 01 October 2017)

Kotoula, Eleni (2012): Kotoula, Eleni, Infrared RTI: Experimentation towards the development of multispectral RTI. https://generic.wordpress.soton.ac.uk/archaeology/2012/11/11/infrared-rti-experimentation-towards-the-development-of-multispectral-rti/ (accessed 01 October 2017)

Leach, Bridget and Tait, John (2000): Leach, Bridget and Tait, John, "Papyrus", in: Paul T. Nicholson and Ian Shaw (eds), Ancient Egyptian Materials and Technology, Cambridge, 227–253.

MacDonald, Lindsay William (2015): MacDonald, Lindsay William, Realistic Visualisation of Cultural Heritage Objects, Unpublished PhD thesis, University College London.

Macfarlane, Roger; Del Mastro, Gianluca; Booras, Steven and Antoni, Agathe (2007): Macfarlane, Roger; Del Mastro, Gianluca; Booras, Steven and Antoni, Agathe, "Update Report on the Use of the Multi-spectral Images of the Herculaneum Papyri", in: Jaakko Frösén, Tiina Purola and Erja Salmenkivi (eds), Proceedings of the XXIV International Congress of Papyrology, Helsinki 2007 (Commentationes Humanarum Litterarum 122), Helsinki, 579–586.

Malzbender, Tom; Gelb, Dan and Wolters, Hans (2001): Malzbender, Tom; Gelb, Dan and Wolters, Hans, "Polynomial Texture Maps", in: Eugene L. Fiume (ed.), Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2001), New York, 519–528. DOI: 10.1145/383259.383320

Mocella, Vito; Brun, Emmanuel; Ferrero, Claudio and Delattre, Daniel (2015): Mocella, Vito; Brun, Emmanuel; Ferrero, Claudio and Delattre, Daniel, "Revealing Letters in Rolled Herculaneum Papyri by X-ray Phase-contrast Imaging", Nature Communications 6, id 5895. DOI: 10.1038/ncomms6895

Mudge, Mark; Malzbender, Tom; Chalmers, Alan; Scopigno, Roberto; Davis, James; Wang, Oliver; Gunawardane, Prabath; Ashley, Michael; Doerr, Martin; Proenca, Alberto and Barbosa, João Carlos Garcia da Cunha (2008): Mudge, Mark; Malzbender, Tom; Chalmers, Alan; Scopigno, Roberto; Davis, James; Wang, Oliver; Gunawardane, Prabath; Ashley, Michael; Doerr, Martin; Proenca, Alberto and Barbosa, João Carlos Garcia da Cunha, "Image-Based Empirical Information Acquisition, Scientific Reliability, and Long-Term Digital Preservation for the Natural Sciences and Cultural Heritage", in: Maria Roussou and Jason Leigh (eds), Eurographics '08 Tutorials, Crete. DOI: 10.2312/egt.20081050

Mudge, Mark; Malzbender, Tom; Schroer, Carla and Lum, Marlin (2006): Mudge, Mark; Malzbender, Tom; Schroer, Carla and Lum, Marlin, "New Reflection Transformation Imaging Methods for Rock Art and Multiple-Viewpoint Display", in: Marinos Ioannides, David Arnold, Franco Niccolucci and Katerina Mania (eds), Proceedings of the 7th International Symposium

on Virtual Reality, Archaeology and Cultural Heritage (VAST 2006), Nicosia, 195–202. DOI: 10.2312/VAST/VAST06/195-202

Mudge, Mark; Schroer, Carla; Earl, Graeme; Martinez, Kirk; Pagi, Hembo; Toler-Franklin, Corey; Rusinkiewicz, Szymon; Palma, Gianpaolo; Wachowiak, Melvin; Ashley, Michael; Matthews, Neffra; Noble, Tommy and Dellepiane, Matteo (2010): Mudge, Mark; Schroer, Carla; Earl, Graeme; Martinez, Kirk; Pagi, Hembo; Toler-Franklin, Corey; Rusinkiewicz, Szymon; Palma, Gianpaolo; Wachowiak, Melvin; Ashley, Michael; Matthews, Neffra; Noble, Tommy and Dellepiane, Matteo, "Principles and Practices of Robust, Photography-based Digital Imaging Techniques for Museums", in: Alessandro Artusi, Morwena Joly-Parvex, Geneviève Lucet, Alejandro Ribes and Denis Pitzalis (eds), VAST 2010: The 11th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage: Short and project papers, Paris, 111–137.

Pal, Kazim; Terras, Melissa and Weyrich, Tim (2013): Pal, Kazim; Terras, Melissa and Weyrich, Tim, "Interactive Exploration and Flattening of Deformed Historical Documents", Computer Graphics Forum 32, 327–334. DOI: 10.1111/cgf.12052

Piquette, Kathryn E. (forthcoming): Piquette, Kathryn E., An Archaeology of Art and Writing: Early Egyptian labels in context, Cologne.

Piquette, Kathryn E. (2016): Piquette, Kathryn E., "Documenting Early Egyptian Imagery: Analysing past technologies and materialities with the aid of Reflectance Transformation Imaging (RTI)", in: Gwenola Graff, Alejandro Jiménez-Serrano and Maxence Bailly (eds), Préhistoires de l'écriture: iconographie, pratiques graphiques et émergence de l'écrit dans l'Egypte prédynastique/Prehistories of writing: Iconography, graphic practices and the forming process of writing in Predynastic Egypt. Actes de la table-ronde de décembre 2010, Aix-en-Provence, 89–112. https://www.academia.edu/22182476/Documenting_Early_Egyptian_Imagery_Analysing_ past_technologies_and_materialities_with_the_aid_of_Reflectance_Transformation_ Imaging_RTI (accessed on 4 June 2017)

Piquette, Kathryn E. (2014): Piquette, Kathryn E., "Scribal Practice and an Early Dynastic Stone Vessel Inscription: Material and aesthetic implications", in: Aidan Dodson, John J. Johnston and Wendy Monkhouse (eds), A Good Scribe and an Exceedingly Wise Man: Studies in honour of W. J. Tait, London, 241–250. https://www.academia.edu/4770946/Scribal_Practice_and_an_ Early_Dynastic_Stone_Vessel_Inscription_Material_and_aesthetic_implications (accessed on 4 June 2017)

Piquette, Kathryn E.; Dahl, Jacob L. and Green, Jack E. M. (2011): Piquette, Kathryn E.; Dahl, Jacob L. and Green, Jack E. M., Exploring Ancient Writings at the Ashmolean Museum with Advanced Digital Technologies. http://www.ashmolean.eu/departments/antiquities/research/ research/rtisad/ (accessed on 4 June 2017)

Seales, William Brent; Griffioen, James; Baumann, Ryan and Field, Matthew (2011): Seales, William Brent; Griffioen, James; Baumann, Ryan and Field, Matthew, "Analysis of Herculaneum Papyri with X-Ray Computed Tomography", in: Proceedings of the 10th International Conference on Non-destructive Investigations and Microanalysis for the Diagnostics and Conservation of Cultural and Environmental Heritage, Brescia. http://www. ndt.net/article/art2011/papers/FIELD%20-%20M%2014.pdf (accessed 29 October 2017)

Seales, William Brent; Parker, Clifford Seth; Segal, Michael; Tov, Emanuel; Shor, Pnina and Porath, Yosef (2016): Seales, William Brent; Parker, Clifford Seth; Segal, Michael; Tov, Emanuel; Shor, Pnina and Porath, Yosef, "From Damage to Discovery via Virtual Unwrapping: Reading the Scroll from En-Gedi", Science Advances 2(9), e1601247. DOI: 10.1126/sciadv.1601247

Stökl ben Ezra, Daniel (2014): Stökl ben Ezra, Daniel, "Interdisciplinary Perspectives from Material and Computer Sciences", Manuscript Cultures 7, 92–103.

Tack, Pieter; Cotte, Marine; Bauters, Stephen; Brun, Emmanuel; Banerjee, Dipanjan; Bras, Wim; Ferrero, Claudio; Delattre, Daniel; Mocella, Vito and Vincze, Laszlo (2016): Tack, Pieter; Cotte, Marine; Bauters, Stephen; Brun, Emmanuel; Banerjee, Dipanjan; Bras, Wim; Ferrero, Claudio; Delattre, Daniel; Mocella, Vito and Vincze, Laszlo, "Tracking Ink Composition on Herculaneum Papyrus Scrolls: Quantification and speciation of lead by X-ray based techniques and Monte Carlo simulations", Scientific Reports 6, id 20763. DOI: 10.1038/srep20763

Tarte, Ségolène M. (2014): Tarte, Ségolène M., "Interpreting Textual Artefacts: Cognitive insights into expert practices", in: Clare Mills, Michael Pidd and Esther Ward (eds), Proceedings of the Digital Humanities Congress 2012: Studies in the Digital Humanities, Sheffield. http://www.hrionline.ac.uk/openbook/chapter/dhc2012-tarte (accessed on 4 June 2017)

Tarte, Ségolène M. (2011): Tarte, Ségolène M., "Papyrological Investigations: Transferring perception and interpretation into the digital world", Literary and Linguistic Computing 26 (2), 233–247.

Terras, Melissa (2006): Terras, Melissa, Image to Interpretation: An intelligent system to aid historians in reading the Vindolanda texts, Oxford, 22–59. DOI: 10.1093/acprof:oso/9780199204557.001.0001

Tov, Emanuel (2015): Tov, Emanuel, Textual Criticism of the Hebrew Bible, Qumran, Septuagint: Collected essays, vol. 3 (Supplements to Vetus Testamentum 167), Leiden. DOI: 10.1163/9789004285569_020

Twycross, Meg (2008): Twycross, Meg, "Virtual Restoration and Manuscript Archaeology", in: Mark Greengrass and Loran Hughes (eds), The Virtual Representation of the Past (AHRC ICT Methods Network: Digital Research in the Arts and Humanities), Aldershot, 23–48.

Ware, Gene A.; Chabries, Douglas M.; Christiansen, Richard W. and Martin, Curtis E. (2002): Ware, Gene A.; Chabries, Douglas M.; Christiansen, Richard W. and Martin, Curtis E., "Multispectral Document Enhancement: Ancient carbonized scrolls", Geoscience and Remote Sensing Symposium (IGARSS), 2000. Proceedings. IEEE 2000 International, vol. 6, 2486–2488. DOI: 10.1109/IGARSS.2000.859615

Webb, E. Keats and Wachowiak, Melvin (2011): Webb, E. Keats and Wachowiak, Melvin, "QuadriFlash Mantis for Reflectance Transformation Imaging (RTI)", Imaging Studio Technical Notes. Smithsonian Museum Conservation Institute. http://si.edu/content/MCIImagingStudio/papers/QuadriFlashMantis_RTI_Technical_Note.pdf (accessed on 4 June 2017)

Author[60]

Dr. Kathryn E. Piquette
UCL Centre for Digital Humanities
University College London
Gower Street, London
WC1E 6BT, UK EU

Email: k.piquette@ucl.ac.uk

Mapping the Words
Experimental visualizations of translation structures between
Ancient Greek and Classical Arabic

Torsten Roeder

Abstract: The article deals with presentation forms of linguistic transformation processes from ancient Greek sources that were translated into classical Arabic from the 9th to 11th century AD. Various examples demonstrate how visualizations support the interpretation of corpus structures, lexical differentiation, grammatical transformation and translation processes for single lexemes in the database project Glossarium Graeco-Arabicum. The database contains about 100,000 manually collected word pairs (still growing) from 76 texts and their translations. The article discusses how the project utilizes Sankey diagrams, tree maps, balloon charts, data grids and classical coordinate systems to point out specific aspects of the data. Visualizations not only help beginners to understand the corpus structure, they also help editors and specialized users to identify specific phenomena. A well-documented interface design is crucial both for usability and interpretative work.[1]

1. Introduction

1.1. Translations as Research Field in the Digital Humanities

Analyses of languages and their structures belong to the fields that have applied computational methods to a humanities-related subject in a very early stage. An open question may stay here whether this was due to the abstract structures of information technology and linguistics, which were particularly easy to combine, or to the large amounts of data that suggested the use of computers, or even to further factors that may have been involved.[2] Eventually it has turned out as a major feature of computational linguistics that digital recording of language-related data would not need to confine itself to the creation of an electronic equivalent of a printed reference book. To the contrary, its great potential lies in the possibility of complex quantitative and structural analyses, which could unveil leads and clues to structures, processes and developments in language.[3]

While digital resources that deal with just one language are available in numerous and diverse forms, bilingual resources still exist only in niches. Yet these possess great importance for investigations of cultural transfer: Assumed that, in any particular cultural context, an interest existed to express or to understand a certain message in another language, translation belongs

---

1 This is a translated, reworked and slightly expanded version of an article which was originally published in the German digital humanities journal "Zeitschrift für digitale Geisteswissenschaften"; see Roeder (2016). Further, the topic has been subject of a lecture in the Digital Classicists Berlin Seminar; see Arzhanov/Roeder (2016).

2 For a thorough comparison of the disciplines of corpus linguistics, literary computing, and computational linguistics cf. Lüdeling/Zeldes (2009).

3 E.g. the Regionalsprache.de (REDE) project (http://www.regionalsprache.de), which uses advanced digital methods to investigate German dialects.

to the most significant facets of intercultural exchange. Not only does the translation process transport information, but it also partly transfers also lexical expressions and structural elements into the other language. In this way, translations provide impetus for cultural development, which is reflected in vocabulary and grammar. Systematically designed bilingual resources are therefore of major relevance for the understanding of cultural transfer. For this reason, the resources that come into question for investigating historical transfer studies are, amongst others, written translations.

Further, such transfer processes normally work only in one direction, i.e. they are not reciprocal in their effect: This would only be the case when an equivalent transfer process also occurred in the other direction, which is usually not applicable in the case of translations (or at least it does not happen in the same area of interchange). Thus, it is often sufficient to consider only one direction of transmission to examine a cultural trend.

Another more special case are receptions of earlier cultures, such as the reception of Ancient Greek literature in the Renaissance or in the medieval Middle East, which took place only unilaterally. In the following, the Glossarium Graeco-Arabicum, a project that deals with Arabic translations of ancient Greek writings, will work as an example of such receptive cultural transfer processes. In addition, a look at a follow-up project will be given, where the special role of early Syriac translations as a bridging language from Greek to Arabic will be the focus of investigation.

1.2.   The Glossarium Graeco-Arabicum

Figure 1: Aristotle teaching, from Kitāb naʿt al-hayawān, 7th century AH (13th century AD).[4]

---

4   British Library, MS Or. 2784, fol. 96r, Reproduced and cited from: Nasr (1976), 50, plate 22. Any copyrights are to be observed.

The Glossarium Graeco-Arabicum (short: GlossGA) undertakes the lexicographical analysis and description of a corpus that is in many aspects of particular interest for studies of cultural transfer. During a flowering period of Arabic translation movements, which lasted approximately from the 3rd to the 5th century AH (9th to 11th century AD), an intensive Arabic reception of classical Greek writings took place in the area of Baghdad.[5] The writings of Aristotle, Plato, Euclid, Galen and many others were translated, summarized and paraphrased into Arabic, mostly by scholars with Christian backgrounds, and then utilized for the study of philosophy, mathematics, medicine, astronomy, and other sciences (cf. figure 1). It is thus a historic transfer process of ancient Greek literature into classical Arabic culture.

Language	Greek	Arabic
Lexeme	λόγος	قول
Root / Stem		q w l – – / –
Part of Speech	noun	noun
Expression	οἱ Σωκρατικοὶ λόγοι	al-aqāwīlu l-mansūbatu ilā Suqrāṭa
Annotation	–	paraphr.
Quotation	οὐδὲν γὰρ ἂν ἔχοιμεν ὀνομάσαι κοινὸν τοὺς Σώφρονος καὶ Ξενάρχου μίμους καὶ τοὺς Σωκρατικοὺς λόγους	wa-ḏālika annahū laysa lanā an nusammiya bi-māḏā yušāriku ḥikāyātu wa-tašbīhātu l-šāʿiri Sūfruna wa-Kasānarḫusa wa-l-aqāwīlu l-mansūbatu ilā Suqrāṭa
Reference	Arist. Poet. 1, 1447b11	220.20
Lexica	❓ Search for translation in Perseus	❓ Search for translation in Perseus

Figure 2: Glossarium Graeco-Arabicum, Word record for λόγος / قول.[6]

The corpus of the Glossarium Graeco-Arabicum[7] is based on 76 original texts and their translations. It includes just over 100,000 manually selected word entries so far. One single entry consists of a word from the Greek original text, the corresponding word (or phrase) in the Arabic translation and additional grammatical properties (base lexeme, part of speech and root). Furthermore, a complete entry (cf. figure 2) contains accurate records of the original quotations. Additionally, the database provides contextual links and data connections to other digital resources in the field of the Graeco-Arabica, such as Perseus Digital Library[8], G2A Web Application for Literary Computing[9] and Digital Corpus for Graeco-Arabic Studies[10]. Since early 2016, the texts are also associated with the names of their authors and (if known) their translators, which are, in most cases, connected to library standard catalog entries (VIAF, GND) and to English Wikipedia pages.

---

5 Cf. D'Ancona (2013).

6 https://telotadev.bbaw.de/glossga/glossary.php?id=131860 (screenshot: January 17, 2016).

7 Glossarium Graeco-Arabicum, European Research Council; Ruhr-Universität Bochum; Berlin-Brandenburgische Akademie der Wissenschaften, http://telota.bbaw.de/glossga (March 5, 2017); the current development version can be found at https://telotadev.bbaw.de/glossga (March 5, 2017). Cf. also Endress et al. (2013); Arzhanov/Roeder (2013).

8 http://www.perseus.tufts.edu/hopper (March 5, 2017).

9 http://g2a.ilc.cnr.it (March 5, 2017).

10 http://www.graeco-arabic-studies.org (March 5, 2017).

Figure 3: File cards used to compile GALex and Glossarium Graeco-Arabicum. Photo: private.

It was the original aim of the project to create the printed lexicon "A Greek and Arabic Lexicon" (short: GALex).[11] Starting in the 1980s, the team began to collect data on handwritten file cards (cf. figure 3), and later, in the course of the "digital turn", scanned them into TIFF images and began to transcribe them manually[12] into a relational database.[13] This database has since been titled Glossarium Graeco-Arabicum.[14] The overcoming of the manifold technical challenges that came with the parallel use of Greek and Arabic characters with diacritics and vowel signs respectively, involving issues from presentation to search indexing, was an essential step towards a stable basis for scientific work with modern usability.[15] Thanks to appropriate funding[16] and infrastructural support, it was possible to redesign gradually the database from a file card management system into a research instrument: It has evolved from a relatively simple and, initially, only internally used database application to a freely accessible digital encyclopedia which now exists parallel to the printed GALex. Indeed, print and digital versions serve different purposes: While the GALex volumes provide consistent entries, currently from letters Alif to Bā', the database covers the full lexical range, while the collecting and editing of data is an open process. Despite its work-in-progress nature, the database is already used regularly for research and frequented by about 30–40 recurrent international users per month.[17]

---

11   Cf. Endress/Gutas (1992).

12   OCR (Optical Character Recognition) and ICR (Intelligent Character Recognition) were not options, as all file cards were filled in manually, with many individual handwriting styles.

13   Cf. Arnzen et al. (2012).

14   The Glossarium Graeco-Arabicum is hosted by the Berlin-Brandenburg Academy of Sciences since 2008.

15   Cf. Roeder (2015).

16   Most recently (2010–2015) under the ERC project "Greek into Arabic. Philosophical and Linguistic Bridges" (Advanced Grant 249431, http://greekintoarabic.eu). The results were presented, amongst other venues, at the international workshop "Plotinus East and West. The Enneads in Arabic and Latin" (Pisa, November 3–6, 2014) by Yury Arzhanov, Gerhard Endreß and Torsten Roeder.

17   As analyzed by Piwik (http://piwik.org), usage data from August 2014 to December 2015.

1.3. Towards Comparative Translation Studies

The availability and the further development of the database's content and functionality is currently guaranteed for the medium term at least, mainly due to a new project titled "Transmission of Classical Scientific and Philosophical Literature from Greek into Syriac and Arabic" which was launched in mid-2016.[18] The planned database "HunaynNet" (named after Hunayn ibn Ishaq, one of the most famous Christian-Arabic translators) will also take earlier Syriac translations into account (cf. figure 4). That those Syriac translations influenced later Arabic versions is a general assumption that still needs to be confirmed. To investigate the question of the role of Syriac in the transmission process and to tackle the problem of continuity from the period of Syrian to Arabic translations, HunaynNet will integrate all records from the Glossarium Graeco-Arabicum to form a trilingual research platform for translation studies. One declared aim of HunaynNet is to reveal how the Syriac translations, which were used primarily by Syriac Christians, might have influenced or prepared the prosperity of the Islamic sciences. The project intends to design the underlying database flexibly enough to incorporate further relevant languages, e.g. Latin, for future projects.

Language	Greek	Syriac	Arabic
Lexeme	λόγος	ܡܠܬܐ	قول
Root / Stem		m l l – – / –	q w l – – / –
Part of Speech	noun	noun	noun
Expression	ὁ λόγος κατὰ τοῦ ὑποκειμένου	ܡܠܬܐ ܕܥܠ ܡܢ ܕܣܝܡ ܬܚܝܬ	القول يحمل على الموضوع
Translation	definition	definition	definition
Annotation	—	—	—
Quotation	ὥστε καὶ τοὔνομα καὶ ὁ λόγος κατὰ τοῦ ὑποκειμένου κατηγορηθήσεται	ܟܕ ܐܦ ܫܡܐ ܘܡܠܬܐ ܕܥܠ ܡܢ ܕܣܝܡ ܬܚܝܬ ܢܬܩܛܪܓ	فيكون الاسم والقول يحملان على الموضوع
Reference	Arist. Cat. 2a26	ed. King, 100.35	BN 160a16
Lexica	Search in Perseus		Search in Perseus

Figure 4: A preliminary example of a tri-lingual entry in the future database HunaynNet, displaying Syriac and Arabic translations for λόγος.

---

[18] The project is directed by Dr. Grigory Kessel, hosted at the Austrian Academy of Sciences (ÖAW) and funded for five years as an ERC Starting Grant, cf. ÖAW, "Vier neue ERC-Starting Grants", November 24, 2015 http://www.oeaw.ac.at/oesterreichische-akademie-der-wissenschaften/die-oeaw/article/vier-nachwuchswissenschaftlerinnen-an-der-oeaw-erhalten-erc-starting-grants-copy-1/ (April 29, 2016).

1.4. Visualization and Interpretation of Big Data

As a collective of pioneer disciplines, the Digital Humanities seek technologically supported alternatives to traditional philological research methods, in order to gain evidence of previously unrecognized or unconsidered phenomena by using digital techniques.[19] This includes also exploratory approaches, e.g. such as "serendipity"[20], which are less appreciated by the humanities, since they are not guided by any predefined, specific interest. Indeed, quantitative analyses have to accept the criticism that their primary results consist of purely numerical statements. However, the oceanic quantities of digitally available information often does not suit hermeneutic approaches, so that alternative methods need to be evaluated, even if eventual insights are impossible to foresee.

Also in the Glossarium Graeco-Arabicum, the steadily increasing amount of data makes it increasingly difficult to gain an overall view of the underlying corpus, or even of excerpts. Although typical database mechanisms, providing search forms and result lists, cover basic requirements, this type of functionality suffices only in rare, special cases: When searching over a corpus of more than 100,000 records, a general search query would in most cases lead to a vast amount of results, unless the user is able to narrow down the results. Thus, the conventional functionality will no longer suffice in the long term, especially while the database is still growing: With extending size, the content of a database becomes more and more opaque for the user, and it becomes increasingly unclear how to evaluate individual information within the context of the whole corpus.

Therefore, in order to provide a solid foundation for research, it is necessary to present the material in alternative ways. Particularly, visually oriented interfaces offer a great opportunity in this respect, as they are able to depict a variety of semantic levels simultaneously, and they offer an attractive method of access for academics or scholars as well as the interested public. A good visualization can give immediate insights to the user into the relationship between data structures and content, and may even give decisive clues to new interpretive possibilities. On this basis, it is possible to generate new impulses for research in the humanities, even from large amounts of data.

---

19   Cf. Oakes/Ji (2012).

20   Cf. Thudt et al. (2012).

2. Method

2.1.   »Mapping« the Words?[21]

**Figure 5: Astronomers working, from Shāhanshāhī-nāmah,
10th century AH (16th century AD).**[21]

The illustration (cf. figure 5) shows an astronomical laboratory in Istanbul in the 16th century. Scholars are handling all sorts of scientific tools, such as quadrant, astrolabe, hourglasses and celestial globe, and transmitting their results on maps and in journals. With their devices, they try to measure and to depict the happenings in the sky, in order to better understand and interpret the movements of the heavenly bodies. This variety of monitoring tools and the detailed recording of the results act here as crucial prerequisites for understanding complex processes. Knowledge production, as pictured here, is thus a process of measuring, mapping and interpretation. Mapping assumes a special role in this process, since it mediates between data collection and measurement on the one side and interpretation and understanding on the other side.

The Glossarium Graeco-Arabicum aims primarily at translation studies, as introduced before, and consists mainly of a large amount of Greek-Arabic word pairs, each associated with their

---

21   University Library Istanbul, MS No. FY 1404. Reproduced and cited from: Nasr (1976), 113, plate 65 (detail). Any copyrights are to be observed.

sources and translators. Geographically, the Arabic side of the corpus concentrates on the area of Baghdad, because in this region the richest production of translations took place. In this respect, the title "Mapping the Words" does not aim at geographical, or even – as the presented image could suggest – at astronomical dimensions, but rather on linguistic structures and their alterations through translation processes. This is the point where visualizations come into play: Graphical illustrations should allow the viewer to recognize (maybe even to "read") structures and information like on a geographical map.[22] Then, in the best case, such maps could form an independent reference for research.

However, some crucial questions remain. In which way should visualizations be designed to support research questions in an actually useful way? Which methods of multidimensional representation are available? In which state does a visualization supply more information, or display it more clearly, than a textual results list of a database query? Do visualizations have the potential to provide a reference for research at all, or are they merely a simplistic view?

Visualizations are not per se useful or useless, but much depends on the quality of its design. Edward Tufte, in his monograph "The Visual Display of Quantitative Information", formulated some guidelines for the design of visualizations under the heading "Principles of Graphical Excellence"[23] and pointed out the following key aspects:

- The informative value of visualization increases with the substantial dimensions displayed and correlated at the same time.
- A good visualization provides information compactly without loss of transparency and diversity.

According to this, formal and structural design is actually a central aspect of visualization. When thinking of the users of a visualization, it seems crucial to explain not only the material, but also the method of presentation. From this perspective, graphical style is less relevant when utilized for visual effects, as well as for suggestive design, which puts only prefabricated statements in the foreground. It is, instead, crucial to raise the viewers' interest by simultaneously displaying possibly correlated information and by doing this transparently, to leave interpretation ultimately to them. This reflects the idea of mapping as demonstrated in the above picture.

The following chapters will exemplify this idea in four use cases from the Glossarium Graeco-Arabicum. As a general strategy, the following aspects need to be clarified before implementing a visualization:

- which specific aspect to investigate into, and which research interest to follow,
- which dimensions are necessary for this, and which correlations are expected in this context,
- which visualization strategies are adequate for this use case, and how the display can be optimized for the viewer.

2.2. Defining Topics and Interests

The development of visualizations for the Glossarium Graeco-Arabicum should support the following aspects of the database:

1. How can an overview of the whole corpus be created, in order to understand the chronological arrangement of the texts and their respective relevance? Here, the scope of the texts needs to be correlated with the history of sources and their translations. This is

---

22 This was demonstrated e.g. for a historic speech corpus of English; cf. Alexander (2010).

23 Cf. Tufte (2001), 51.

of particular importance for new users, who are not yet familiar with the database corpus (cf. section 3.1: Corpus Overview).

2. Which lexical differences exist between sources and their translations in general? Do translations use a broader vocabulary and tend to circumscribe phrases with more words, or do they tend to synthesize information? Here it would be necessary to compare the lexical variance of source and target language (cf. section 3.2: Terminological Consistency).

3. How did the texts transform on the grammatical level? How did the translators react to the structural differences between Greek and Arabic? Which differences exist among the translators? Which Greek parts of speech transformed into which Arabic ones, and which differences become visible by comparing authors or sources to one other (cf. section 3.3: Part of Speech Transformations)?

4. How can the origins and translation variants for a specific lemma, including the relevance of authors and translators in the process, be displayed at a glance? The aim would be to show how the various translators treated a specific word in correlation to its usage by different authors or in different sources (cf. section 3.4: Translation Variants).

2.3. Choosing Visualization Techniques

After having defined the topics and the specific research interests, it is the right time to decide for a data visualization. Which visualization techniques are actually adequate to map the above-mentioned aspects of the database? Obviously, typical two-dimensional diagrams, e.g. bar charts or pie charts, will not suffice to depict the mentioned correlations without compromising the subject's complexity. Therefore, the following approaches favor visualizations that depict more than two parameters simultaneously.

The implementations presented in the next chapter will apply three basic models of visualization:

- Coordinate systems, which represent information usually as two-dimensional points that are further qualifiable by color, size and shape, resulting in five simultaneous dimensions (or even more, if e.g. different patterns or line widths would also be applied).
- Tree maps, which represent information as areas of quantity-related size, arranged in a way that reflect hierarchic relationships.[24] Colors, forms and patterns are suitable to display further dimensions.
- Flowcharts, which illustrate multistage processes, and amongst these especially Sankey diagrams[25], which depict quantitative connections between nodes.

From the variety of visualization approaches shown here, it becomes evident again that the decision for a suitable visualization is already determined by the data itself and by the intention of which aspect to investigate. It also becomes clear that visualizations not only reflect quantities, but also qualities like typologies, relationships, and processes.

---

24 For a historical overview and an introduction to the various algorithms, cf. Shneiderman/Plaisant (2009) and Leydesdorff/Welbers (2011).

25 Named after the Irish engineer Matthew Henry Phineas Riall Sankey (1853–1925), who developed a graphical representation of simultaneous data flows that displays direction and quantity at the same time. A collection of application examples can be found on the website "Sankey Diagrams. A Sankey diagram says more than 1000 pie charts", http://www.sankey-diagrams.com (April 29, 2016).

3. Implementation

On the technical level, all following implementations rely on the API *Google Chart Tools*[26], which provides a highly reliable, sufficiently flexible and no-cost way to create visualizations on a web site.

3.1. Corpus Overview

The researcher looks at the recently acquired encyclopedia in the bookshelf. Intuitively, he recognizes the scope of this publication in comparison to other books on the shelf. He takes one of the volumes and browses through the pages, receiving a quick insight into the content structure. Most databases do not provide an equivalent possibility: In many cases, the information about scope and structure of a database corpus remains hidden, and overviews for the users are rarely implemented. However, especially in large corpora, this entrance procedure is of great importance, as it can guide the user's expectations into the right direction before research takes place. For the diversified corpus of the Glossarium Graeco-Arabicum a tree map was implemented to provide exactly such a preliminary overview, to show to the user what he can actually expect to find in the database.
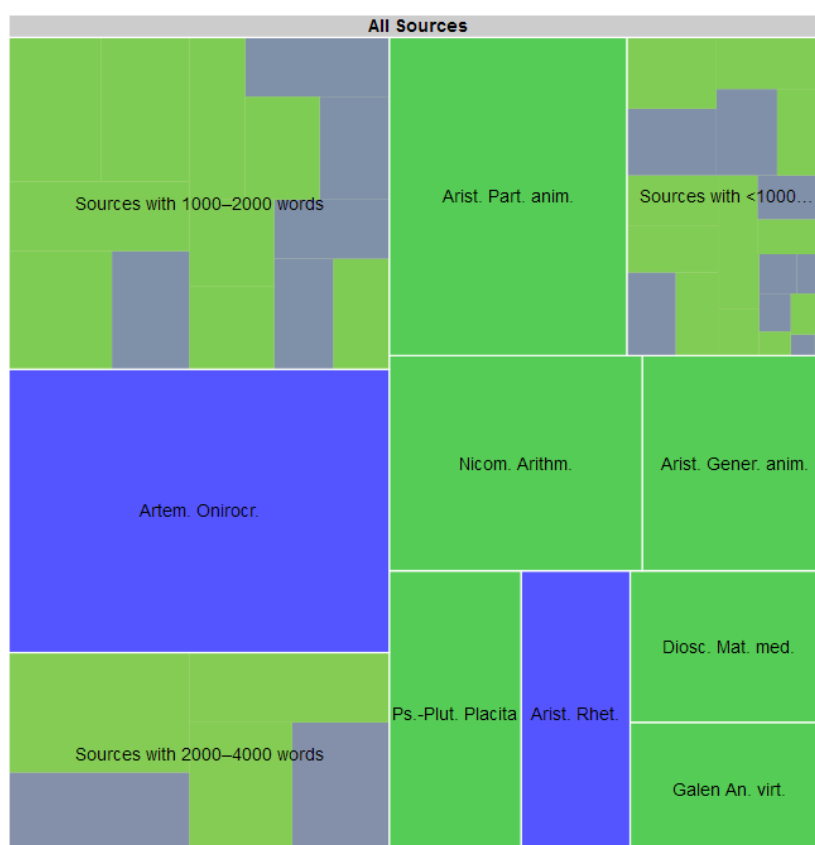
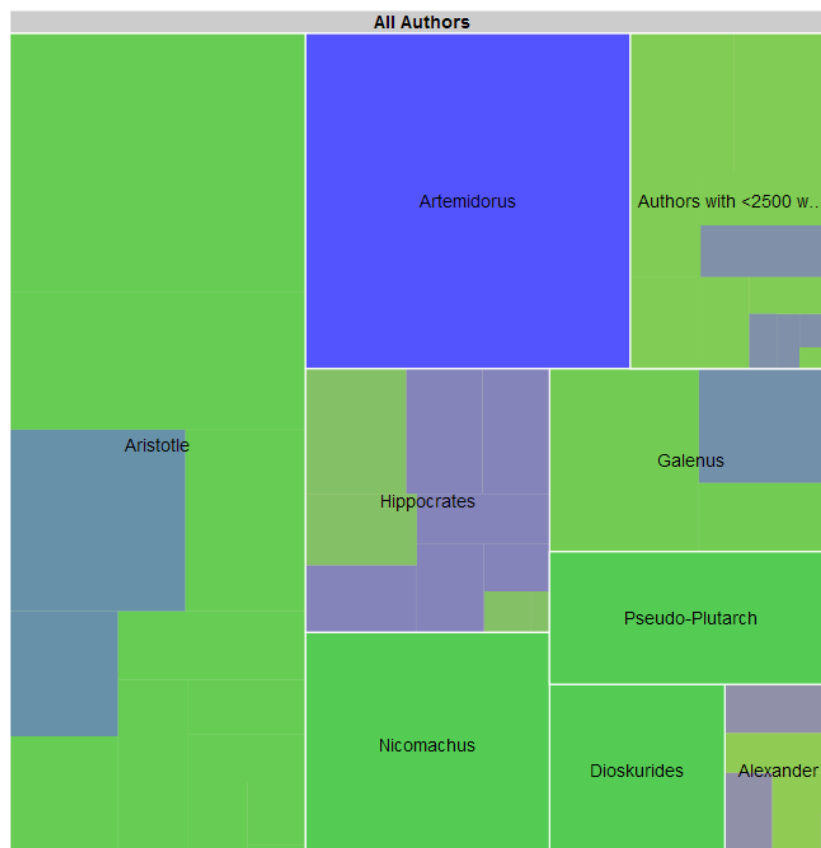**Figure 6: Glossarium Graeco-Arabicum, Corpus Treemap by Source.**[27]

---

26 Google Chart Tools, version 44, February 23, 2016 https://developers.google.com/chart/interactive/docs/release_notes#-february-23-2016 (April 29, 2016).

27 https://telotadev.bbaw.de/glossga/source_charts.php?chart_type=corpus-treemap (screenshot: January 16, 2016).

The tree map (cf. figure 6) is a display of all 76 source texts of the corpus at the same time. Each area in this visualization represents one source, the size being proportional to the number of recorded words. For a better overview, the sources were divided into three groups of size (2,000–4,000 words, 1,000–2,000 words and less than 1,000 words). It is recognizable that about one-half of the corpus is determined by eight sources with a very high number of records. It should be noted here that the number of recorded words does not necessarily correspond to the actual length of a source's full text, because some sources are only represented by a representative selection of words.

In addition to the size, a color code provides information on the translator: green indicates a source with a known translator, while blue represents an unknown translator. The tree map demonstrates that the vast majority of translators is known, especially of the more extensively recorded sources; overall, the translators are known in 46 of 76 cases (about 60%).
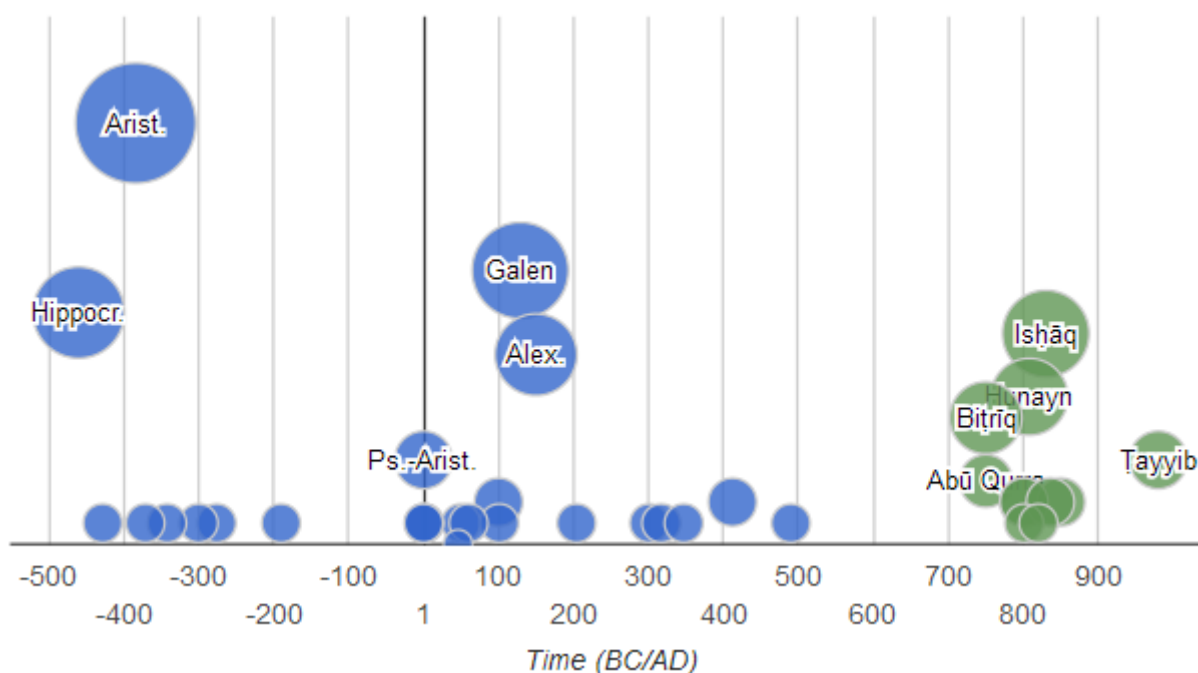

Figure 7: Glossarium Graeco-Arabicum,Corpus Treemap by Author.[28]

The alternative arrangement of the tree map by authors (cf. figure 7) basically gives a similar impression, but the information is focused on the relevance of the authors. Similar to the previous example, the image is determined by a few large areas, which is even more extreme in this case. However, it also becomes clear that e.g. Aristotle and Galen are represented by a large variety of texts, while Artemidorus, the author of the most comprehensive text, is represented only by that one single source. In addition, it becomes apparent that some translators, mainly those of Artemidorus and Hippocrates, are not yet identified. By clicking on the area, all texts related to that author are displayed as a tree map again (like a zoom-in on the corpus tree map).

Figure 8: Glossarium Graeco-Arabicum, Author and Translator Time-map.[29]

For a chronological perspective that also takes the relevance of the respective author or translator for the corpus into account, a "Balloon Chart" was designed (cf. figure 8).[30] Each circle on the underlying timeline represents one person, while size and vertical position are proportional to the number of corresponding sources. This produces a clustering effect in the most important periods, while the more relevant authors and translators stand significantly out through the balloon effect, supporting legibility.

3.2. Terminological Consistency

When comparing texts and their translations, it is natural that some translators tend to differentiate the original terminology, while others make more use of generalizations. While a translation ideally would maintain terminological consistency as well as possible, a special interest of the translator, possibly related to the personal background, could determine a different

---

29   https://telotadev.bbaw.de/glossga/author_charts.php?chart_type=timeline (screenshot: January 16, 2016).

30   This visualization type is related to the so-called poppy chart, implemented by Valentina D'Efilippo and Nicolas Pigelet on http://www.poppyfield.org (April 29, 2016).

terminological rendition in the translation. The question at hand is whether such a tendency existed when the Greek texts were translated into Arabic, and to what extent this depended on translators or even on the authors. A first approach to answering this question is comparing the numbers of distinct words (counting each word only once, regardless of how often it is used) in each source and its respective translation. A comparison of these numbers could indicate to which extent the lexical diversity of the original text was modulated by the translation.
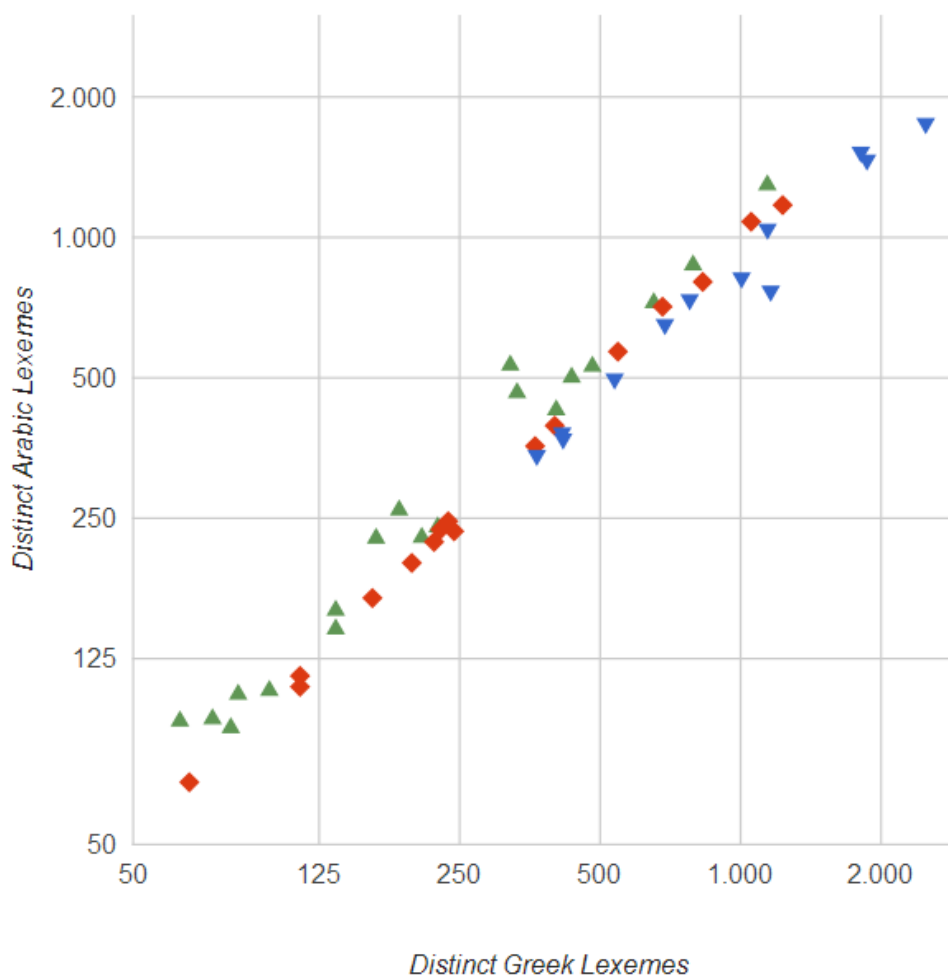
Figure 9: Glossarium Graeco-Arabicum, Distinct Lexemes by Source.[31]

This diagram (cf. figure 9) will help to compare this phenomenon throughout the whole corpus. Each shape corresponds to one text. The x-axis represents the number of distinct Greek lexemes, while the y-axis corresponds to the same in Arabic. Due to the numerical distribution, a logarithmic scale is applied. An exactly equal distribution of distinct lexemes (inflections ignored) in the Greek source and the Arabic translation would appear on the center diagonal; a trend to generalization in the Arabic text would result in a position toward the lower right, while the opposite case, a trend to diversity, would result in a position toward the upper left. In addition, a shape/color coding is used: A variation of less than 5% is highlighted as a red diamond (◆), while higher deviations are marked as a green triangle pointing up (▲) for more diversity in Arabic, or a blue triangle pointing down (▼) for the opposite case.

First, it becomes clear that the majority of translations tend towards a more diversified terminology. Yet, this trend is visible mostly in the lower half of the diagram (less than 500 words),

---

31   https://telotadev.bbaw.de/glossga/source_charts.php?chart_type=-distinct-lexemes (screenshot: January 16, 2016).

while it is balanced in the middle range (500–1,000 words) and reverted in the upper region (more than 1,000 words). However, this does not imply that texts with more recorded words are terminologically less diverse: In larger texts, a disproportional number of overlaps levels the clear tendency towards differentiation in smaller texts, and thus creates a statistical counter effect.
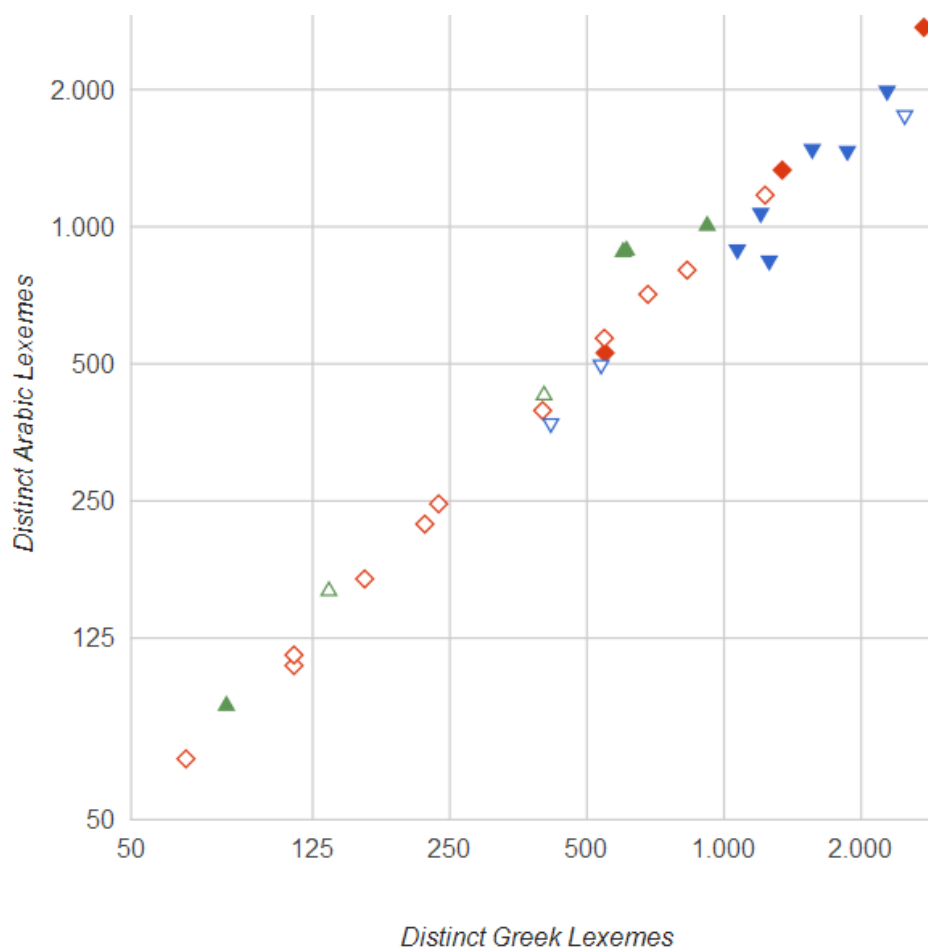
Figure 10: Glossarium Graeco-Arabicum, Distinct Lexemes by Translator.[32]

A diagram variant which displays the same setting, now focused on translators (cf. figure 10), demonstrates that this is not only a random spike in the statistics, but also confirms the latter assumption. As all texts by one author are merged into a larger group, resulting in some very large sets of words, a clear tendency towards generalization is visible; and this occurs despite a contrary tendency in the single texts.

A real statistical glitch is instead caused by the texts *De virtutibus et vitiis and Divisiones quae vulgo dicuntur Aristoteleae*, which both display a very high lexical diversity in Arabic. This can be explained by the fact that this text was translated two times, namely by Theodore Abū Qurra (ca. 750–ca. 823) and later by Ibn al-Tayyib (980–1043).

Overall, these observations give some important insights about how to interpret a statistical statement on terminological consistency: It is relative to the total amount of compared words, while the absolute ratio does not serve as an indicator. However, future experiments should take this into consideration to normalize the statistical behavior. Additionally, an abnormally high divergence from the average points very probably towards special conditions that are unrelated to a single author or translator.

---

32 https://telotadev.bbaw.de/glossga/author_charts.php?chart_type=distinct-lexemes (screenshot: January 16, 2016).

3.3. Part of Speech Transformations

This example investigates linguistic structures and their transformations through the translation process. The grammars of Greek and Arabic are significantly different from each other, so that in many cases the part of speech of a word had to be changed. For example, a direct equivalent of the Greek gerund does not exist in the Arabic grammar, forcing a translator to choose another part of speech. Additionally, the idiomatics of a language and the interpretative function of the translator exert additional strong influences on part of speech transformation. This creativity in the translation process can also be visualized for (or communicated to) users who do not possess expertise in either or both of the languages.
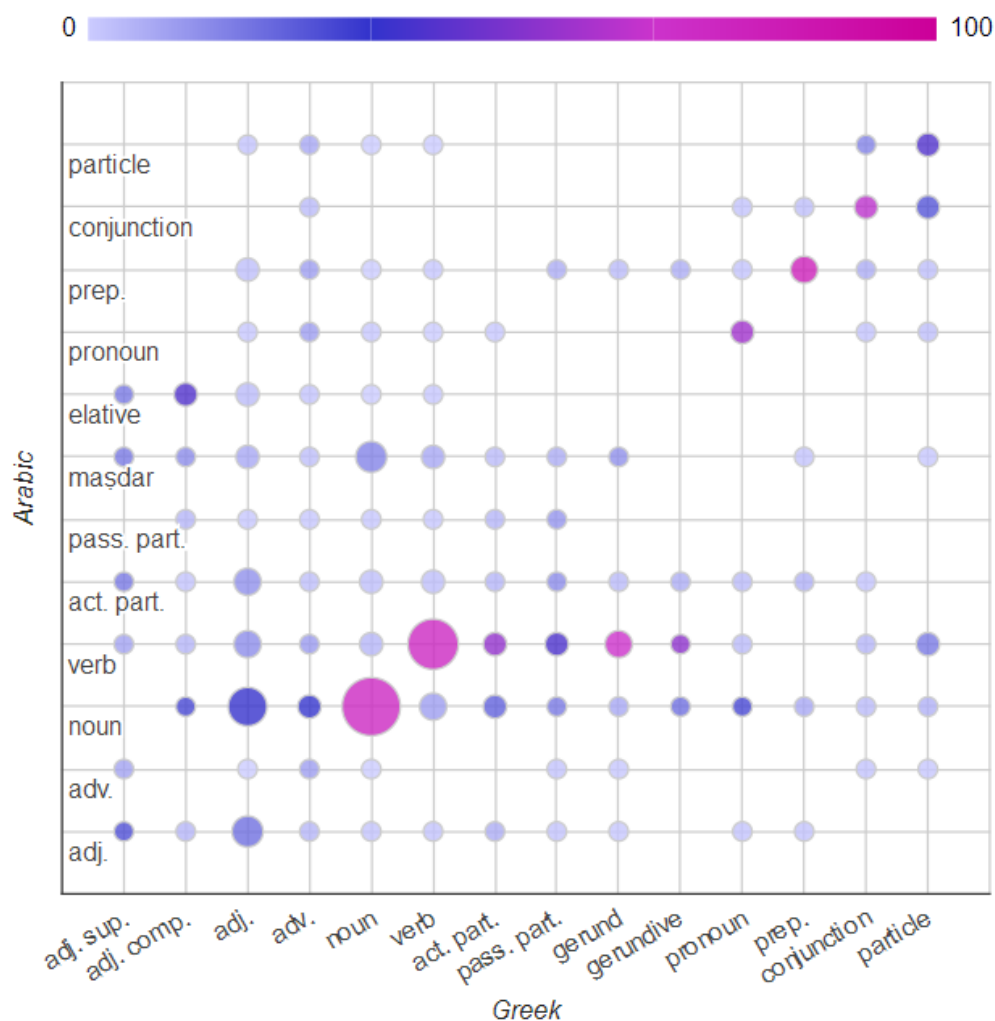
Figure 11: Glossarium Graeco-Arabicum, Compared Parts of Speech for Ḥunayn b. Isḥāq.[33]

The diagram (cf. figure 11) correlates Greek parts of speech (on the x-axis) to Arabic parts of speech (y-axis) based on the frequency of transformation, in this case focused on the translator Ḥunayn b. Isḥāq (808–873 AD). The size of each intersection represents the absolute frequency, while the color reflects the relative frequency for the respective part of speech. It becomes visible, for example, that Hunayn transformed Greek gerundives and gerunds most frequently into finite verbal forms; and among the parts of speech that exist in both languages, he transformed adjectives often into nouns, and nouns sometimes into maṣdar (a nominal form that

has similar qualities to gerund). Among the more syntactic parts of speech (pronouns, prepositions, conjunctions, particles), there is a lot of transformation within this group, as the "island" in the upper right corner displays. Generally, there is also a strong tendency of transformation into nouns or verbs, as can be seen in the strong accumulations on the respective horizontal lines. Furthermore, the visualization offers a link to the underlying data: A click on an intersection leads to a list of all respective word pairs in order to investigate them on a more detailed level.
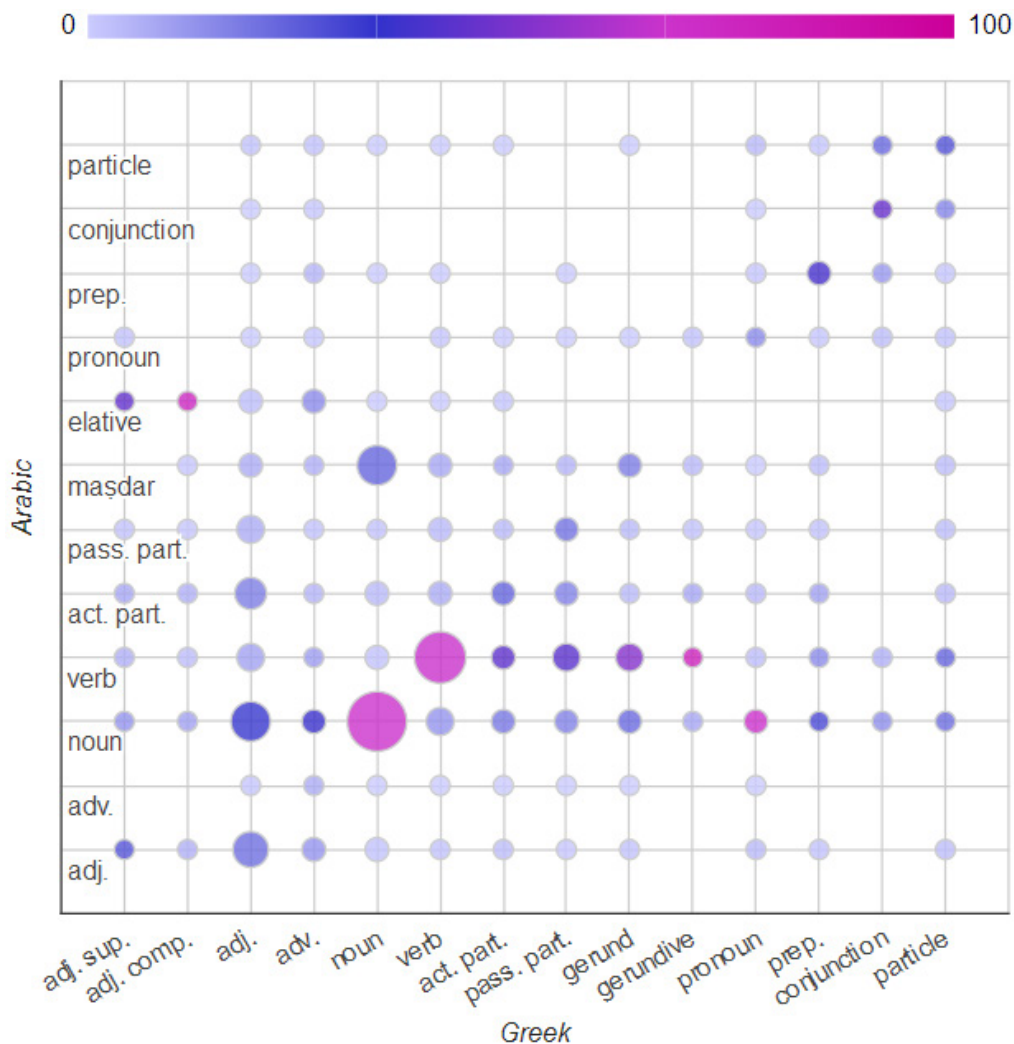
Figure 12: Glossarium Graeco-Arabicum, Compared Parts of Speech for Isḥāq b. Ḥunayn.[34]

The diagram for Isḥāq b. Ḥunayn (son of Ḥunayn b. Isḥāq, 830–910 AD, cf. figure 12) is significantly different: The "syntax island" in the upper right is less recognizable, while the tendency towards noun and maṣdar is much stronger, as well as the frequency of elative. This reveals that part of speech transformation is potentially a distinct characteristic of a translator.

---

34   https://telotadev.bbaw.de/glossga/author_charts.php?person_id=100026&chart_type=compared-pos-grid (screenshot: January 18, 2016).

3.4. Translation Variants

The last example shifts from the quantitative and grammatical to the semantic level, focusing on the translation processes of individual words. It belongs to the most important tasks of a bilingual dictionary to depict the variety of translation possibilities. The reason for variety is partly due to the different contexts in which a word is used, and secondly, to the interpretation of the translators. To start a deeper investigation of this issue, the example experiments with a Sankey diagram.
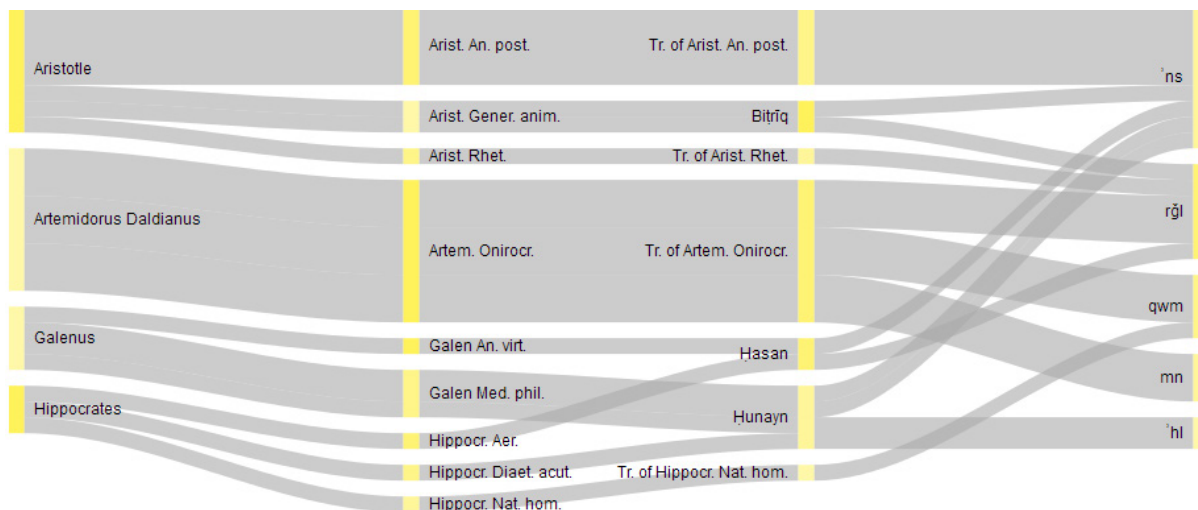
Figure 13: Glossarium Graeco-Arabicum, Translation process scheme for ἄνθρωπος.[35]

Read from left to right, the diagram (cf. figure 13) shows the translation process for the word ἄνθρωπος (anthropos, "man"). It appears in various texts by Aristotle, Artemidorus, Galen and Hippocrates, most frequently (as seen in section 3.1) in the extensively represented sources *Oneirocritica* and *Analytica posteriora*. In most cases it is translated with أنس ('ns, "human"), but also often with رجل (rğl, "man").[36] Apparently, the translators Biṭrīq and Ḥasan covered exactly these two cases, while Hunayn preferred the variant أهل ('hl, "people, group"). The variants من (mn, "someone") and قوم (qwm, "crowd") were particularly used by the unknown translator of *Oneirocritica*.

Within this diagram, the user can click on a node (i.e. a vertical bar) to highlight the corresponding connections to this node. When looking at the nodes on the right side, a greater diversification of variants أنس ('ns, "man") and رجل (rğl, "man") can be noticed, while the other variants, in almost every case, derive from only one translator each. This suggests that the latter versions are individual cases, due to a special meaning in the original text or a special interpretation by the translator.
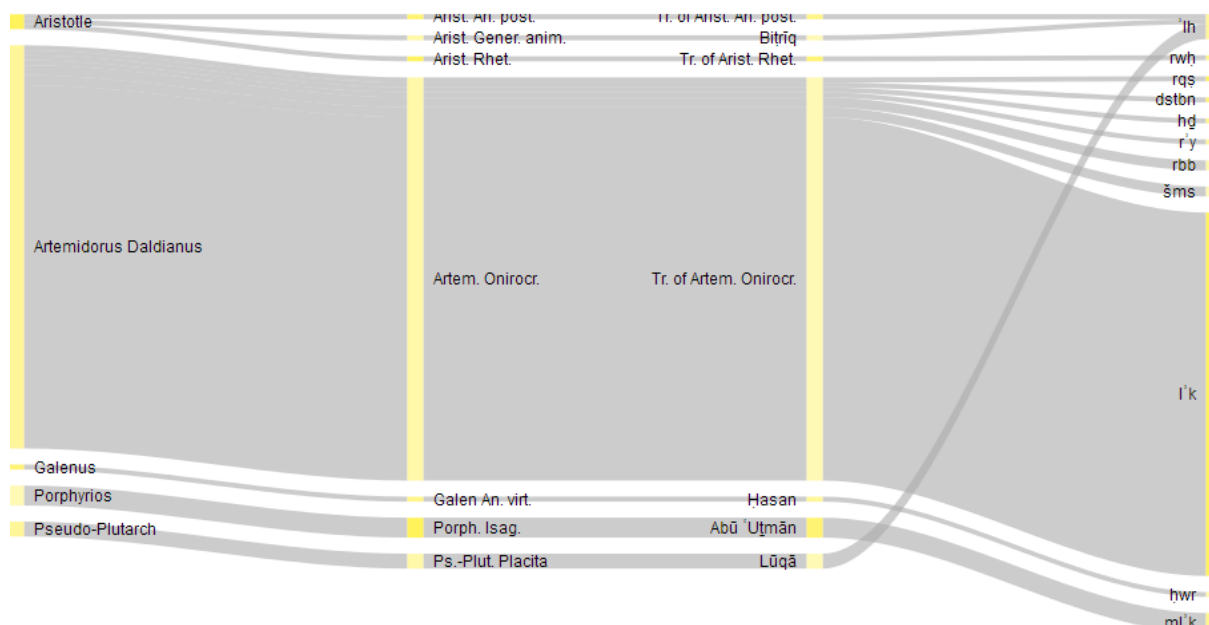
Figure 14: Translation process scheme for θεός.[37]

For a comparison, the Sankey diagram for the word θεός (theos, "god") is considered (cf. figure 14). Here, the scheme displays a completely different scenario: The translation of *Oneirocritica*, in which θεός translates most frequently into ملك (l'k, "angels"), dominates the diagram. This is a monotheistic interpretation of the Greek word "god", in the original context used in a plural form, degrading the Greek "gods" to Arabic "angels" (all under one God). For "god" in a singular meaning, which is less frequent in the Greek texts, also the today familiar Arabic form الله ('lh, "Allah") can be found, though it occurs mainly in the translations of Pseudo-Plutarchian and Aristotelian writings.[38]

However, Sankey diagrams have their limitations, as they produce less clarity and much more confusion in cases that are more complex: A display of a huge amount of variants would result in an unreadable, giant, spaghetti-like diagram. Data aggregation, clustering similar variants into groups, would help in such cases. Another desideratum in this case is the inclusion of chronological data; however, the chosen API is not yet sufficiently configurable (whether it will provide these options in one of the future releases remains an open question).

4.   Results

The four presented examples demonstrate from very different angles how to visualize translation structures on various application levels, and how these visualizations can help to investigate various areas of research interest. Concerning corpus structures, lexical and grammatical differentiation, grammatical transformations and translation processes as well, visualizations help to identify phenomena that would not have been recognizable and comprehensible though a textually oriented interface. These revealed structures raise new questions regarding the material. Especially for a complex and large data collection like the Glossarium Graeco-Arabicum, which already tends to grow beyond human understanding, visual "distant reading" seems a fruitful approach to get a fresh perspective on the material.

---

37   https://telotadev.bbaw.de/glossga/glossary.php?gr_lexeme=θεός (screenshot: January 18, 2016).

38   GALex, however, gives still more records for "Allah" and covers more sources.

In addition, those structures are recognizable both for professionals as well as for amateurs, and thus fulfill both a professional and didactic purpose, from the corpus overview down to the single lexemes: Nonprofessionals can make a quick start through visual aids, while an expert can detect special phenomena, which could be worth a closer investigation. In all cases, the visualizations provide a good degree of transparency, as the underlying data is accessible through a simple link on the interactive display. Finally, yet importantly, even the database editors can profit from the visualizations, as they give an opportunity to check the integrity of the data and the consistency of the whole corpus.

Visualizations create impetus and inspiration through their direct, immediate effect. It is, however, crucial that such presentation forms provide also sufficient information about the database structure and the data representation. If those structures are not properly explained, there is a high risk that visualizations suggest correlations that later reveal as fallacies. Transparency is as a key here: Misunderstandings are avoidable by accurately documenting and explaining structures and collection methods of a corpus. Still, visualizations – understood in the meaning of mappings – will neither explain nor interpret anything about themselves: Instead, they recite data in their very special way, which still needs to be interpreted by the user, and explanations can only be given based on in-depth studies. Visualizations can support this in their function of a mediator, as long as it is clearly understandable what is actually being displayed and how it is displayed. The more consistently they do this, the more they will become a compact graphical narration and actually get closer to the ideal of becoming an independent scholarly media form.

5. Literature

Alexander (2010): Marc Alexander, "The Various Forms of Civilization Arranged in Chronological Strata. Manipulating the Historical Thesaurus of the OED", in: Michael Adams (ed.), Cunning passages, contrived corridors. Unexpected Essays in the History of Lexicography, Monza, 2010.

Arzhanov/Roeder (2013): Yury Arzhanov and Torsten Roeder, "The Glossarium Graeco-Arabicum. Linguistic Research and Database Design in Polyalphabetic Environments". Lecture at the Digital Classicists Berlin Seminar on November 19, 2013. http://hdl.handle.net/11858/00-1780-0000-0022-D548-B (permalink).

Arnzen et al. (2012): Rüdiger Arnzen, Yury Arzhanov and Gerhard Endress, "Griechische Wissenschaft in arabischer Sprache", in: RUBIN Wissenschaftsmagazin, spring 2012, 14–21. http://rubin.rub.de/de/griechische-wissenschaft-arabischer-sprache (March 5, 2017).

D'Ancona (2013): Cristina D'Ancona, "Greek Sources in Arabic and Islamic Philosophy", in: Edward N. Zalta (ed.), The Stanford Encyclopedia of Philosophy, 2013. http://plato.stanford.edu/entries/arabic-islamic-greek/ (March 5, 2017).

Endress/Gutas (1992): Gerhard Endress and Dimitri Gutas, A Greek and Arabic Lexicon (GALex), Leiden, 1992ff.

Endress et al. (2013): Gerhard Endress, Rüdiger Arnzen and Yury Arzhanov, "Griechische Wissenschaft in arabischer Sprache. Ein griechisch-arabisches Fachwörterbuch der internationalen Wissensgesellschaft im klassischen Islam", in: Studia graeco-arabica 3 (2013), 143–156. http://www.greekintoarabic.eu/uploads/media/BOCHUM_SGA_3-2013.pdf (March 5, 2017).

Leydesdorff/Welbers (2011): Loet Leydesdorff and Kasper Welbers, "The semantic mapping of words and co-words in contexts", in: Journal of Informetrics 5, No. 3 (July 2011), 469–475.

Lüdeling/Zeldes (2009): Anke Lüdeling and Amir Zeldes, "Three Views on Corpora: Corpus Linguistics, Literary Computing, and Computational Linguistics", in: Jahrbuch für Computerphilologie 9 (2009), 151–180. http://computerphilologie.tu-darmstadt.de/jg07/luedzeldes.html (March 5, 2017).

Nasr (1976): Seyyed Hossein Nasr, Islamic Science. An Illustrated Study, World of Islam Festival Publishing, 1976.

Oakes/Ji (2012): Michael P. Oakes and Meng Ji (ed.): Quantitative Methods in Corpus-Based Translation Studies. A practical guide to descriptive translation research (= Studies in Corpus Linguistics 51), Amsterdam/Philadelphia, 2012.

Roeder (2015): Torsten Roeder, "Alpha into Alif. Schnittstellen zwischen Schriftkunde und Informatik am Beispiel von Unicode im Glossarium Graeco-Arabicum", in: Studia graeco-arabica 5 (2015), 345–363. http://www.greekintoarabic.eu/uploads/media/14-SGA_V_2015_ROEDER.pdf (March 5, 2017).

Roeder (2016): Torsten Roeder, "Mapping the Words. Experimentelle Visualisierungen von Übersetzungsstrukturen zwischen Altgriechisch und Hocharabisch", in: Zeitschrift für digitale Geisteswissenschaften. http://dx.doi.org/10.17175/2016_006 (March 5, 2017).

Roeder/Arzhanov (2016): Torsten Roeder and Yury Arzhanov, "Experimental visualizations of translation structures between Ancient Greek and Classical Arabic", Lecture at the Digital Classicists Berlin Seminar on January 19, 2016. http://hdl.handle.net/11858/00-1780-0000-0029-C04E-B (permalink).

Shneiderman/Plaisant (2009): Ben Shneiderman and Catherine Plaisant, "Treemaps for space-constrained visualization of hierarchies", 2009–2014. http://www.cs.umd.edu/hcil/treemap-history (March 5, 2017).

Thudt et al. (2012): Alice Thudt, Uta Hinrichs and Sheelagh Carpendale, "The bohemian bookshelf: supporting serendipitous book discoveries through information visualization", in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12), New York, 2012, 1461–1470. http://www.alicethudt.de/BohemianBookshelf/material/ThudtCHI2012.pdf (March 5, 2017).

Tufte (2001): Edward R. Tufte, The Visual Display of Quantitative Information, Creshire, 2001.

Author[39]

Torsten Roeder
Julius-Maximilians-Universität Würzburg
Institut für Musikforschung
Domerschulstraße 13
D-97070 Würzburg

Email: torsten.roeder@uni-wuerzburg.de

Gods, graves and graphs – social and semantic network analysis based on Ancient Egyptian and Indian corpora

Frederik Elwert, Simone Gerhards & Sven Sellmer

Abstract: In this paper, the authors show the application and use of automated text network analysis based on ancient corpora. The examples draw from Ancient Egyptian sources and the Indian Mahābhārata. Different text-based network generation algorithms like "Nubbi" or "Textplot" are presented in order to showcase alternative methodological approaches. Visualizations of the generated networks will help scholars to grasp complex social and semantic text structures and serve as a starting point for new research questions. All tools for applying the methods to ancient corpora are available as open source software.

1. Introduction

Network analysis as an analytical approach is quite popular in the digital humanities. Its methodological foundation, mathematical graph theory, is a generic way of modelling entities (nodes) and their relations (edges), and can be adopted to a variety of application domains. For humanists, inspiration often stems from the sociological branch of network research, Social Network Analysis (SNA). Modelling the connections between historical[1] as well as fictional[2] persons is an obvious application of SNA as a means to study historical eras and literary works in a distant reading fashion. But network analysis is not limited to studying personal networks. For philological research, approaches borrowed from computational and corpus linguistics can also help to highlight the connections between concepts in a text or a corpus.

The aim of this paper is to give an overview of some of the very different ways in which we can use network analysis in order to study ancient texts. Given the nature of such a methodological overview, we will be able to only briefly touch on the details of the different approaches. Examples from our own research on Ancient Egyptian and Indian corpora will be used to show the practical value of the different approaches and highlight their differences. Our paper will follow a path from social network analysis on the one end, which is probably better known, to semantic text network analysis on the other end, and various combinations in between.

---

1   Gramsch (2013).

2   Moretti (2011).

2. The Project

The research presented in this paper is based on our work in the project "semantic and social network analysis as a means to study religious contact" (SeNeReKo). SeNeReKo was a joint project between the Center for Religious Studies in Bochum and the Trier Center for Digital Humanities from 2012 till 2015.[3] The team comprised one computer scientist, one Egyptologist, one Indologist, and one scholar of religion. The aim of the project was to develop and apply new computational methods for the study of religious history. One major issue when applying computational text analysis methodology to historical corpora is that we work with languages for which relatively few tools and linguistic resources are available, compared to modern languages. This poses a challenge when relying on techniques from computational linguistics for distant reading approaches. At the same time, this allows us to evaluate the requirements for a real-world application (given the issues mentioned) of different methods for the study of historical corpora.

3. The Sources

On the Indian side we worked with two large corpora of a little more than 1.5 million lexical units[4] each:

1. a collection of Buddhist canonical texts composed in the Middle Indian language Pāli, the so-called Pāli Canon;
2. the Sanskrit epic *Mahābhārata*

Here, we will focus on the latter text. It is traditionally counted as an "epic", but differs in several respects from the European representatives of this genre. It is not only much longer than the Homeric epics (ten times the length of the Iliad), but also considerably more diverse: apart from the main plot (a family feud) and lengthy battle descriptions, it also features numerous stories and tens of thousands of lines containing philosophical and ethical teachings.

Coming to the technical side, both corpora are available in digital form, but only as plain text files – which means that they have to be pre-processed (lemmatised etc.) before they can be used for analyses of the presented type. This is a difficult task, because in addition to the problems connected with flective languages in general, Sanskrit and Pāli pose further, very substantial ones, especially the phonetic changes called *sandhi* (see fn. 4). Luckily, for the *Mahābhārata* we were able to use a lemmatised text (partly with part of speech information) that was prepared by the pioneering computer program SanskritTagger of our kind colleague Oliver Hellwig.[5] For the Egyptian material we used the text corpus from the database of the Thesaurus Linguae Aegyptiae from the Berlin-Brandenburgische Akademie der Wissenschaften. They kindly provided us their digitized and annotated database, which contains more than

---

3 The SeNeReKo project was funded by the German Federal Ministry of Education and Research, project number: 01UG1242A. The authors of this paper are responsible for its content.

4 Speaking of "lexical units", we refer to the surface forms of words, which in Sanskrit differ according to the phonetic context. E.g., the nom. masc. sg. form devaḥ ("god") may appear in the text as: devaḥ, devas, devaẖ, devaś, devo, or deva. This phenomenon is known as "sandhi".

5 Hellwig (2010).

one million lexical units. The texts go back as far as the third millennium BCE. The corpus consists of different genres like religious hymns, biographic inscriptions or literary and medical texts. The length of the texts varies from a few words to more than 8 000, so that one can speak of a heterogeneous text corpus. Originally, the texts were written on papyrus, stone, ostraca or, for instance, temple walls.

4. An Introduction to Network Analysis

Applying network analysis requires building a network model. A network model consists of *nodes* and *edges*. As with any kind of modelling, this requires an abstraction from the given data and reducing it to its core features.[6] In the most common case of social networks, nodes are people, and edges are relations between them. In the case of online social network platforms like twitter, registered users are nodes. And if user A follows user B, that constitutes a relation, or edge. In this case, a certain network model is already given and technically enforced when using the platform's functionality. But of course, social networks as a social phenomenon pre-date the corporate adaptation of that term. Network ties are also constituted by letters between people,[7] or by political alliances,[8] or by mutual support[9]. But networks in the humanities don't have to consist of people. Everything that can be described as a set of relations between units of some sort is a network. In the remainder of the article, we will start with social networks, i.e. networks between people. Due to the project's background of religious studies, our personal networks will also include gods. From there, we will then move more and more away from pure social networks to other types of networks that one might study in the humanities, like conceptual and text networks. While the former allow us to study the use of single words, the latter represent complete texts as networks.

Whatever our nodes and relations are, the formal language of mathematical graph theory allows us to mathematically analyse a given network, and ask questions like: "Which node is the most central or important one in the network?" Or, "Which smaller groups can I identify in my network?" A network model is attractive for humanities research, since it highlights the relational aspects of our data: The importance of a person in relation to others, the constitution of the meaning of a word from its relation to other words, and the meaning of text as an emergent property of the relation of the words it is made of. Relational modes of thinking resonate with recent theoretical developments in different areas of the humanities. However, network models are not the only kind of model that captures relationality, and a network might not be the best kind of model for any given research question. Other methods like topic models or neural word embeddings can be used for similar purposes and might be more appropriate when large amounts of textual data are available.

---

6   What core features are is not absolute: It depends on the type of model (e.g., a directed or an undirected network), but also on the research question. In this sense, core features for a network model are much more constructed than found in the data.

7   Winterer (2012).

8   Gramsch (2013).

9   Düring (2015).

5. Social Network Analysis

To start with an example of social network analysis as applied to literary texts, we strived to obtain an overview of the important persons in the *Mahābhārata* and their mutual relations. Here, we had to solve two problems:

1. How to identify persons?
2. How to define their relations?

Regarding the first problem, it was impossible to take into account coreferences because this would have required an enormous amount of analysis by human experts. But even when looking only at names, the task is more tricky than one might expect, because important figures tend to appear not only under their primary names, but also under (in many cases multiple) epithets; in addition, sometimes the same epithet is used for several persons. These problems had to be, therefore, tackled with the help of manual checking.

As to the second question, it is clear that, for the present purpose, a relation must be recognizable by a computer, so we chose a very technical criterion: a relation (represented by an edge in our social network) exists where two persons appear in the same verse (i.e., in the case of the *Mahābhārata*, mostly in the space of two lines).

On this basis a social network can be quite easily constructed once the data are prepared. Figure 1 shows the result for the most frequent persons of the *Mahābhārata*.

Figure 1: Social network of the main protagonists in the *Mahābhārata*.

This graph is mainly meant to demonstrate the ability of the computer to automatically detect so-called "communities", i.e. clusters of nodes/persons that are more closely connected with each other than with the rest of the network. This becomes nicely visible in figure 2.
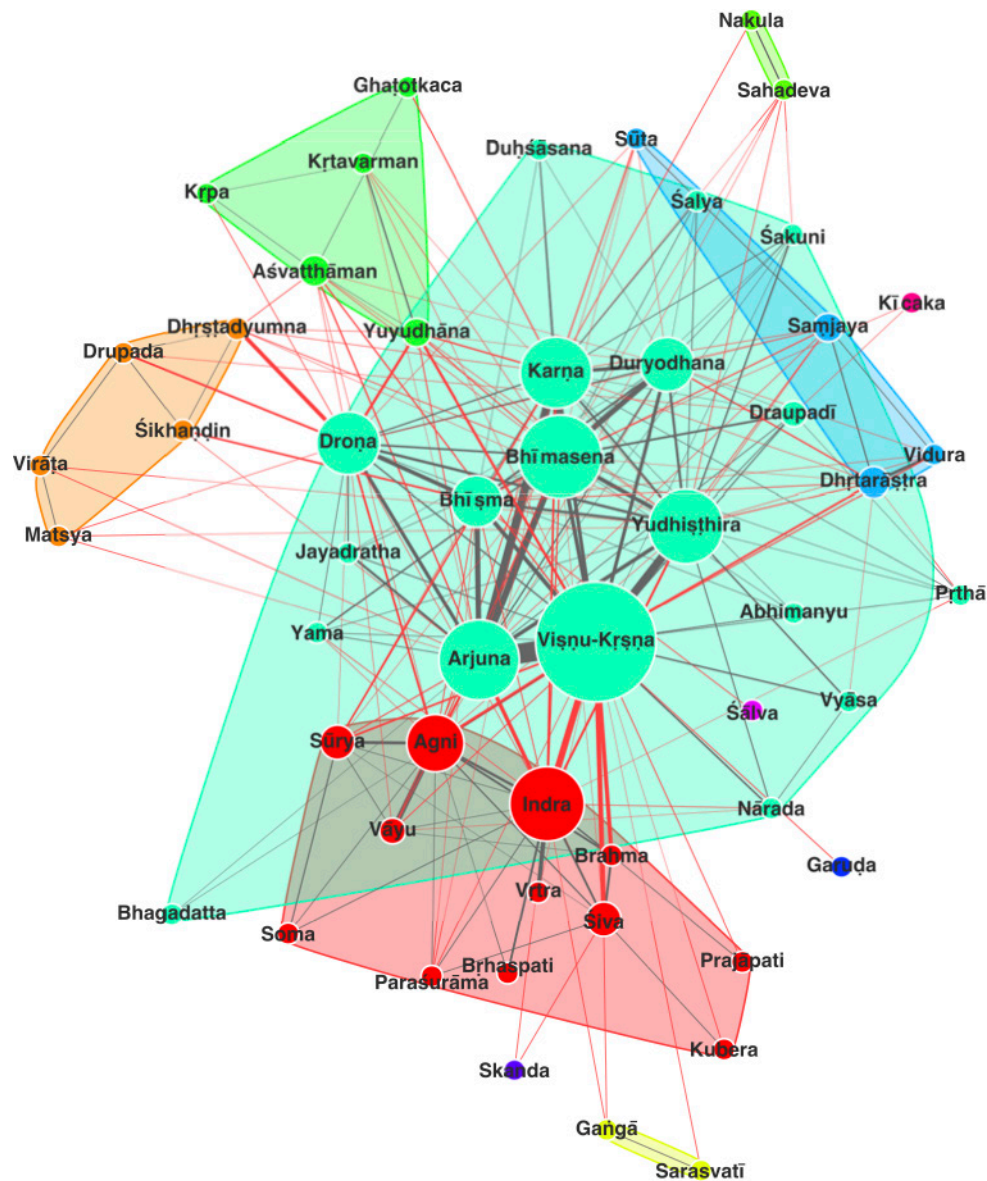
Figure 2: Communities in the *Mahābhārata*.

Here, the software (automatically!) groups the gods together (red), with the exception of Yama, the god of death, who belongs to the central green community, which comprises most of the principal actors. The small subgroups consist of heroes that are particularly closely linked among themselves. The singletons represent such persons that do not form part of the main plot but play an important role in one specific episode.

6. Semantic social networks

The previous example provides a good impression of the social structure of the *Mahābhārata*. But social structure is not everything that we want to study. Especially in the humanities, but of course also in the social sciences, we are interested in content, in semantics: What is this network actually about? Who are these people, and of what quality are their relations? Instead of the simple kind of network model introduced above, a semantic social network can be used that contains additional information about the connotations associated with its elements.

One way to add semantic information to networks is what we call the typed edges model:[10] The network model can be enriched by adding information about the kind of relations that we observe. So we could distinguish between friendship, teacher-student-relations, and enmity. This usually requires to build a typology of relations that guides data collection. The researcher has to decide which kinds of relations are taken into account and define criteria for their identification. Depending on the research philosophy, this deductive approach can be an issue. In our case of ancient cultures, we did not want to impose a given typology (which might be derived from modern-day western concepts) on our material. Instead, we were interested in discovering the differences in relations that were expressed by the sources themselves. As a consequence, we followed a more inductive way of creating a typed model of the *Mahābhārata's* social network.

The basis for this is an algorithm called "Nubbi".[11] Nubbi utilizes topic modelling, a machine learning technique that allows to identify latent semantic structures in texts.[12] A topic in this sense is expressed by a list of thematically related words. Topic models use the word distribution across documents (or document sections) as information to automatically assign words to topics. A typical word list produced by a topic-modelling algorithm might consist of the words "ratha" (chariot), "śara" (arrow), "raṇa" (battle) and "han" (to kill). Labelling such topics is always an interpretative act. The labels used here, e.g. "fighting" for the given word list have been assigned manually.

Nubbi extends this model by distinguishing between topics that describe entities, or nodes, and topics that describe relations, or edges. For this purpose, it makes use of the text that surrounds the occurrence of an entity or a relation in the corpus. When a single entity is found in the text, the surrounding text is used for finding entity topics. When multiple entities are found, the text contributes to the topics describing the relations between those entities.[13] We assume that the words near the mentioning of a relation in the text can be used to infer the quality of the relation, and of the entities that are part of it. This translates roughly to the idea of different types of actors and types of relations in a network, but it is more like a thematic connotation. Figure 3 illustrates this.

---

10   In network analysis terminology, this is often called a multiplex network, if multiple relations of different kinds are allowed between two nodes.

11   Chang et al. (2009).

12   Brett (2012).

13   The actual model treats relation documents as mixtures of entity and relation topics: In addition to the relation topics, also the entity topics of both individual entities contribute to the word distribution.
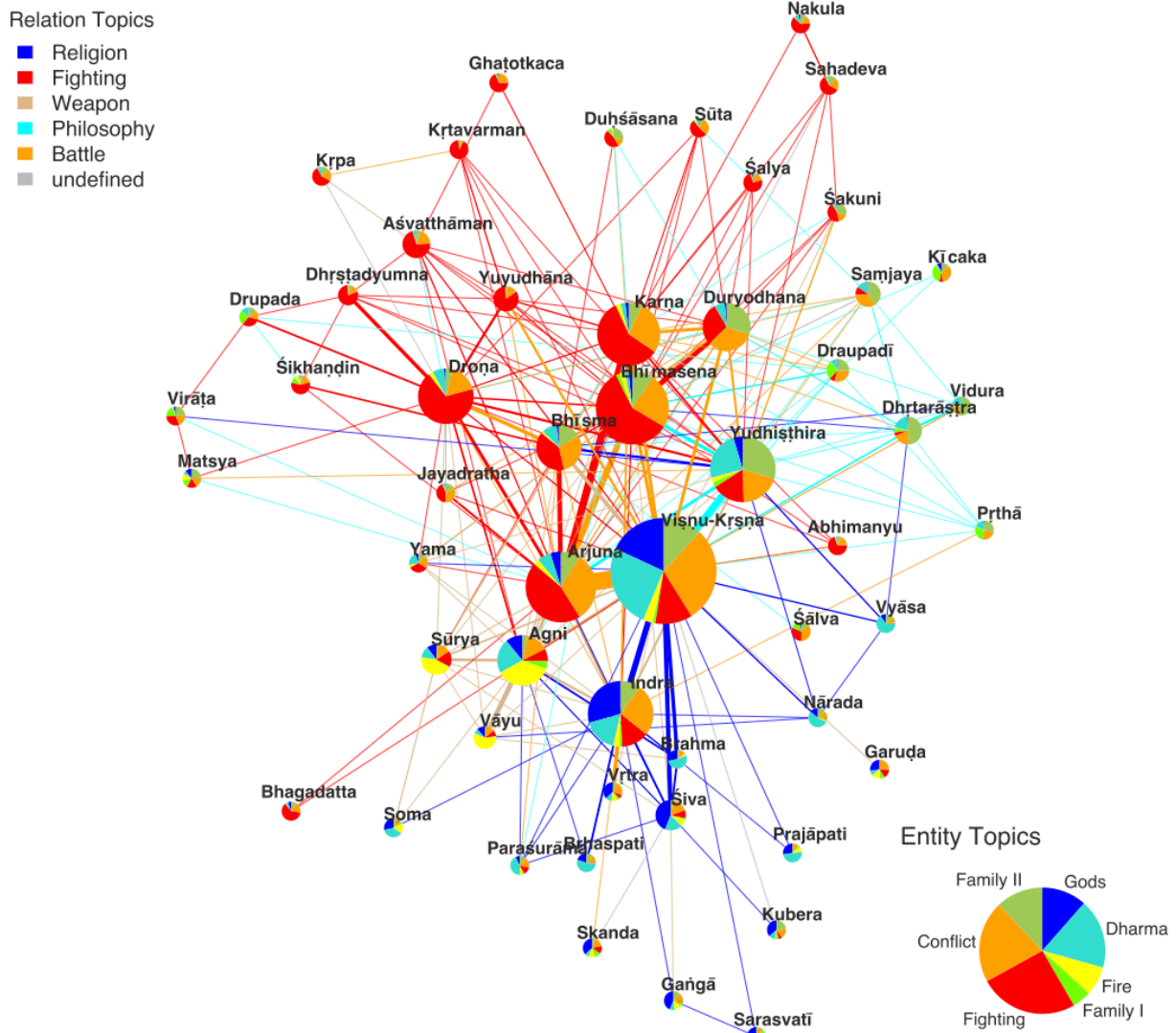
Figure 3: Semantically enriched social network of the *Mahābhārata*.

Here we have the same network as before, but with the information generated by Nubbi added through colouring the nodes and edges. Let us start with the persons that are now represented by pie diagrams. The size of the differently coloured sectors corresponds to the percentage with which the persons are associated with the single topics. The general distribution of the topics is visible on the pie in the bottom right corner. It is important to note that our domain expert has added the names of the topics. The program only gives a list of words that belong to (or constitute) a topic. Sometimes these "computer topics" are rather surprising or even unintelligible to the human interpreter, but in the present case they were more or less humanly understandable, so it was possible to attach a kind of "title" to each of them.

As just mentioned, Nubbi also extracts topics that are characteristic for the *relation* of two entities (therefore we call them "relation topics" – see left upper corner). They are represented by the colour of the edges. Because most edges are rather small, we chose to use only the colour of the predominant topic. So, for example, looking at the red connections – which symbolise fighting – one can easily identify the main enemies.

Now, one could continue to enumerate the pieces of information hidden in this graph – mostly intuitively convincing for an expert of the *Mahābhārata*, but sometimes also astonishing and intriguing – but it seems better to point to two general observations we made in the present context:

1. Persons may appear both as actors and in other, more figurative functions. E.g., when it is said that a hero "shines brightly like Sūrya" (the sun god) or that one warrior sends another "to Yama's abode" (= to the god of death, i.e., kills him), then Sūrya and Yama have a very different role from the usual ones of agent or patient. Since these metaphorical uses do not constitute interpersonal relations in a classical sense, they might be undesirable. (Indeed we decided to remove the Yama verses of that type from the network, keeping only the relation between the killer and the killed.)

2. Certain kinds of relation topics are structurally underrepresented, especially the philosophical ones, because often lengthy philosophical instructions are prompted by a simple question, but follow only in the subsequent lines and are therefore not recognized as belonging to the relation questioner – answering person.

These, and cognate, phenomena call for future improvements, but even now we hope to have shown that by refining simple co-occurrence networks it is possible to model and visualize the semantic aspect of social relations (as reflected in textual content) to a useful and, according to our impression, sometimes astonishing degree.

7. Semantic Context Networks

In the previous case of the semantic social network, a network model is used only to capture the interpersonal relations. The semantic dimension is analysed using a different kind of model, here a topic model. Both kinds of models are integrated in a way that the social network informs the topic-modelling algorithm. Topic-modelling is also relational in the sense that it is based on co-occurrence of words that form the context of the elements of the network. The words that surround a reference to a person or a relation – or any other entity – in a text are not purely coincidental; we can assume they have some sort of semantic relationship to that entity. However, the internal state of the model is somewhat opaque, making it difficult to analyse these co-occurrences on a local level. For a specific entity, we get only a list of words associated with that entity as a result of the learning process, but we cannot inspect the nature and form of those associations. But if we regard these context words as related to the entity in question, then we can describe them as a network as well. We call this kind of network representation of semantics the "semantic nodes model": Here, not only social actors, but also words are nodes of the network. The semantic information is thus contained in the network structure itself. This is suitable for examining the semantic context in more detail than what a topic model allows.

To build the context network, we used a co-occurrence based algorithm.[14] These algorithms assume words to be related if they appear close to each other, possibly adding extra weight to the edges based on the proximity of the words. Then, directly neighboured words would have a heavier connection than words in greater distance. Such algorithms have been used to model whole texts.[15] In our research, we found that these networks are often difficult to interpret on the global level once the underlying texts become too large or too diverse. This makes the method less suitable for studying medium to large corpora. However, they are useful for modelling the local neighbourhood of words. Thus, we apply co-occurrence networks mainly to study the semantic context of individual words or entities. This use resembles techniques used for word sense induction in computational linguistics, but with a different research question. Here, we

---

14   Paranyushkin (2011).

15   Lietz (2007).

aim not at inferring the different senses of an ambiguous term as an intermediary step in text processing, since our corpora already contain disambiguated lemma information. Rather, we aim at studying the connotations associated with an entity or word in historical corpora.

An example will explain better what this means in concrete terms, so let us have a look at the semantic context of the god Horus from the pyramid texts.

The pyramid texts are a collection of ancient Egyptian religious spells from the Old Kingdom. They are written in hieroglyphic script and are inscribed on the walls of the pyramids from Saqqara, i.e., from about 2.350 till 2.100 BCE. In the Old Kingdom, the use of the texts was exclusively reserved for the king; after the Old Kingdom, copies of the spells can be found on tomb walls etc. of non-royal persons as well. The spells are concerned, for instance, with the protection of the body, the preservation of the name and the ascend to the heaven. Furthermore, they could be used to call gods to help the king. There are in total about 750 different spells, which are never used all together in one single collection. The whole corpus preserves the largest body of inscriptions known from that age.[16] To obtain a representative result, we chose that corpus as basis for the semantic network, because it is comparatively large, spans a rather short time period and belongs to one text genre. Horus is one of the oldest Egyptian gods and can for example be represented as a falcon or a falcon-headed human. He has many functions in the Egyptian pantheon, but has in general an affiliation to the sun, war and protection.

For the following network we used all spells in which the lemma "Horus"[17] is mentioned. But we considered every spell just once, no duplicates were used. To create the network, we used a special kind of co-occurrence algorithm.
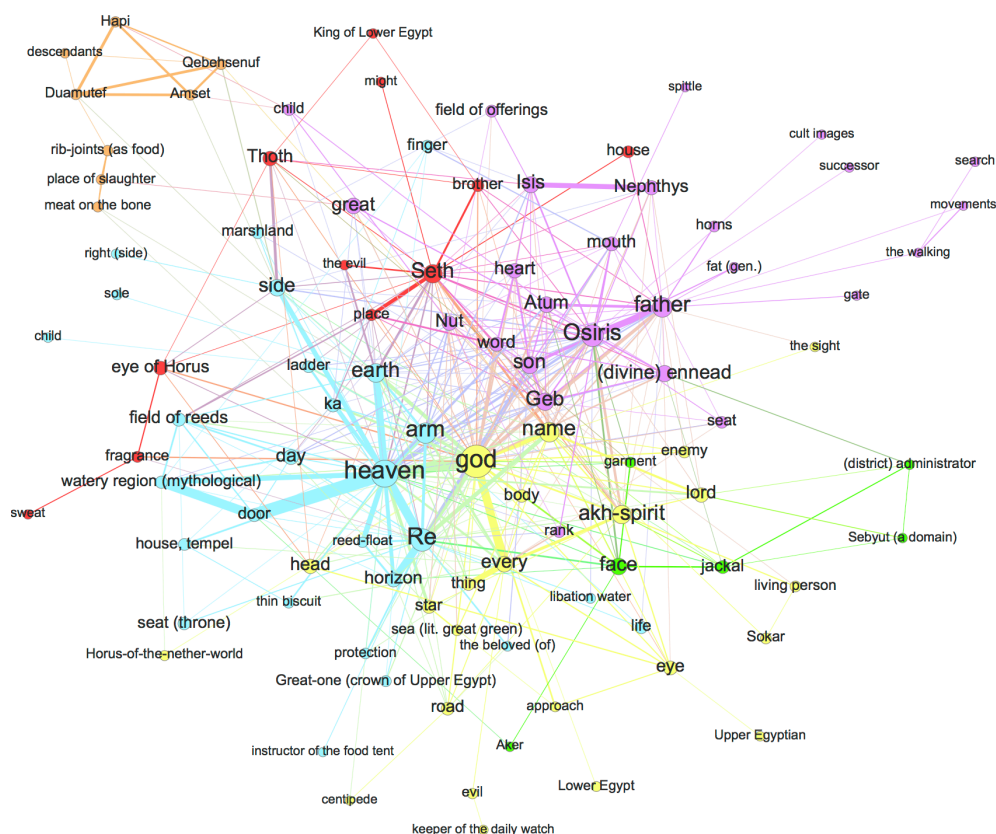

Figure 4: Context network of Horus in the pyramid texts.

16   Allen (2015).

17   TLA lemma entry No. 107500.

In order to identify different contexts in which the lemma is used, a community detection algorithm was applied, assigning the nodes to different groups. This structure, therefore often called a community structure, describes how the network is compartmentalized into sub-networks (see figure 4). As a result, all words[18] are marked by one of six colours, every colour standing for a different context. As for the size of the nodes, the bigger a node, the more central it is for the context of Horus. So the words "god", "heaven", "Re", "name", "Osiris", "father", "arm" and "Seth" have the highest degree. Here, the question may arise: "Where is Horus represented in the network?" The answer is, "nowhere". Because he is by definition connected to all words, it is not necessary to show Horus explicitly in the network. Now we take a look at the details. First, we will have a look at the violet community, which consists most of the other gods and of information about their relationship to Horus. Osiris is the father of Horus, so Horus is his son. Isis, in turn, is the mother of Horus, and the sister of Nephthys. Furthermore, we find the divine siblings Geb and Nut, and Atum. The node "child" leads to a second community of relatives, the orange one of Horus' children (Hapi, Duamutef, Kebekhsenuef and Amset). In the pyramid texts, one of their main purposes is to supply the descendent with food, as visible in the network. So they do not occur in the same context as the other gods. In the centre of the network are body parts that are semantically connected to Horus like "arm", "mouth", and "heart". The blue community shows Horus' affiliation to heaven and afterlife. Here appears the sky divinity Re, and the heavenly region, but also words that show the way to the afterlife like "door" or "ladder". The last community to be mentioned here is the red one, which belongs to the god Seth. He is the uncle of Horus, but also his competitor. In Egyptian mythology, Seth is portrayed as the usurper who killed and mutilated his own brother Osiris. Horus sought revenge upon Seth, and the myths describe their conflicts. In the pyramid texts, Seth occurs as the (evil) counterpart who needs to be defeated.[19] To sum up the main points, the network shows us the words most relevant for Horus in the pyramid texts. In 1916 Thomas George Allan wrote his egyptological Dissertation about "Horus in the Pyramid Texts"[20] where he analysed on a semantical basis the relations of Horus to other divinities, body parts or, for instance, the king. It is very interesting that he obtained the same results that our network shows at a glance: Horus' strong connection to heaven, his function of helping the dead king going to "heaven", the connection to his father Osiris and the connection to body parts like "arm", "mouth" and "heart", and the important role of Seth.

8. Text Networks (textplot)

The previous example used a co-occurrence algorithm to model the semantic context of an actor, the god Horus. But the same technique can be used to study the context of any given word. Following this path, the network model used no longer resembles to a social network, but a word network. The application of this kind of networks is not limited to the study of word contexts, but can also be used to model larger units like entire texts. Of course, a network model of a text is an abstraction, and one loses a lot of detail that is contained in the syntactic structure of the sentences. But following the idea of distant reading,[21] a text network should highlight some information that is harder to grasp otherwise.

---

18   The English words are based on the TLA translation of the lemma entries.

19   Meurer (2002), 99 passim.

20   Allen (1916).

21   Moretti (2013).

As mentioned above, co-occurrence networks have indeed been used to this end.[22] However, we found that the basic principles of co-occurrence do not scale well: The larger the text, the more blurred and hard to interpret the results become. While they work very well to capture local structures, they are less suitable to express the macro-structure of larger textual units. This problem is tackled by a different approach of building text networks called textplot.[23] It uses an intermediary abstraction to highlight global word relations instead of local phenomena. The basic idea is to add an edge between two words if they appear in the same passages throughout the text. To this end, it defines relations between words in terms of similarity of their distribution across the whole text.

Technically speaking, a kernel density estimate is used to model the distribution of a word as a smoothed curve. Then, the overlap between every pair of curves is calculated. For each word, a link to a given number (here: ten, following the original paper) of words with the most similar distribution curve is created. The resulting network is a suitable representation of the broad thematic structure of the text.
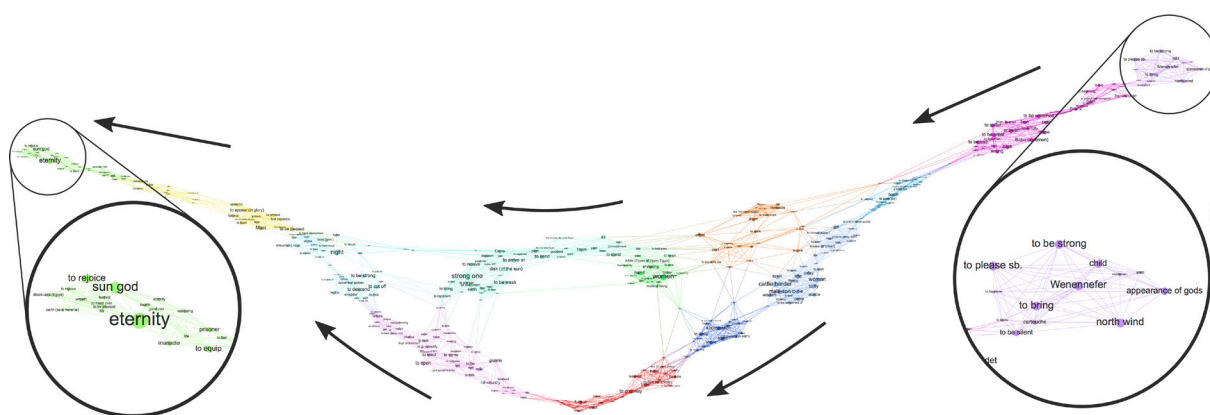
Figure 5: Text network of "the contendings of Horus and Seth".

The network represents the text of "the contendings of Horus and Seth".[24] It is a good example for the textplot method, because it is with 4820 lexical units the longest, coherent narrative in the TLA database and therefore provides enough data for testing the algorithm. The text deals with the battles between Horus and Seth for the succession to the throne of Osiris (see the remarks about the struggle between Horus and Seth above). The specific time of the contendings is a period during which the fighting has temporarily stopped and Seth and Horus have brought their case before the tribunal of the divine ennead. Throughout the story, Horus and Seth compete in several ways in order to find out who will be king.

This elongated network visualizes the process of the story very well (see figure 5):[25] The beginning of the story is a sort of a trial when Seth and Horus plead their cases, gods appear and the divine judges state their opinion. At the End of the story, the trial starts up again between Horus and Seth and finally, the situation is resolved when Horus is determined to be rightful king of Egypt. So beginning and end of the story are similarly structured.

But in the middle of the story something happens: The upper part represents the sections in which the gods are discussing who should be the next king and heir of the crown. The lower part of the network visualizes all the little sub-stories where the goddess Isis is involved and

---

22   Lietz (2007).

23   McClure (2014).

24   pChester Beatty I, recto (Dublin, Chester Beatty Library).

25   Based on TLA data of the text.

where she tries to manipulate the action-packed battles between Horus and Seth. The textplot algorithm is able to recognize that the story consists of these two different "plots", which are not that clearly structured in the actual story.

To sum up, constructing networks with textplot is a useful method for distant reading and for making visible the inner text structure.

9. Conclusion

Network models are an interesting approach to capture the relational nature of many of the phenomena that are of interest for humanities research. As a formal tool, they can be used to model anything that can be expressed as a set of nodes and edges. Picking the right network model and deciding what these nodes and edges are in a specific case has to be guided by the research question. In this paper, we presented a series of network modelling techniques that are suitable for studying ancient texts. Starting with social network analysis of literary characters, we showed how the semantic dimension of text, i.e., its content, could also be modelled as a network.

The findings presented here stem from research by the SeNeReKo project. Using ancient corpora as differing as the Indian *Mahābhārata* and Egyptian pyramid texts, we evaluated several network creation techniques. The tools we created to apply these methods to historical corpora are available as open source software.[26]

As can be expected, the methods described in this paper and similar ones proved to be particularly helpful in the case of large texts because firstly, the manual gathering of, e.g., the data used in the networks based on the *Mahābhārata* would have required a virtually unmanageable amount of human work; but more importantly, graphical representations of such data (as shown above) enable the scholar to grasp complex social and semantic structures at a glance that could not – or only very imperfect – be noticed by traditional reading. Our research taught us that graphs of that kind, as a rule, do not provide final answers by themselves, but trigger new questions and are excellent starting points for further research.

---

26   https://github.com/SeNeReKo.

10. References

Allen (2015): Allen, James P., The ancient Egyptian pyramid texts (Writings from the ancient world 38), Second edition, Atlanta, GA.

Allen (1916): Allen, Thomas George, Horus in the Pyramid Texts, Dissertation, University of Chicago.

Brett (2012): Brett, Megan R., "Topic Modeling: A Basic Introduction", Journal of Digital Humanities 2/1, http://journalofdigitalhumanities.org/2-1/topic-modeling-a-basic-introduction-by-megan-r-brett/.

Chang et al. (2009): Chang, Jonathan, Jordan Boyd-Graber und David M. Blei, "Connections between the lines: augmenting social networks with text", in: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '09), New York, 169–178.

Düring (2015): Düring, Marten, Verdeckte soziale Netzwerke im Nationalsozialismus. Berliner Hilfsnetzwerke für verfolgte Juden, Berlin.

Gramsch (2013): Gramsch, Robert, Das Reich als Netzwerk der Fürsten: politische Strukturen unter dem Doppelkönigtum Friedrichs II. und Heinrichs (VII.) 1225–1235, Ostfildern.

Hellwig (2010): Hellwig, Oliver, "Performance of a Lexical and POS Tagger for Sanskrit", in: Girish Jha (ed.): Sanskrit Computational Linguistics, Lecture Notes in Computer Science, Berlin/Heidelberg, 162–172.

Lietz (2007): Lietz, Haiko, "Mit neuen Methoden zu neuen Aussagen: Semantische Netzwerkanalyse am Beispiel der Europäischen Verfassung", http://www.haikolietz.de/docs/verfassung.pdf (7 March 2017).

McClure (2014): McClure, David, "(Mental) maps of texts", http://dclure.org/essays/mental-maps-of-texts/ (7 March 2017).

Meurer (2002): Meurer, Georg, Die Feinde des Königs in den Pyramidentexten (Orbis biblicus et orientalis 189), Freiburg, Schweiz.

Moretti (2011): Moretti, Franco, "Network Theory, Plot Analysis", New Left Review 68, 80–102.

Moretti (2013): Moretti, Franco, Distant Reading, London.

Paranyushkin (2011): Paranyushkin, Dmitry, "Identifying the Pathways for Meaning Circulation using Text Network Analysis", http://noduslabs.com/research/pathways-meaning-circulation-text-network-analysis/ (7 March 2017).

Winterer (2012): Winterer, Caroline, "Where is America in the Republic of Letters?", Modern Intellectual History 9/3, 597–623.

Authors[27]

Frederik Elwert, Dr.
Frederik Elwert, Dr.
Center for Religious Studies (CERES)
Ruhr University Bochum
Universitätsstr. 90a
D-44789 Bochum

Email: frederik.elwert@rub.de

Simone Gerhards M.A.
Institute for Ancient Studies
Egyptology, FB 07
Johannes Gutenberg University Mainz
D-55099 Mainz

Email: gerhards@uni-mainz.de

Sven Sellmer, Dr. habil.
Chair of Oriental Studies
Adam Mickiewicz University
ul. 28 czerwca 1956 nr 198
61–485 Poznań
POLAND

Email: sven@amu.edu.pl.

Social Networks in Late Hellenistic Northern Etruria: From a multicultural society to a society of partial identities

Raffaella Da Vela

Abstract: This contribution concerns the perception of local cultural identities in Hellenistic Northern Etruria. I aim to answer the following question: how did the evolution of the economic and politic relationships between settlements condition the perception and the expression of the local identities? During the Romanization, the Region of my case study presents a complexity of patterns, as result of the interaction of local needs and backgrounds with the global process of institutional and economic unification of the Mediterranean following. The Social Network Analysis (SNA) has been chosen as methodology to approach this complexity. I propose some parameters to detect local identity markers, then I present an integrated method to relate their development to the evolution of the relational networks between the communities.

"[…] Aren't all these who are speaking Galileans? 8 Then how is it that each of us hears them in our native language? 9 Parthians, Medes and Elamites; residents of Mesopotamia, Judea and Cappadocia, Pontus and Asia, 10 Phrygia and Pamphylia, Egypt and the parts of Libya near Cyrene; visitors from Rome 11 (both Jews and converts to Judaism); Cretans and Arabs […]"[1]

1. Research questions and methodology

This contribution introduces my research project[2] on the transformation of local cultural identities during the Romanization.[3] Because the project is still in progress, I will present the research questions, the methodological approach and some examples within a selected case study to illustrate the proposed procedure.

---

1   Bible. Acts 2,7–11 . NIV.

2   My ongoing post-doctoral research project started in 2015 and has so far been partially financed by the fellowship at the DFG-Training Research Group 1878 "Archaeology of Pre-Modern Economies" at the University of Bonn and Cologne. The project was presented and discussed during the meetings of the research training group itself, as well as at the 5th Seminar of the Mommsen-Gesellschaft "Local Responses to Global Change in the Ancient World", moderated by Hans Beck. The methodological applications have been discussed in Berlin, during the lecture series "Digital Classicist" of Topoi and DAI, and in some meetings, among others: The CAA-2015 in Siena, the "Digital Humanities" Seminar in Rennes; the workshop "Méthodes quantitatives et outils numériques appliqués à l'Antiquité et au haut Moyen Âge" in Grenoble and the TRAC 2016 in Rome in the panel session organized by Francesca Diosono and Dominik Maschek: "Beyond Hybridity and Code-Switching. New Approaches to Archaeology of Late Hellenistic Rome, Italy and the wider Mediterranean". These presentations have not been published, but I wish to thank all participants for the stimulating discussions. English text revised by Henry Heitmann-Gordon.

3   The term Romanization here denotes the complex process of adoption of Roman political institutions and Latin as official language in Etruria cf. Terrenato (1998a), 54. 94.

The aim of my research is to detect patterns of interaction between the perception of identities within local communities and the transformation of the relational networks between them. Generally speaking, the transformation of the relational network between local communities is governed by both local factors and interaction dynamics and by external forces and factors. This is particularly pertinent during periods of high personal mobility and frequent cultural contacts. The project further aims to understand whether changes in the perception of local cultural identities could be due to the impact of external forces and factors, such as Roman military occupation, the building of colonies and new infrastructures as well as the reorganization and unification of the commercial sea-routes after the fall of Carthage and the Roman conquests in Greece. How did the evolution of the economic and political relationships between settlements condition the perception and the expression of the local identities and how did global events[4] and dynamics affect the perception of cultural identities in the local communities? To approach these research questions a *glocal*[5] perspective has been adopted: the analysis of the local perception of identities within the communities has been related to the evolution of the network of the connections between them. The *local* dimension (local cultural identities) is detected by identifying specific identity markers within the archaeological dataset of the local communities. The global dimension (transformation in the networks of the communities[6]) is analysed through Social Network Analysis (SNA). The networks concern the geographic connection between communities, the social connections brought about by personal mobility, and the use of shared assemblages of objects (material exchange and agreements) and ideas (cultural transmission, political and religious affiliation). The unification of the local and the global dimension in a glocal perspective is obtained by combining the results gleaned from these two stands of analysis. This method is here put to the test in a concrete and specific case study: the Romanization of Northern Etruria.

2. Cultural Identities

The opening quotation from the Acts of the Apostles gives an intuitive impression of what I mean by the perception of cultural identities, here in the multicultural ancient community of 1st century CE Jerusalem.[7] The perception of different cultural identities depends on the use of different languages, on the geographic provenience, as well as on political and/or religious affiliations. This complexity makes it necessary to define more precisely the use of the term "cultural identity" in this contribution.

The phrase "local cultural identities" here denotes the collective identities of the local communities on a cultural level, meaning the system of values, codes and behaviors they share as a group.[8] These values and behaviors are considered to be specific to and characteristic of any given community, and used in a diacritical function respect other communities. The term 'identities' is deliberately used in the plural, implying that different cultural identities

---

4 For an application of globalization theory to the material culture of Hellenistic contexts cf. Hoo (2015) 37 f. For the interpretation of the process of Romanization as a global event/globalization factor: cf. Terrenato (1998); and Versluys (2014) [[DOI:10.1017/S1380203814000038]].

5 For the possibility to apply this terminology to archaeological contexts cf. Hoo (2015) nt. 30, 40. Pitts/Versluys (2015).

6 For the definition of the network of relations ("interconnectedness") of the human societies as global level cf. Attema (2010) 8.

7 For the chronology of this source: cf. Abbolito Simonetti (1995) 13. 55 f.

8 Cf. Hall (2012) 351.

can coexist within a group. This is the case, e.g., in multicultural communities,[9] where the collective identity is the result of a deal struck between a plurality of sub-identities present in the communities. The local cultural identities further present an internal complexity, linked to the social segmentation of the community and to the possibility of fragmented individual identities[10] or multiple identities.[11] The complexity of these layers and of their stratification is not fully assessable through the archaeological and epigraphic sources.[12] Individual identities in particular are but sporadically documented. For these reasons I decided to use the collective identities of the local communities as the smallest unit of my research, which can be determined by identifying the most frequently shared values and behaviors, expressed, for instance, in funerary ritual, in imagery and in the choices related to material culture.[13]

3. Introducing the case study: Hellenistic Northern Etruria

The Hellenistic Northern Etruscan communities (fig. 1) have been chosen here, because the related data set shows some very interesting peculiarities. In the Hellenistic period, this region presents a complex mesh of different patterns. The increase in personal mobility generated more opportunities for cultural contacts. The local communities adopted new customs and lifestyles without giving up their Etruscan cultural background. These "multicultural"[14] societies switch slowly, at the end of the 2nd century BCE, to societies of partial identities,[15] where the Etruscan language and traditions became confined to the private sphere, while the Latin language and Roman institutions dominated public life.[16] These preconditions (complexity, rich dataset of archaeological and epigraphic sources related to personal mobility, attested changes of behavior), within a restricted geographic area, make the case study ideally suited to detecting the evolution of the perception and construction of local collective identities.

---

9   To describe these communities the neutral terms of "composite identities" cf. Wallace-Hadrill (2011) or "Mixed Identities" cf. D'Ercole (2011) are often adopted.

10   On the fragmentation of personal identities as a consequence of globalization cf. Bradley (2007).

11   Multiple Identity in archaeological context cf. Blake (1999), 35. For a review of the studies on the subject cf. Luhrmann (2001).

12   Especially because the tombs with grave goods represent a limited group within the real societies cf. Berrendonner (2007).

13   For the role played by material culture in the construction of collective identity (an image that a group constructs to allow the identification of its members) cf. Versluys (2013) 431 with further bibliography. On the Italic Identities during the Romanization and the possibility to individuate assertion of local distinctiveness: Stek (2013) 348–352.

14   Eriksen (2015) 28 f.

15   For the social function of the partial identities (Teilidentitäten) cf. Straus (2002) 263–169.

16   Benelli (2009); Benelli (2012); Maggiani (2014).
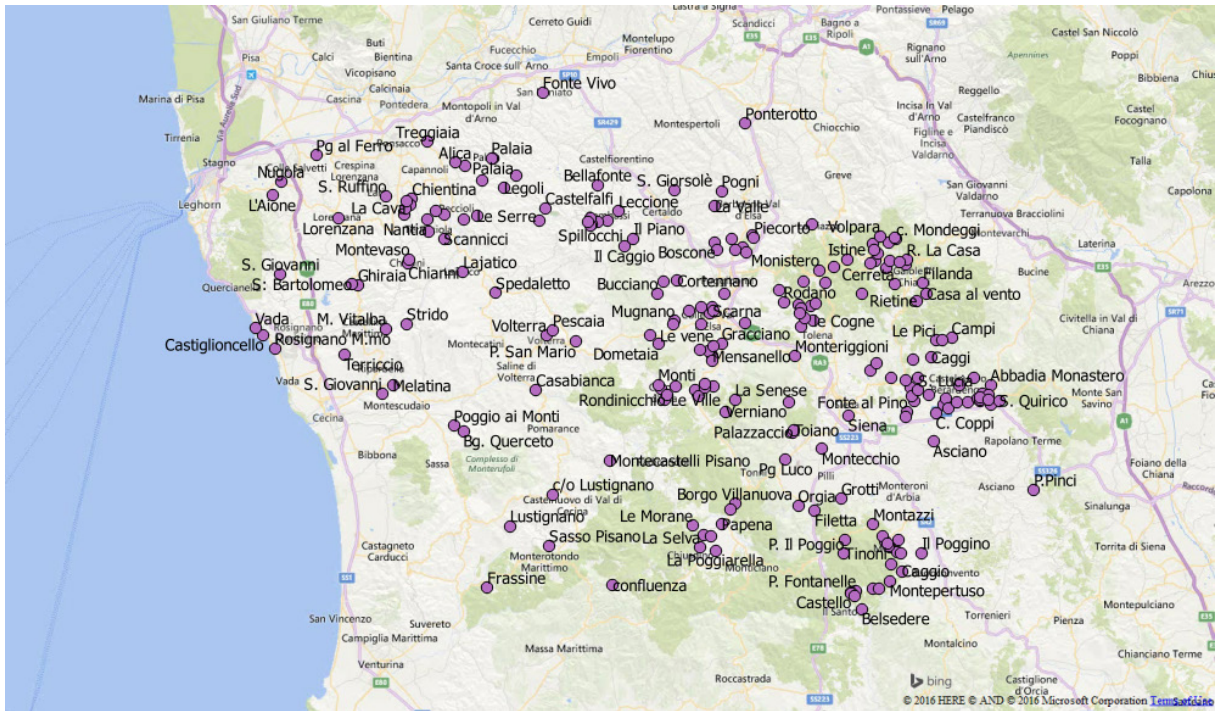
Figure 1: The attested archaeological places in the territory of Volterra during the Punic Wars (QGIS 2.12.3, using Bing Maps).

Naturally, the case study also has its limits. As the study of ancient society can rarely draw upon complete datasets, it may be helpful to present the principal gaps in the extant sources and to discuss how one can accommodate the challenges of an incomplete dataset and to what degree this affects the analysis. The archaeological sources (geographic position, buildings, material culture) are related to settlements and to necropoleis. The position of the settlements in space and their chronology are mapped through survey archaeology.[17] The surveys allow for excellent spatial analysis, to understand the occupation of the landscape and its modalities (dispersed settlements, cities, etc.). The chronology is frequently determined simply by collecting pottery shards and other materials on the surface layer. The consequence is a partial knowledge of the layers, which were obliterated by the most recent phase of activity. Nevertheless, it seems possible to detect some trends and patterns of occupation in Hellenistic northern Etruria, because the occupation of the landscape is well mapped for the whole Hellenistic period and this coverage allows one to analyze the connectivity of the settlements. A complete dataset for the material culture within the settlements is not yet available and exists only for the largest cities. The necropoleis, however, offer a very rich documentation of the material culture. Their geographic relation to some settlements frequently allows one to reconstruct also the material culture of the communities. In some cases, settlements are still unknown but the material culture of the local community and their geographic positions are already documented by their necropoleis. For this reason, the smallest unit of research is the local community and not the settlement. Unfortunately, analysis of the material culture of the necropoleis is hampered by the fact that many of them were excavated between the 18th and the beginning of the 20th century; consequently, a part of the contextual data was not collected in conformity with the current scientific standards. The history of the research biases the data set, especially regarding the proportion of the collected materials. For this reason, no quantitative analysis of the data will be conducted, at least for the moment; only the presence/absence of some classes

---

17   Cherici (1987); Cherici (1992); Torelli (1992); Valenti (1995); Cambi (1996); Valenti (1999); Campana (2001); Nardini (2001); Felici (2004); Botarelli (2005); Cenni (2007); Paolucci (2008); Acconcia (2012).

of materials is taken into consideration. Because the interest is here mainly in assessing the exchange of information and the cultural and economic contacts between communities, it will be probably necessary to choose a threshold to establish the presence/absence of some aspects of the material culture in the local communities. This threshold allows one to weight the impact of the quantity on the exchange of information, assuming that repeated and more numerous contacts are able to share more or to consolidate the shared information[18].

A large amount of data is finally available from epigraphic sources; in particular, the data related to the marriage politics and to the circulation of people between communities offer an appropriate set for network analysis. The frequent use of matronymics and patronymics, as well as gamonymics[19] in the funerary inscriptions, often allows one to trace the arrangements between families of the upper classes of the local communities, especially in the early and middle Hellenistic period, as well as the migration of people from other Italic regions, as well as Northern Africa and Greece.

These data are very accurate, but cannot illuminate the entire area of the case study: They are available exclusively in a land-locked sub-region, the area around Chiusi, Perugia and Volterra. In the coastal communities there are far fewer inscriptions, because of the differences in funerary customs: inscriptions were set up here only in exceptional cases. The consequence is a sub-regional limitation of the network of mobility. Nevertheless it seems useful to apply SNA in the sub-regions, where data are available.

4. The local dimension: Identifying identity markers

In order to be able to detect local cultural identities one needs to select specific parameters that can trace markers of these identities. Doing so is quite difficult in the Hellenistic period, because of the wide diffusion of koinè elements in the material culture: Once the local productions have been identified, the presence of objects or ideas which do not belong to these productions (outsiders) can be explained as consequences of economic transaction or of cultural contact. The use of objects is frequently not enough to identify the cultural identity of the owners of the objects, nor of their community.

I therefore developed four complex parameters in order to distinguish markers of identity: 1. the association of objects in context; 2. the existence of diacritical patterns in the material culture; 3. the assignment of the semantic value of objects in the social and religious rites of the communities; 4. and last, but not least, the shift in the use of language (fig. 2).

---

18   The repetition of the message as a factor able to effect the impact of the communication has been studied in Social Psychology: cf. Chaiken/Trope (1999) 307 f.

19   In form of the husband's or wife's name.
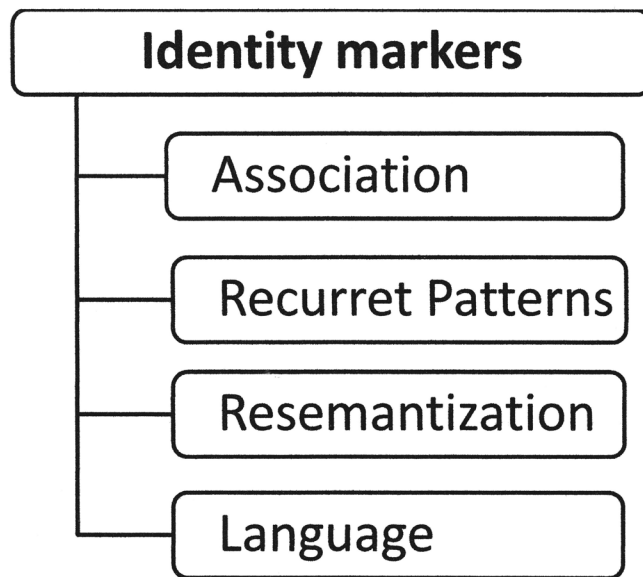
Figure 2: Parameters to distinguish Identity Markers.

The first parameter, the association within the grave goods, finds its application at micro level, in the contextual association in single tombs or in groups of tombs in the necropoleis of local communities.

Figure 3: Campanian Dinos and cinerary urn in the Tomb of Fastia Aemilia Praesenti in Perugia (Cipollone (2009) 155).

The second parameter, the individuation of diacritical patterns in the material culture, allows one to cluster specific cultural aspects of local communities even during the period of homologation of the grave goods and productions due to the Hellenistic koinè.[20] The individuation of local peculiarity and of general patterns is useful as it helps specify the degree to which local communities differ from one another and, at the same time, to individuate single divergent cases within a local community: The number of divergent cases is higher in communities with multicultural or composite identities (Inclusion Communities), lower in communities with univocal or homogeneous cultural identity (Exclusion Communities).[21]

The third parameter used to individuate markers of identity is the change of meaning. A change of meaning is constituted by the semantic editing of an object or class of objects, which has been transmitted from one cultural background to another. Changes of meaning in the funerary context often serve to distance the object's original meaning and function and to place it in a symbolic relation with a new constructed cultural identity.[22]

The fourth parameter to detect markers of local cultural identity is the choice of language,[23] which is one of the parameters most commonly adopted to identify ancient culture groups.[24]

5. Some examples of identity markers in Hellenistic Northern Etruria

The following examples show the detection of the identity markers by means of the proposed parameters, within the selected case study.

The first parameter, the association of grave-goods, has been applied in a tomb of the necropolis of Santa Caterina in Perugia (fig. 3).[25] Here in the family tomb of the *gens Praesnti*, a cinerary urn of the noble woman *Hastia Aemilia Praesenti* was deposed. The cinerary case was decorated with a relief showing a mythological scene: The conversation between Odysseus and Circe, set in Campania. The assemblage of grave goods included earrings, imitations of south-Italian models, and a bronze mirror. These objects can be dated to the 1st half of the 1st century BCE. However, the tomb contained also a second cinerary urn, dated to the 5th century BCE, which can be attributed to an ancestor of *Hastia Aemilia*. This vase, a dinos produced in Campania, probably indicates that the family was of Campanian origin. Therefore, its translation into the new grave may have had a strong symbolic significance. In this case-study the association between the old Campanian vase, the iconography of the recent cinerary case with its allusion to the Campanian landscape, as well as the earrings which imitate a south-Italian model, constitute a semantic association with a strong allusion to a Campanian identity and can be assumed as a document of a complex identity in a family, whose members are writing in Latin, using an Etruscan cinerary urn and carrying an Etruscan name (*Praesnti*).[26]

---

20   Cf. Stefanie Martin-Kilcher's research on the Late Republican necropolis of the Maggiore Lake: Martin-Kilcher (1998), as well as Milinda Hoo's research on Hellenistic Ai-Khanum (English version forthcoming): Hoo (2015).

21   Cf. Bruhn (2011); also Allman (2013).

22   Cf. Da Vela (2016) with bibliography.

23   The central importance of language in constructing cultural identity has been analysed in linguistic anthropology: cf. Bucholtz/Hall (2004) 369–394.

24   Wallace-Hadrill (2011) 422; D'Ercole (2011) 438–442.

25   Reconstruction of the grave goods association cf. Benassai (2002).

26   Iozzo (2011).

The second parameter, the detection of specific patterns in the funerary rituals, can for instance be applied to the coastal necropolis of Castiglioncello.[27] Here it has been possible to assume, for the Hellenistic period, a presence of Ligurian refugees and migrants, marked by their characteristic grave goods of weapons and La-Tène fibulae (fig. 4).[28]

Figure 4: Tomb 7/97 of the necropolis of Castiglioncello with Ligurian markers of identity (Aut MIBACT/Soprintendenza Pisa, prot. 9606; 28.13.10/26).

27 Gambogi/Palladino (1999).

28 Maggiani (2013) 241–243.

The third parameter used to detect markers of identity, resemantization can be exemplified in the use of commercial amphorae as cinerary urns in the necropoleis of the Apuanic Ligurian and in the Northern Etruscan necropoleis situated in contact zones with Liguria. The co-presence of different burial customs in these Northern Etruscan centers is an example of cohabitation of differently connoted groups within the local communities. In the late apuanic Ligurian culture, between the end of the 3rd century and the beginning of the 2nd century BCE, several burials re-used cut Greco-italic *amphorae* as containers of the cinerary olla, while in Etruria, Greco-italic amphorae were deposed as an element of the banquet set, relating to the consumption of wine in the symposium (fig. 5).

Figure 5: Apuanic Ligurian burials showing amphorae used as containers of the cinerary vase. Museo Archeologico Nazionale di Villa Guinigi, Lucca, Tomb from Ponticello di Marlia, 170–130 BCE (Da Vela (2016) fig. 7, 38).

The fourth parameter is the use of the language. In the selected case study, not just the language is an expression of cultural identity, but the choice of alphabet as well.[29] The alphabet adopted can be a local Etruscan alphabet, a Latin alphabet or a mixture of the two. The opportunity to use parameter is exemplified particularly well by the necropolis of Balena, near San Casciano dei Bagni, in the Province of Siena,[30] which was located in a frontier zone between the territories of the Etruscan cities of Chiusi and Orvieto. Here the integration of new elements by the local population is well documented in the late Hellenistic loculi-tombs of the necropolis. The grave goods were deposed in burial recesses, carved into the local sandstone from a shared dromos. Each burial recess was closed by a funerary tile, on which a funerary inscription was incised (fig. 6). In every one of the 6 recently excavated and published dromoi, the inscriptions show

---

29   For the value of the choice of the alphabet in function of the cultural identity cf. Prost (2002) 319–324.

30   Salvini (2014).

use of both the Etruscan and Latin language and alphabet, as well as mixtures of the two, all in different combinations. It emerges that the choice of alphabet and language expressed the intention of the family of the deceased to define their cultural identity as Etruscan or Roman. The pattern shows a high degree of variability over time, as attested by the grave goods in the burial recesses, and as regards the origins of the family, attested by the name. The typology of the grave goods is standardized (mainly unguentaria, lagynoi and strigiles). In the burial recesses three types of cinerary urns were deposed: Etruscan urns of Chiusi, with a lid shaped like a banqueter or sleeper, bell urns, exclusive to Chiusi and the surrounding area, and ollae, covered with a dish or bowl, reversed on the top. When the grave goods and the inscribed tile were preserved in association, it has been possible to determine a concordance between the type of the urn, the origin of the deceased and the linguistic choices. So e.g. the traditional form of the cinerary case was normally associated with local families, which are well attested in the main center of Chiusi, and had inscriptions written in Etruscan language and alphabet, while cinerary ollae were associated with Umbrian or Italic names and inscriptions in Latin language. In this case the correlation between the shift in the use of the Latin over time and the presence of new elements in the population is very remarkable. The necropolis' main period of use is between 175 and 40 BCE. The oldest grave goods relate to people with Italic names and were sealed by a tile bearing inscriptions in Etruscan alphabet and Latin language. The new elements of the population seem to have sought to adapt to the local identities and it is also possible that they let local artisans or intellectuals write their names, which they themselves are able to spell, but not write. In the following years these families, as well as the families of Etruscan origin, begin to write in the Latin alphabet.[31] Another important piece of information, which can be deduced from this social context is that the buried people weren't relatives or members of the same family. This is a big break in the Etruscan funerary habitus. The study of the architecture and form of the grave and their confrontations with similar contexts around Lake Trasimene has allowed us to trace an evolution in the agricultural communities of central Etruria after the Punic Wars and to attribute the necropolis to recently founded rural villages, where local elements and migrants were living together and building a network of new local communities which initially possessed a multicultural identity. This identity was slowly transformed into a new, shared and uniform identity, perceivable in the adoption of the Latin language.

Figure 6: Inscribed funerary tile in the Necropolis of Balena (San Casciano dei Bagni, Siena) (from: Salvini (2014) fig. 3, Tav. VI, 202).

---

31 Maggiani (2013) 56.

6. The global dimension: Constructing a multilevel network

The methodology chosen for this part of the study is Social Network Analysis (SNA), as it allows for the analysis of complex systems without losing details relating to their individual components.[32] The project contemplates the construction of a longitudinal multilevel network to analyze different kinds of connections between local communities.[33] 'Longitudinal' here means that the network is articulated in different time segments, which are compared to analyze the evolution of the relationships over time.[34] The smallest units of my network (nodes) are the local communities. The agents of the connections between communities are people: I assume that, on average, people living in the same place have a high frequency of contacts and consequently a high level of shared information, wishes and points of views. The connections between communities, established by the people living in them, are the ties or edges of the network. Many different human interactions and forms of contact allow information to be shared, with different levels of intensity, ranging from the superficial to the deepest level and are able to influence the perception of identity. I selected 7 different kinds of relationships, or relational ties, which, in my opinion, influence the question of the identity. I build a network inside my geographic area for each one of these ties (fig. 7): 1- connectivity, i.e. the possibility to move in the geographical space between communities; 2- mobility as the actual movement of people between two communities; 3- exchange of material goods, via trade, gift exchange, etc.; 4- exchange of immaterial goods as cultural transmission of ideas, lifestyles, images; 5- political association between communities, documented through literary sources; 6- interaction between people living in different communities, visible in agreements to buy and sell land; 7- religious behavior, i.e. shared rituals, cults and cult places.
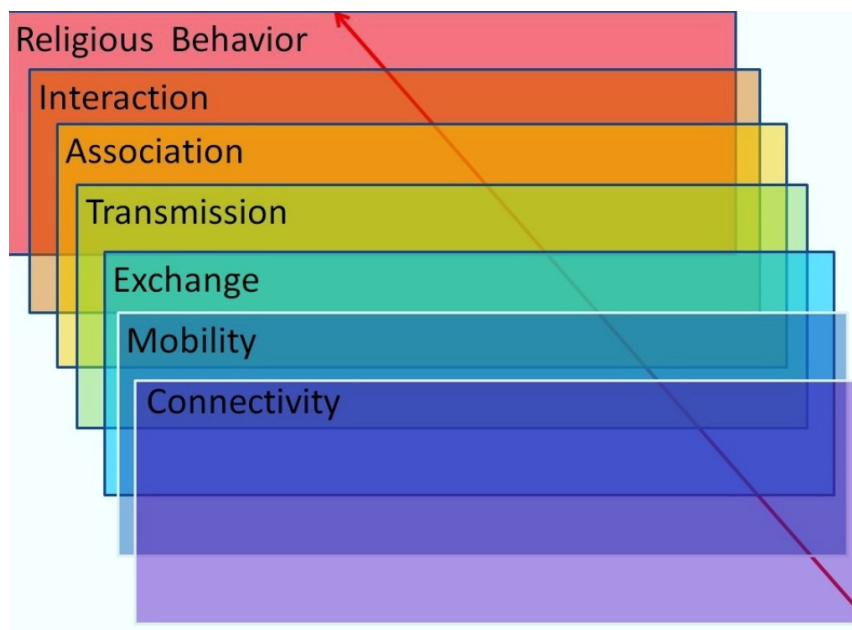

Figure 7: Plan of a Multilevel network.

---

32   Cf. Knappet (2013) 3–14.

33   Multi-level intended here as multi-tie network: cf. Mol (2014) 253 f.

34   For the application of the longitudinal network in archaeology: cf. Mizoguchi (2013).

7. Some examples of the network of connectivity between Northern Etruscan local communities

I present some graphs of the SNA of a part of my dataset, for the tie connectivity.[35] Networks of connectivity can be constructed on a number of different criteria. I have confronted three of these: a) empirical b) k-nearest neighbors 3) linear distance with a threshold of proximity.

a) The connections are established when people of a community A can reach a community B without having to cross a third community C. The presence/absence matrix is symmetrical and the graph undirected, because I assumed a reciprocate probability that people would travel from A to B and from B to A.[36] The routes are empirically established, through personal knowledge of the topography and of the possible routes. This method could be however biased by different factors, for instance by difference in the perception of practicable/impracticable routes in the past (fig. 8).
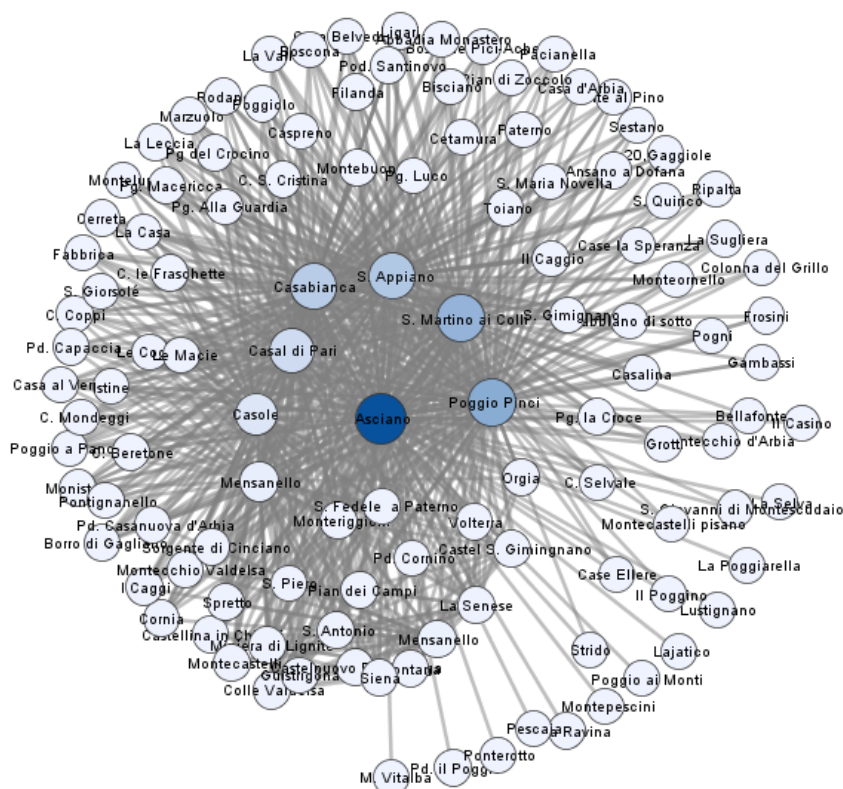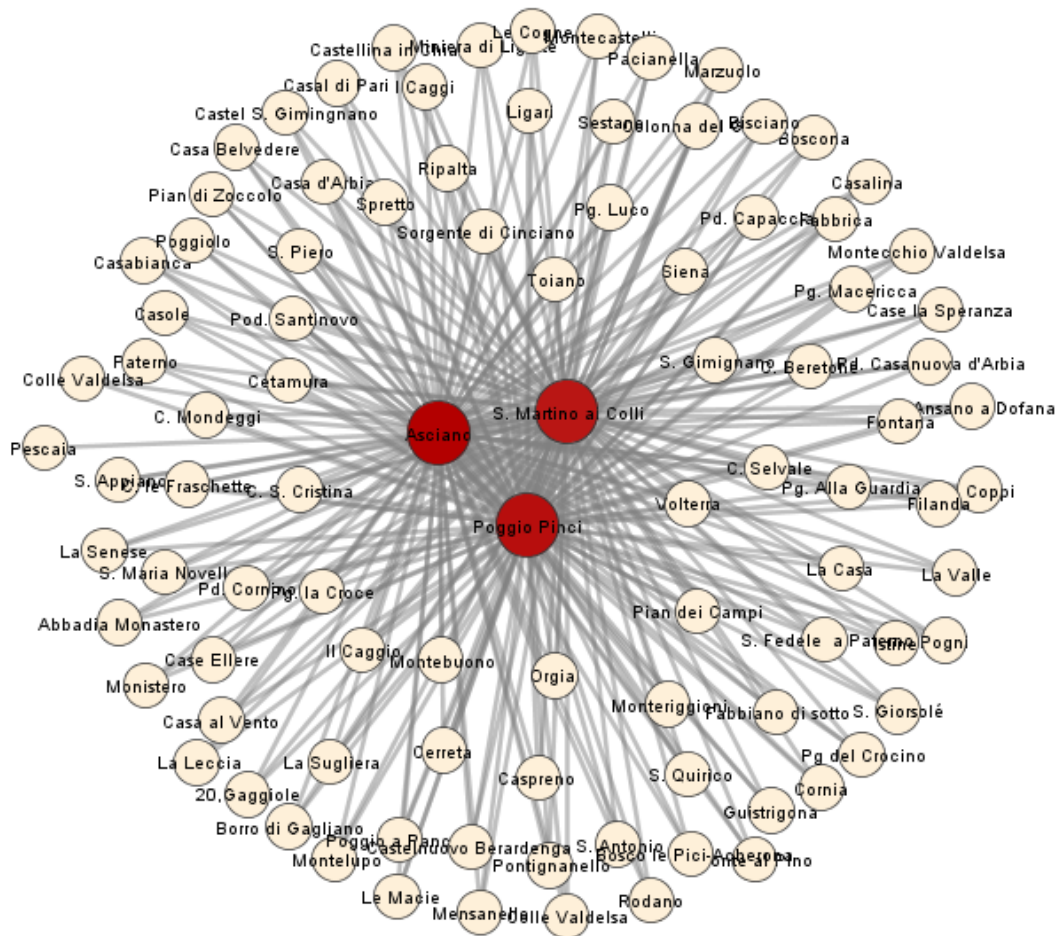

Figure 8: Graph of the connectivity, criterion (a) applied to a partial network in the territory of Volterra (Processed with the Software Visone 2.9.2).

---

35   The complete dataset is composed of ca. 500 communities, distinguished by elaborating the survey-data for three time periods (350–264 BCE; 264–146 BCE; 146–80 BCE). I cannot here present the mobility network, because the database is still in development: I have collected a database of 2300 Hellenistic funerary inscriptions for the territory of Chiusi, but the presentation of the network for this area is still in progress. The prosopographic network seems one of the most promising for my case study and has been productively used in other research (as these of Bastian Still for his PhD at the University of Leiden); nevertheless the high complexity and the level of documentation of the intermarriages between communities required an accurate evaluation of the reliability of the sources: each inscription has to be divided into components in order to understand the direction of the movement of people between communities, the chronology of the inscription has to be determined, and finally the possibility weighed that the migration of people could have preceded the real appearance of the name in the grave. The preexisting studies of Enrico Benelli: Benelli (2009); Benelli (2012) and Adriano Maggiani: Maggiani (2014) are very valuable in that they help me collect and evaluate data for the network set.

36   For the moment, I did not consider any hierarchy between communities in the geographic network. After this first elaboration, I will try to insert an "attractiveness index" of the biggest communities, taking in consideration the dimension and the economic role of them.

b) The second method is based on the choice of a k-number of nearest neighbours. I have chosen to select the three nearest.[37] A is connected with only three communities B, C, and D, which are those closest to community A. The distance is established on the linear proximity. The presence/absence matrix is symmetrical and the graph undirected (fig. 9).

Figure 9: Graph of the connectivity, criterion (b) applied to a partial network in the territory of Volterra (Processed with the Software Visone 2.9.2).

---

37   Nearest neighbor cf. River/Evans (2013) 107.

c) I inserted in the matrix the air-line distances between communities and I have arbitrary chosen a maximum distance (10 km) which allows one to consider two communities connected:[38] A is connected with all communities contained in a circle with a radius of 10 km and centered on A. The presence/absence matrix is symmetrical and the graph undirected (fig. 10).
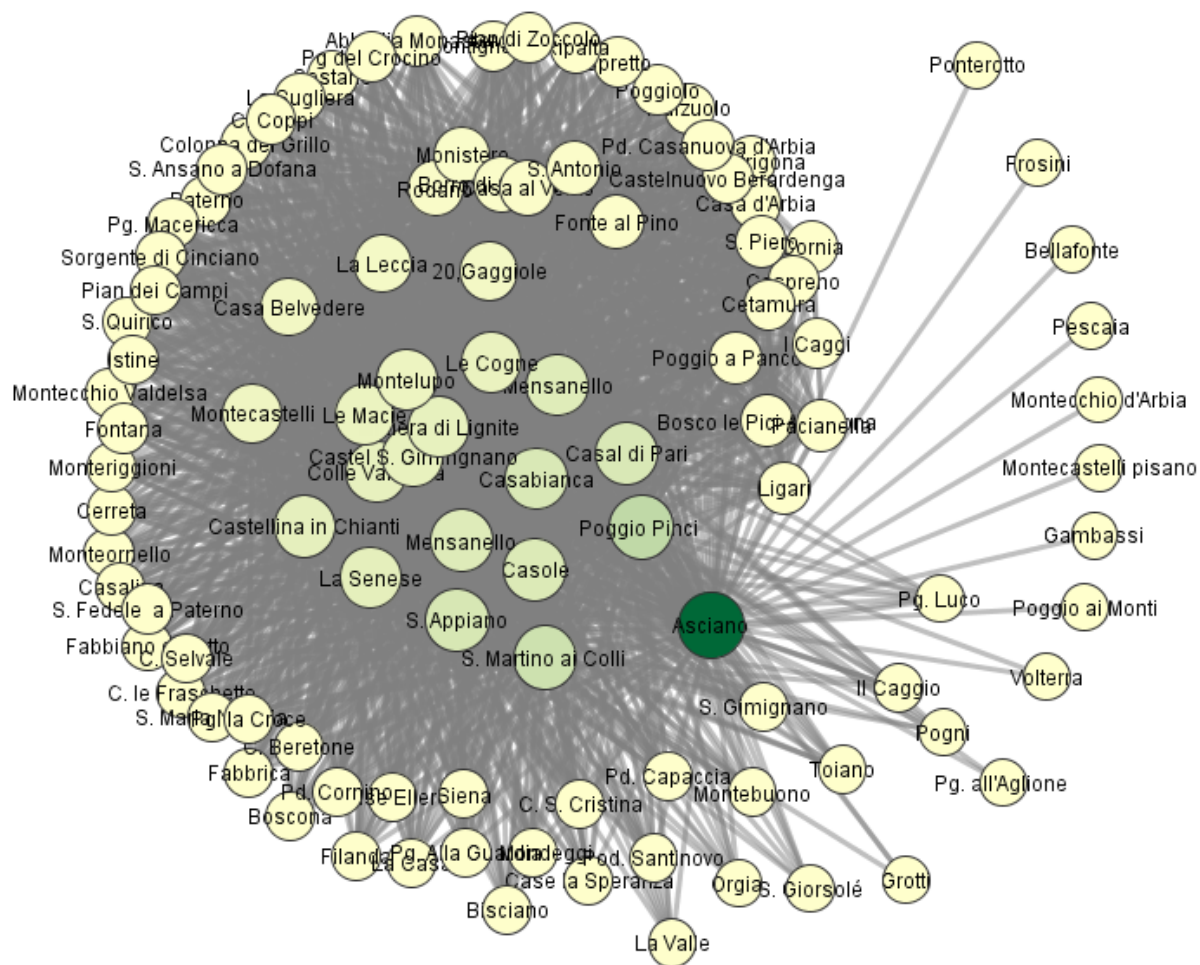
Figure 10: Graph of the connectivity, criterion (c) applied to a partial network in the territory of Volterra (Processed with the Software Visone 2.9.2).

Some observations can be advanced by confronting these three different criteria. The structures of the whole networks obtained with the 1st (fig. 8) and 3rd (fig. 10) criterion are similar, while the network dependent on the 2nd criterion, gives a different structure (fig. 9). The comparison of the ranking of the centrality measurements shows a higher variability in the detection of the role of the communities in the 3rd network (Tab. 1). In conclusion, the 1st and the 3rd criteria are more suitable for the aim of the project, both allowing to underline the role and the position of the communities in the network. The 3rd criterion is furthermore more appropriate for the conditions of the dataset, because in absence of conflict, the communication between two connected communities is not hindered by the presence of other communities on the path between them.

---

38   For the Geographic Threshold Criterion (Relative Neighborhood Network) cf. Jiménez-Badillo (2012).

Site Name	1-clos. %	2-clos. %	3-clos. %	1-betw.%	2-betw.%	3-betw.%
Asciano	38.541	35.155	30.48	1.596	1.992	1.515
Poggio Pinci	16.883	32.96	7.523	1.446	1.971	1.383
S. Martino ai Colli	14.992	31.885	5.869	1.413	1.951	1.383
S. Appiano	10.196	0	4.739	1.332	1.012	1.371
Casabian-ca	9.156	0	4.457	1.303	1.012	1.348
Casal di Pari	5.489	0	4.457	1.186	1.012	1.337
Casole	2.928	0	4.1	1.082	1.012	1.361
Mensanel-lo	1.099	0	3.855	0.979	1.012	1.348
La Senese	0.143	0	3.778	0.903	1.012	1.315

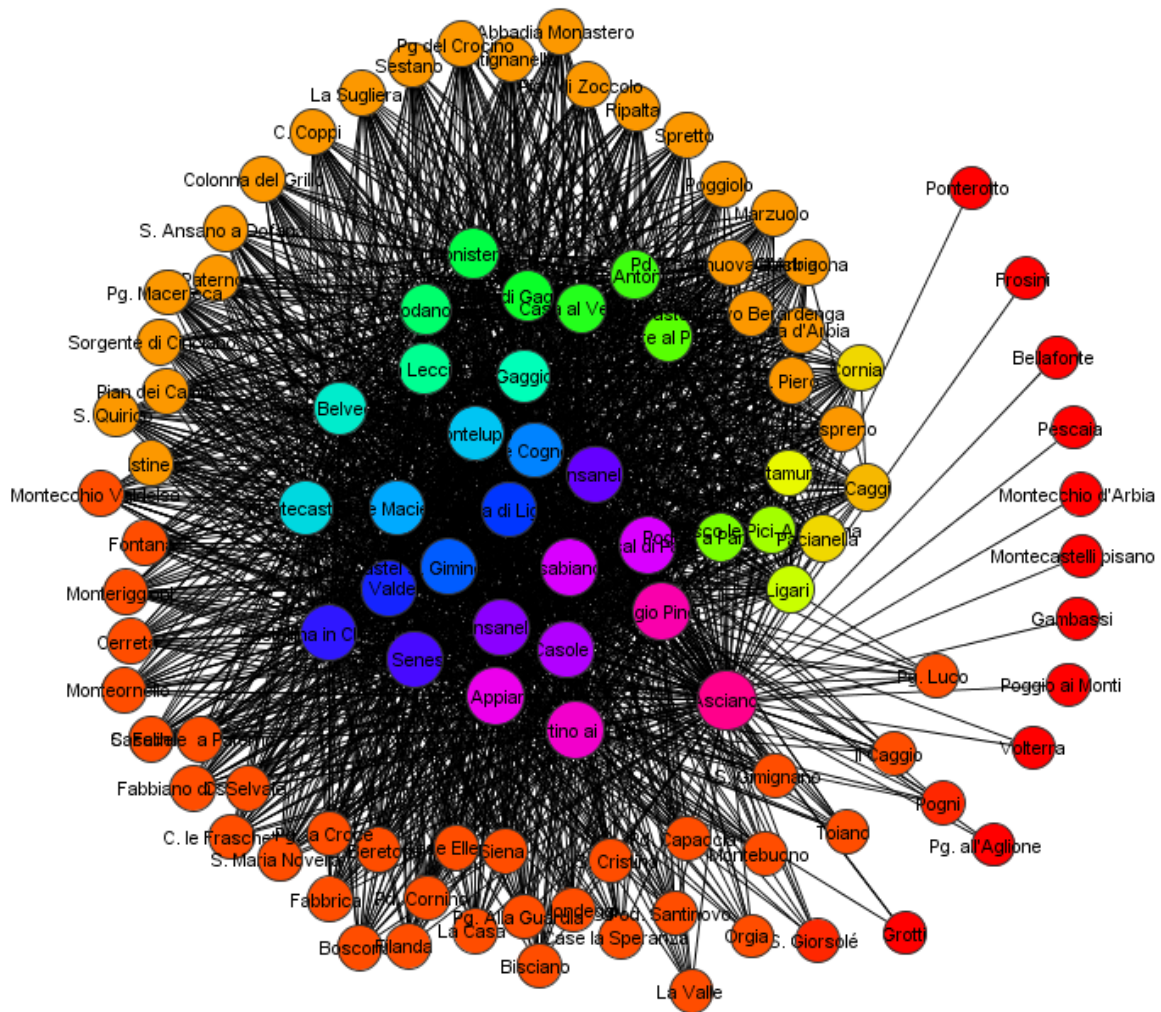## 8. How to relate the change in the perception of the identities with the change in the network?

Since the elaboration of the network and the detection and collection of the identity-markers is still in progress, I am, for now, presenting only a number of methodological propositions as to how one might relate the transformation of the connections between settlements with changes in the perception of cultural identities within them.

I propose to apply two centrality measurements of the role of the nodes in the network: the closeness centrality, which shows the role of a settlement in receiving and transmitting information (fig. 11) and the betweenness centrality, which shows the role of a community in brokering information between the other communities (fig. 12). These measurements are dependent on the number and weight (strength) of their ties.

Figure 11: Graph of the connectivity of the 2nd chronological Phase. Closeness Centrality (Processed with the Software Visone 2.9.2).

Figure 12: Graph of the connectivity of the 2nd chronological Phase. Betweenness Centrality (Processed with the Software Visone 2.9.2).

These measurements are more accurate in the network obtained with the 3rd criterion (figg. 10–12), which presents an higher number of connections.

The transformation of the role of the communities is then empirically compared with the transformation of the markers of the cultural identity within them (fig. 13).

This comparison allows one to detect recurrent patterns of interaction, e.g. if a settlement's greater capacity to receive and transmit information of correlates with evidence of cohesion or fragmentation in the perception of the cultural identities of the inhabitants. Since the individuation of recurrent patterns requires a complete analysis of the data, it will not be possible to answer these research questions before the research project concludes.
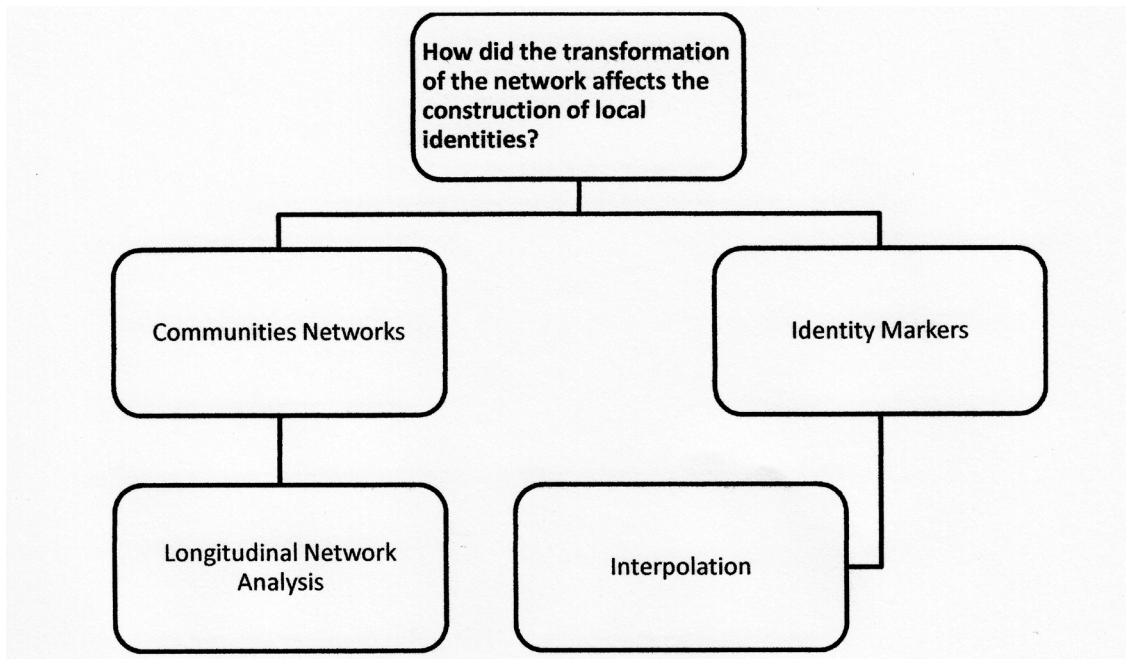
Figure 13: Diagram of intersection of network analysis and identity markers.

In a perspective of exploratory network analysis,[39] it could be possible to compare the changes within the network of the Etruscan communities to the evolution of many random networks, generated on specific parameters, to avoid the bias of expectation in the analysis of the coincidence between transformations of the real network and evolution of their properties. I currently do not expect to incorporate such experiments into my research. Instead, I adopt as an element of control the recurrences of similar patterns of reactions or responses to the transformation of the network of the identities (properties of the nodes) within the local communities (nodes).

9. Synthesis

My aim is to relate the impact of the evolution of the social network of the local communities to the perception of cultural identity within them. In the specific case study of Hellenistic Northern Etruria I aim to understand and describe whether and how the dynamic of Romanization in the social network of the local communities was able to affect the perception of the cultural identity of their components. In particular, I wish to discover how the transformation of multicultural communities into communities of partial identities works. To answer this question I propose an integration of two different methodologies: the detection of identity markers of the local communities by applying a set of parameters to archaeological and epigraphic sources, and the description and analysis of social networks of local communities through Social Network Analysis. The diachronic variations of the role of the local communities within the network are then related to the markers of identity collected within them. The identification of recurrent patterns of response by the local cultural identities to the evolutions of their network offers a new way of reading the impact of global political and economic challenges on the life of the local communities.

---

39   Knappett (2013) 8.

10. Bibliography

Acconcia (2012): V. Acconcia, "Paesaggi etruschi in terra di Siena. L'agro tra Volterra e Chiusi dall'età del Ferro all'età romana. British Archaeological Report, BARInternational 2422, Oxford.

Allman (2013): D. Allman, "The Sociology of Social Inclusion", Open Sage 3.1, 1–16. DOI:10.1177/2158244012471957 (08.01.2013)

Attema (2010): P. Attema, "Regional Pathways to Complexity", in: Peter Attema, Gert J. Burgers, Martijn van Leusen (Eds.) Settlement and Land-Use Dynamics in Early Italy from the Bronze Age to the Republican Period. Amsterdam Archaeological Studies 15, Amsterdam.

Abbolito Simonetti (1995): G. Abbolito Simonetti, Venerabile Beda. Esposizione e revisione degli Atti degli Apostoli, Roma.

Benassai (2002): R. Benassai, "Un cinerario campano a Perugia", Annali della Fondazione per il Museo Claudio Faina 9, 525–540.

Benelli (2009): E. Benelli, "La società chiusina tra la guerra annibalica e l'età di Augusto. Osservazioni archeologiche ed epigrafiche", Ostraka. Rivista di antichità 18, 303–322.

Benelli (2012): E. Benelli, "Chiusi da città etrusca a municipio romano", in: Simona Marchesini (Ed.), Matrimoni misti e identità in cambiamento, Trento, 103–112.

Berrendonner (2007): C. Berrendonner, "La società di Chiusi ellenistica e la sua immagine. Il contributo delle necropoli alla conoscenza delle strutture sociali", Etruscan Studies 10, 67–78.

NIV: The Holy Bible. New International Version. Biblica. Biblica Inc. (2011).

Blake (1999): E. Blake, "Identity-Mapping in the Sardinian Bronze Age", European Journal of Archaeology 2.1, 35–55.

Botarelli (2005): L. Botarelli, Carta Archeologica della Provincia di Siena VII. Radicofani. Botarelli, Lucia, Siena.

Bradley (2007): G. J. Bradley, "Ancient Italy. Regions without Boundaries", in: Guy J. Bradley, Elena Isayev u. Corinna Riva (Eds.), Romanization: the End of Peoples of Italy?, Exeter, 295–322.

Bruhn (2011): J. C. Bruhn, The Sociology of Community Connections, London.

Bucholtz/Hall (2004): M. Bucholtz u. K. Hall, "Language and Identity", in: Alessandro Duranti (Ed.), A Companion to Linguistic Anthropology, Oxford, 369–394.

Cambi (1996): F. Cambi, Carta Archeologica della Provincia di Siena II. Amiata (Abbadia San Salvatore), Siena.

Campana (2001): S. Campana, Carta Archeologica della Provincia di Siena V. Murlo, Siena.

Cenni (2007): F. Cenni, Carta Archeologica della Provincia di Siena VIII. Buonconvento, Siena.

Chaiken/Trope (1999): S. Chaiken u. Y. Trope, Dual-Process Theory in Social Psychology, London.

Cherici (1987): A. Cherici, "Cortona struttura e storia. Materiali per una conoscenza operante della città e del territorio", in: Giancarlo Cataldi (ed.), Materiali per una carta archeologica del territorio cortonese e Materiali per una carta archeologica di Cortona, Arezzo, 141–181.

Cherici (1992): A. Cherici, "L'insediamento antico nel territorio aretino. Carta Archeologica F. 114 II", Rivista di Topografia Antica 2, 23–90.

Cipollone (2009): M. Cipollone, "Necropoli di S. Caterina Vecchia, tomba 6", in: Marco Saioni (ed.), Invito al Museo. percorsi, immagini, materiali del Museo Archeologico Nazionale dell'Umbria, Perugia, 155 f.

D'Ercole (2011): M. C. D'Ercole, "Sharing New Worlds. Mixed Identities around the Adriatic (Sixth to Fourth Centuries B.C.E.)", in: Erich S. Gruen, Cultural Identity. In the Ancient Mediterranean, Los Angeles), 428–451.

Da Vela (2016): R. Da Vela, "Cultural Transmission and Semantic Change of Ceramic Forms in Grave Goods of Hellenistic Etruria", Distant Worlds Journal 1, 27–54.

Felici (2004): C. Felici, Carta Archeologica della Provincia di Siena VI. Pienza, Siena.

Gambogi/Palladino (1999): P. Gambogi u. S. Palladino, Castiglioncello. La necropoli ritrovata. Cento anni di Scoperte e Scavi (1896–1897), Rosignano.

Hall (2012): J. M. Hall, "The Creation and the Expression of Identity. The Greek World", in: Susan E. Alcock u. Robin Obsborne (eds.), Classical Archaeology", Chichester2, 350–367.

Hoo (2015): M. Hoo, "Maanvrouwe aan de Oxus: Baktrisch Ai Khanum als casus voor globalisering in de Oudheid", Tijdschrift voor Mediterrane Archeologie 54, 34–40.

Eriksen (2015): T. H. Eriksen, "Multiculturalism, Anthropology of,", in: Neil J. Smelser u. Paul B. Baltes (Eds.), International Encyclopedia of the Social and Behavioral Sciences, Amsterdam, 28–33.

Iozzo (2011): M. Iozzo, "Il beauty-case di Thana Plecunia", in: Simona Rafanelli u. Paola Spaziani (Eds.), Etruschi. Il privilegio della bellezza", San Sepolcro, 206–209.

Jiménez-Badillo (2012): D. Jiménez-Badillo, "Relative Neighbourhood Networks for Archaeological Analysis", in: Mingquan Zhou, Iza Romanowsa, Zhongke Wu, Pengfei Xu u. Philip Verhagen (Eds.), Revive the Past. Computer Applications and Quantitative Methods in Archaeology (CAA Beijn 12th–16th April 2012), Amsterdam, 370–380.

Knappett (2013): C. Knappett, Network Analysis in Archaeology. New Approaches to regional Interaction. Knappett, Oxford.

Luhrmann (2001): T. M. Luhrmann, "Identity in Anthropology", in: Neil Smelser u. Paul B. Baltes, International Encyclopedia of the Social and Behavioral Sciences, Amsterdam, 7154–7159.

Maggiani (2013): A. Maggiani, "Mercenari liguri?", Annali della Fondazione per il Museo Claudio Faina 20, 233–255.

Maggiani (2014): A. Maggiani, "La necropoli di Balena. Una comunità rurale alla periferia del territorio di Chiusi in età medio e tardo ellenistica (II-I sec. a.C.), in: Monica Salvini (ed.), Etruschi e Romani a San Casciano dei Bagni. Le stanze cassianensi, Rom, 51–57.

Martin-Kilcher (1998): S. Martin-Kilcher, "Gräber der späten Republik und der frühen Kaiserzeit am Lago Maggiore", in: Peter Fasold, Thomas Fischer, Henner von Hesberg u. Marion Witteyer (Eds.) Bestattungsitte und kulturelle Identität. Grabanlagen und Grabbeigaben der frühen römischen Kaiserzeit in Italien und den NOrdwest-Provinzen (Kolloquium in Xanten vom 16. Bis 18. Februar 1995). Xantener Berichte 7, Bonn, 191–252.

Mizoguchi (2013): K. Mizoguchi, "The evolution of prestige good systems: an application of network analysis to the transformation of communication system and their media", in: Carl Knappett (Ed.), Archaeology of interaction: network perspectives on material culture and society, Oxford/New York, 151–180.

Mol (2014): A. A. A. Mol, The connected Caribbean. A socio-material network approach to patterns of homogeneity and diversity in the pre-colonial period. Leiden.

Nardini (2001): A. Nardi, Carta Archeologica della Provincia di Siena IV. Chiusdino, Siena.

Paolucci (2008): G. Paolucci, Carta Archeologica della Provincia di Siena. IX. Chianciano Terme, Siena.

Pitts/Versluys (2015): M. Pitts – M. J. Versluys (Eds.), Globalisation and the Roman World. World History, connectivity and material culture, Cambridge.

Prost (2002): F. Prost, "L'alphabet des Déliens à l'époque archaïque", in: Christel Müller u. Francis Prost (Eds.), Identités et cultures dans le monde méditerranéen antique. Études en honneur de Francis Croissant, Paris, 305–328.

Rivers/Evans (2013): R. Rivers u. T. Evans, "Network Models and Archaeological Spaces", in Andrew Bevan u. Marc Lake (Eds.), Computational Approaches to Archaeological Spaces, Oxon, 99–126.

Salvini (2014): M. Salvini (Ed.), Etruschi e Romani a San Casciano dei Bagni. Le stanze cassianensi, Rom.

Stek (2013): T. D. Stek, "Material Culture, Italic Identities and the Romanization of Italy", in: Jane DeRopse Evans (Ed.), A Companion to the Archaeology of the Roman Republic", Chichester, 337–353.

Straus (2002): F. Straus, Netzwerkanalysen. Gemeindepsychologische Perspektiven für Forschung und Praxis,Wiesbaden.

Terrenato (1998): N. Terrenato, "The Romanization of Italy: global acculturation or cultural bricolage", in: Colin Forcey, John Hawthorne, John u. Robert Witcher (Eds.), Theoretical Roman Archaeology Conference and International Roman Archaeology Conference Notthingam April 1997, Oxford, 20–27.

Terrenato (1998a): N. Terrenato, "Tam Firmum Municipium: The Romanization of Volaterrae and its cultural implications", Jounal of Roman Studies 88, 94–114.

Torelli (1992): M. Torelli (Ed.), Atlante dei siti archeologici della Toscana., Rom.

Valenti (1995): M. Valenti, Carta Archeologica della Provincia di Siena. I. Il Chianti senese Castellina in Chianti, Castelnuovo Berardenga, Gaiole in Chianti, Radda in Chianti, Siena.

Valenti (1999): M. Valenti, Carta Archeologica della Provincia di Siena III . La Valdelsa (Colle Valdelsa, Poggibonsi), Siena.

Versluys (2014): M. J. Versluys, "Understanding objects in motion. An archaeological dialogue on Romanization", Archaeological Dialogues. 20.1, 1–20.
DOI: https://doi.org/10.1017/S1380203814000038 (16.05.2014)

Versluys (2013): M. J. Versluys, "Material Culture and Identity in the Late Roman Republic (c. 200–c. 20)", in: Jane DeRose Evans (Ed.), "A Companion to the Archaeology of the Roman Republic, Chichester, 429–440.

Wallace-Hadrill (2011): A. Wallace-Hadrill, "Pompeian Identities. Between Oscan, Samnite, Greek, Roman and Punic", in: Erich S. Gruen (Ed.), Cultural Identity in the Ancient Mediterranean, Los Angeles, 415–427.

Author[40]

Dr. Raffaella Da Vela

Universität Leipzig
Institut für Klassische Archäologie und Antikenmuseum
Ritterstraße 14
D-04109 Leipzig

Email: davela.network@gmail.com