

SNAP:DRGN Advisory Board (AB)

2nd meeting Skype (voice only) 2014-08-27

Present: Øyvind Eide (ØE, chair), Fabian Koerner (FK), Laurie Pearce (LP), Charlotte Roueché (CR), Rainer Simon (RS), Gabriel Bodard (GB) (principal investigator)

Apologies: Sonia Ranade, Robert Parker.

The meeting lasted one hour.

Minutes written by Øyvind Eide based on notes from Laurie Pearce.

1. Call to order (15.00)

ØE welcomes. Call for other business: none.

2. Updates from PI (GB) (15.05)

Main development is the release of Cookbook 1.0. It required testing, getting base data sets into RDF. Most of work was directed on showing/recommending to others how to put data into SNAP format. There was a major worksprint in Edinburgh where priorities for future funding of the project was also discussed, as the project is currently only funded for this calendar year.

Based on the important distinction of two kinds of data sets, the project has to make decisions about the next stages and priorities in the development. Two types of sets of data are distinguished:

1. Prosopography: this is information about persons, intended to disambiguate them (even it disambiguation is not always successful). This is the kind of data SNAP would import.
2. List of attestations: SNAP will not import and assign URIs to such data. The data owners are invited to annotate such datasets with SNAP ids. However, might test second integration to incorporate the data and annotate at a later stage.

SNAP is in discussions with VIAF about useful association between the two. There is a subset of about 2000 person references from VIAF, with dates before 1000 AND wrote in either Latin or Greek. Those w/o dates or languages are omitted from this set. This small subset has been imported to SNAP.

How SNAP can help VIAF: VIAF is not interested in all references, just those who are authors according to the library catalogues. If SNAP had a field for role/occupation, contributors who has data about persons being creators/actors/painters/poets/theologians etc. can be asked to provide it in order to flag relevant persons for VIAF. VIAF would then assign identifiers to these persons, even if no real information beyond the fact that the tombstone says “painter” is available.

Additional datasets: SNAP has received data from the British Museum and VIAF, and are in advanced conversation with others, including the Hellenistic Babylonia, PBW, Smith Dictionary, RIB and the Zenon catalogue.

Working on the triples to show functionality: this is the slowest part to get ready. RDF requires much work. Triples store that had been recommended was not capable of handling the data, and had to start with a more robust triple store. As a result, many elements of the API that were specified haven't been built yet. SNAP has been in touch with contributors, have made mock up RDFs which are being tested, but no further production imports yet.

3. Discussion of update (15.15)

CR asked about the VIAF relationship. For example, for Julius Caesar, the VIAF record might put in one role only, and that role might not be author(ship). Have to consider the specific relationship

as creator of work. There is RDF relationship between individuals and things they create.

GB replied that there is nothing to preclude assigning more roles, but are building subset that is minimalistic to work with other projects; not building a prosopography.

ØE: On more general level, more databases will have specific things each is interested in. Based on the simplicity of the SNAP project, these things will not apply to top layer. So callbacks to local databases may be necessary, but this is not simple. In order to get to a situation where one can access more detailed information from the local databases, one would have to map into something more complex than SNAP. One needs more a more advanced ontology to be able to connect into more complex prosopographies.

GB replied that there are not so many fields left that are not accounted for in SNAP that are reducible, and none of the providers does that level of reduction anyway. So this is currently not a relevant problem.

ØE asked about the discussion in Edinburgh that might have focused on future funding. He asked GB to share his notes/impressions with SNAP, even if the notes are brief. GB agreed to do so.

- person-search as a research tool
- graph-search as a crosswalk channel
- speccing full annotation, certainty & disagreement
 - *and getting uptake*
- Pelagios-style harvester for oac annotations pointing in to snap
- infrastructure and optimization

GB: Items that remain for future processing: to integrate into a SNAP graph: a new scholarly statement identifying a name instance as a specific person, and to indicate the authority of who is making that identification and who disagrees. Will not yet have many references pointing in to SNAP by the end of the year. However, it would be useful to have the Pelagios harvester with “here are all the persons” and “these are the datasets that have been annotated with references to these persons.”

Getting infrastructure working and triples store working: Sesame is not powerful, but still needs much more memory, say the equivalent of 10 usual projects in a campus institution. (Migrating to FourStore led to much improved performance, but this is still a relatively small dataset.) Have to consider how to optimize and get more computing power.

ØE: Might look to supercomputing as a possible source of funding – most work in the humanities does not need this, but there is funding available if one can document a need.

CRMinf is an extension to CRM covering argumentation and inference making Link to documentation that is currently under development:

CRMinf: the Argumentation Model. An Extension of CIDOC-CRM to support argumentation.
http://www.cidoc-crm.org/technical_papers.html

4. Discussion of Cookbook (15.30)

The PDF version does not have much supporting prose. GB asks for pointers that could be added to the paragraphs to clarify what is intended and/or necessary to help users who are not able to read and understand RDF easily. He notes that the cookbook does not include a soup-to-nuts example, one that takes a user from start to finish. Should the cookbook include full markup of persons as examples? The meeting agreed this was a good idea, either in the cookbook or linked to from it.

LP asked about the preferred means for providing comments/feedback.

GB: Email is good for straightforward corrections, but please use the ancient-people email list for discussion. ØE noted that the raising of specific points on the ancient-people list is good way to flag

issues.

CR found the cookbook to be very clear and helpful. However, as she went through the list of items she lost the overall picture. It would be good to have examples of minimum structure needed. Some potential contributors may not be certain of whether they have prosopographical data in the SNAP sense.

GB: Can show minimal sets with only date and name. He wants to include a description of what constitutes a prosopography in the SNAP sense.

ØE suggested that illustrations might be useful in order to understand the contents better.

RS found the cookbook clear. He brought up the topic of name properties. What makes a name important enough to get richer encoding? GB would like minimal encoding, such as birth name. Whether one would use additional properties depends on whether the contributing database has controlled vocabularies of names, as Trismegistos and LGPN have. One could and should contribute variant names to SNAP, but SNAP prefers the primary name.

RS brought up the annotating of documents with SNAP URIs as in the Pelagios use case. What is the boundary between a name and attestations to it? When does it become attestation and use RDF in cookbook, contrary to annotating images on inscriptions

GB: This comes down to whether data is truly prosopographical in the SNAP sense, as discussed above. Only some contributors have URNs for names. If you have, contribute them; otherwise, SNAP still wants the names. Attestations are links from SNAP to other data sets. Annotations are links for other data sets to SNAP.

ØE: Two-way links could allow for ingesting lists also from non-prosopographies.

GB: It is a question if your data is prosopographical or not, but this is not thought fully through. The intent of your annotations/attestation is central.

ØE: The issue of date (discussed in the first meeting) is now well-defined, and is not complex. Date is understood as a time period/point that overlaps with the life of person. It would be good to have an equally simple and clear definition for place.

GB: This should be linked to importance. He will ponder on a formulation.

FK: This should be left to the provider. We must keep in mind the choice of place will be difficult.

GB: One can include more than one place, if that's the case.

FK: Would it be good?

Gabby: It will probably not hurt. More than one place means that all of them are significant.

ØE: We should stop now and continue the discussion at the ancient-people list.

5. Any other business (15:50)

None.

6. Summing up (15.55)

ØE asked about the SNAP ontology: should this be discussed at Skype or another format?

GB: It would be good to discuss it in more detail, but it is dependent the participation of Faith Lawrence and Hugh Cayless.

ØE: It could either be the topic of the third AB meeting or an additional ad hoc meeting on the topic. The AB will agree on how to proceed via email.

ØE thanked the participants and closed the meeting.