SNAP:DRGN Advisory Board (AB)

2nd meeting Skype (voice only) 2015-02-23

Present: Øyvind Eide (ØE, chair), Fabian Koerner (FK), Robert Parker (RP), Laurie Pearce (LP), Charlotte Roueché (CR, until 17:35), Rainer Simon (RS), Gabriel Bodard (GB, principal investigator)

1. Call to order (17.00)

2. Updates from PI (GB) (17.05)

The project ended officially at the end of December 2014, but the granting organisation has approved a budget extension until the end of September 2015 with no additional funding. This means that the remaining approx. £ 4,000 can be used for travel for outreach and work with partners. GB asked for suggestions for how to best put the remaining funds to use, e.g., travel and/or a small workshop.

There are three main outputs from the project:

1) Cookbook. This is ready for release of 2.0 with substantial development from 1.0. Cosmetic changes are still needed in order to improve development form, but the cookbook is functional.

2) Ontology. This is to a large extent based on reuse from other ontologies, including LAWD, FOAF and SKOS. It contains a number of relationship types between people. This is adequate for current data but experience shows that it must be expanded whenever new datasets are added.

3) Triplestore. This includes less data than expected, only five datasets so far. It turned out to be more difficult than anticipated to get data in from providers. However, another 10 or so datasets are waiting in line, held up by different obstacles.

The only way of finding a person is currently through a URI or through a query to the SPARQL endpoint. Better search facilities are needed. As there is no more funding for salaries left it is limited how much can be done with this for the time being, but a group of nine students are ready to start working on the project. Whether this will be mainly a training exercise for the students or if they will be able to do significant work on some of the around 40 identified remaining micro-tasks remains to be seen. Work on improving the API would be especially welcome.

SNAP is keen to identify new prosopographies and partners. These can be of three types:

1) Actual databases with information about persons, from full prosopographies to those of the type found in LGPN.

2) Sources of names with identifiers, including library catalogues, lists of authors, lists of emperors.

3) Users of the SNAP system for disambiguation of local resources. This includes projects and texts with lots of references to personal names needing disambiguating tools as well as out-of-date works that are still being used and cited; some of which have even been supplanted by more recent print versions.

3. Discussion of update (17.20)

Some travel ideas were suggested:

- LP suggested a possible visit to BPS. She also pointed out that the HBTIN corpus is still 12-18 months from a stage where it will be ready for SNAP.

GB replied that this sounds interesting; however, given travel distance and costs, it may be better if SNAP team member Hugh Cayless make this visit.

- ØE suggested that some funding could be spent on working with the co-reference development and the establishment of family relationships and roles in CIDOC-CRM, e.g., through a workshop in London in late April/early May – some of the people working on this is located in London. He also suggested a presentation of the SNAP ontology at the next CRM SIG meeting in May would be a good idea.

GB was positive to the idea and commented that it would be good to bring team member Faith Lawrence into the conversation about family roles and relationships. There are many new non-family relationships being identified, esp. from Smith Dictionary. Examples include 'appointed' and 'killed'.

- FK suggested a Berlin visit to meet with the Akademie about PMBZ and other topics of common interest.

GB said that this would be good and that he hopes there will be more data to get from the Akademie. He had a meeting with Matthäus Heil in December, who was keen to work with SNAP.

- CR suggested that the online version of the British Prosopography of the Byzantine Empire, which first version 2001 on CD, should also be presented in Berlin. John Bradley is close to finalising the work of turning it into a web publication. It includes all relevant PMBZ ID numbers.

4. Future work (17.40)

GB presented his ideas for the future. He is currently putting together a grant for a further project, consisting of:

1. Data creation and ingestion of existing data sets.
2. Community building. This is needed because the plans for the inclusion of data sets in the original proposal was too optimistic.
3. Identify new types of input, e.g., OCR of and automatic extraction of data from already scanned out of copyright sources.

FK asked about the feasibility of this. GB replied that it would be essential to define how much resources (time) to be put into such experiments as they could easily swallow large resources without significant results.

4. Experiments with the establishment of new prosopographies, e.g., for gods.

Gods, mythological persons, and historical persons are not the same but still interconnected; there was some discussion of the feasibility of such a project, but the idea was taken positively. ØE pointed out that in the cultural heritage context of the CIDOC-CRM gods and mythological persons are concepts.

5. Scholarly use of the resources created; what will people do with this? Building better search functionality and finding out what functionality researchers want is important. This should be done in the form of workshops with rapid prototyping so that the researchers can play with possible solutions to understand better what functionality they really want.

6. For the data set itself, it should develop into a gold standard for future work, with the possible use of name tables for spell checking and for morphological parsing.

7. As a scholarly contribution GB also suggested to define how to model opinions, develop systems for structured commentaries. One could have 2–3 groups of people, e.g., scholars, graduate students, the general public, to contribute disambiguation and co-references.

ØE commented that the draft standard CRMinf (CRMinf: the Argumentation Model. An Extension of CIDOC-CRM to support argumentation. URL: http://www.cidoc-crm.org/docs/cidoc_crm_sig/CRMinf-0.4.pdf) could be useful here.

5. Any other business (17.50)

GB asked if anyone would be interested in a discussion about the SNAP ontology with Faith Lawrence and Hugh Cayless? Nobody opposed this. GB will take upon himself to organise it.

GB also signalled that he would invite the group to become members of the AB of the next project. Formal invitations will follow in due time.

6. Summing up (17.55)

ØE thanked GB on behalf of the AB, GB thanked everybody in the AB for their contribution. A general round of thanks.