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).

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

30

DCO 3,2 (2017)

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).

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

34

DCO 3,2 (2017)

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.

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.

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

---