

People People

<http://goo.gl/9gqpNt>

Present:

Alessandro Adamou (AA) - OU: Knowledge Media Institute

Agiatis Benardou (AB) - DCU: Europeana Research / DARIAH-EU / ARIADNE

Gabriel Bodard (GB) - KCL: SNAP Project,

Hugh Cayless (HC) - Duke University / SNAP Project

Anna Collar (AC)

Tom Gheldof (TG) - KU Leuven / DARIAH-VL / Trismegistos / EAGLE-Europeana / SNAP2

Ethan Gruber (EG): American Numismatic Society

Chris Hilton (CH) - Wellcome Library, London

Lorna Hughes (LH) - School of Advanced Study, University of London/ nedimah.eu

Faith Lawrence (FL) - KCL: SNAP Project,

Richard Light (RL)

Charlotte Roueché (CR) - KCL

Rainer Simon (RS) - AIT: Pelagios Project

Pau de Soto (PS) - Univ. of Southampton: Pelagios Project

Fiona Maguire (FM)- Uni of Birmingham: Estoria De España Project

Problems with People:

- RL: Identifying unique individuals
 - CR: Everyone has a source!
 - GB: People may not be indivisible people (although most should be) e.g. Gods
 - GB: People may not actually exist/be human
 - CH: Also groups, families, organisations...
 - FL: Titles etc
 - EG: Named "creator" may be 'school of' rather than 'the person themselves'
- RL: Names can be false friends as not unique
 - GB: Not everyone has a name (for reasons)
-

TG: TM uses attestation as primary source and then links attestation to person

GB: Does this solve the Bradley's "Personhood" problem?

FL: QuAC: Qualified Authenticated Coreference (for when you have something that looks like a person and acts like a person but which may or may not be a person)

Personhood problem working group (GB, HC, AC, FL, RL, VV):

Assumption: Each person(*) needs a URI

Assumption: Each person has 1+ attestations

Problem 1: a claim that some attestations belong to a person who can't be referenced by URI (negative claim? e.g. that a statement is not true)

Problem 2: a claim that attestation is not an attestation, or attests multiple people, etc.

Proposed solution 1:

- (cf TM behavior: Start from attestations, then claim creates person co-refs)
- recommend new data projects be granular to the level of attestation
- persons are created by claims that attestations co-reference (or not)
- Problem remains if claim affects atomicity of attestation:

Solution 2:

- recommend new projects be granular to level of source
- only create attestations by claim that source => attest => person

Questions:

1. what would an existing prosopography do—can't start again from scratch?
2. how does this relate to factoid model?

We have been thinking about the issue of identifying both individual people and person-like entities (cf. mythical beasts, gods, families, non-specific supernatural forces, corporate bodies, and so on) and the problems this raises in describing these in LOD formats. We start from two assumptions: first, in LOD, each person (which we use henceforth as a shorthand for any person-like entity as above) needs a URI. Second, person entities only exist in LOD as a result of them having one or more attestations in the sources. This works very well in giving us unique and indivisible person identities to which we can refer and make statements.

The first major problem with this model comes when a new scholarly claim contradicts the assignment of an attestation or group of attestations to a single person, in particular if this claim would require reference to a new person (for whom no URI yet exists) or would simply be a negative claim that an attestation is not of said person.

We believe, partly inspired by the current functionality of the Trismegistos database, that this problem would be addressed if instead of considering person-records to be the unique and indivisible unit of the dataset, we first assign URIs to each attestation. Person records would then be created, and given URIs, as the result of a claim that one or more attestations refer to a single person. Competing claims could then be recorded that create different persons, to whom URIs would also be assigned. Each claim would have to be annotated with information about provenance, author, argumentation, bibliography and so forth, so that judgements can be made about which claims to trust in a given context, counter-claims can explicitly argue against them, etc.

This solution however leaves a further problem, namely that a claim might also affect that existence of attestations themselves. A particular source might be reinterpreted as a place rather than a person reference, or a name and filiation rather than name and surname (so

attesting to two persons). If the attestation is the unique and indivisible unit of the dataset, this sort of claim has the same problems that claims about persons do in the problem posited above.

We propose that a solution to this second problem is to make the lowest, URI-bearing unit of the dataset the source (text, text fragment, image, etc.) of the person-information. A claim (which can be addressed, annotated, disagreed with, etc.) about the source engenders an attestation record, with a URI; a further claim about one or more attestations engenders a person record, with a URI.

A major unresolved question arising from this discussion is how do we expect an existing prosopography to address these problems—given that they will likely not be in a position to start their entire project from scratch with our new data model. We would also like to consider a bit more closely how the two solutions proposed above relate to, are compatible with, and even reproduce the “Factoid Model” of prosopography.

AC + FL: URIncorns - sort after mythical creature of linked data

FL: Divisible vs Indivisible persons

FL: Divisible persons have an additional layer of conceptualisation (fictionality, religious belief, shared position etc)?

1. What does the community need to increase connectivity?

- Stable URIs
 - RS: URIs can be used for clusters of attestations
 - AA: the moment attestations grow to include unambiguous individuals, one should keep stable CoolURIs if they choose to be opaque
- What person authorities do we have (<https://goo.gl/kZ1pWP>) :
 - ULAN
 - VIAF (100s of library catalogs, includes ULAN)
 - CR - Bad on Greek & Cypriot
 - SNAC (based on EAC-CPF)
 - OrcID
 - ISNI (International Standard Name Identifiers)
 - SNAP (Classical + related)

- GB: Includes some VIAF people and will be feeding creator data back to VIAF
- ADS have modern archaeologists
- TTT
- Lives of WW1 (defacto URIs for identified people)
- [and NB: large amounts of personal / corporate / family data out there in archive / library / museum systems, currently not available as LOD at all: in proprietary formats or XML]
- GB: How can we make connections/co-references between datasets?
 - RL: side-by-side comparisons
 - CH: Make it easy to find identifiers and embed/use identifiers
 - EG: reconciliation service
 - EG: Make it require a little effort as possible for users
 - CR: cf. identity suggestions in GENIE (sp.?)
 - EG: [Open Refine](#) ?
 - Works on Freebase (RIP)—port to Wikidata/VIAF?
 - EG: Training needed
 - HC: Guidance also needed (#this issue)
 - Can get RDF output

2. What can we achieve today, this month, in one year?

(What have we achieved today?

1. asked ourselves questions
2. Tasked WG to draft “nature of identity” position text (GB, RL, AC, HC => John Bradley)
3.)

In next month:

1. list of person-authorities (do they want to receive our data too?)
 - a. are these guidelines/recommendations of interest to AHRC?
 - b. FL to create GoogleSpreadsheet with list of person authorities above, circulate (<https://goo.gl/kZ1pWP>)
- 2.

In next year:

1. Think about requirements behind HC's concern about attaching provenance to triples?
2. Guidelines for creation of data (FL + AC + RL + CR + Sopracasa):
 - a. SNAP Cookbook *plus*
 - b. New Byzantine seals scholar at KCL *plus*
 - c. AC new database *plus*
 - d. Genealogy experience
 - e. INCLUDE guidance on achieving "wow-factor" (widgets and other available resources)

RL: Turn discussion of nature of identities into a few paragraphs of prose

HC: address concerns about LOD infrastructure vis a vis disagreement/provenance

FL: Triples or Quads?

GB: or everything reified! (Turtles all the way down.)

CR: guidelines for creating new material (learning from past projects)

LH: develop and demonstrate use-cases (exploit the Wow-Factor)

(FL: even simple things, the awld.js widget)

CH: use cases enable archivists / librarians / museum cataloguers etc to put better pressure on system suppliers to build these things into commercial systems

Collect widgets and tools that can be used: AWLD.js; Doug+Ceri's widget;

3. How can contribute to LOD ecosystem?

FL: Give your [GB: historical] data to SNAP/Use SNAP Identifiers/Tell SNAP how we can help you ;-)

<http://snapdrgn.net/>

4. What should community invest in?

CH: influence creators of data into using LOD; into improving quality of data and metadata

FL: lean on research councils to demand good practice => Open data->LINKED open data

GB (white papers?)

CR identifiers for sources

document/object (PN/TM)

person

bibliography :: Worldcat Works

