

Jouni Tuominen: EMLO Prosopographical Data Model – Towards a Biographical Conceptual Reference Model

During my STSM in Oxford in March/April 2016 I engaged in the development of a data model for representing biographical information for

prosopographical research. The resulting Biographical Conceptual Reference Model (Bio CRM) is a general, event-based and role-centric

schema, which can be applied to different prosopographical databases, by extending it with particular event and role classification

schemes. The design of the data model is based on the anticipated use cases for prosopographical research, in the sense that the data

model should support the principal tasks relevant for the researchers in this area.

The work started with analyzing the current data model of the Early Modern Letters Online (EMLO) database managed in Oxford. The EMLO

data model is an event--based, person-- and role--centric model for representing the activities a person has participated in during his

life, essentially forming his biography. A such activity involves one or more participants, acting in specific roles, place(s), and a

time expression. The EMLO data model is based on PROV for representing the activities and the participant roles. Further, related data

models for representing (biographical) events were inspected and compared, including: CIDOC CRM, BIO, Relationship, BIBO, ORG, Event

Ontology, LODE, VIVO, Simple Event Model (SEM), ULAN, and PROSO.

The outcome of the work is a data model with the following features:

- Act as an extension of CIDOC CRM for compatibility with other cultural heritage datasets.

- Separation of the general biographical data model from the EMLO--specific event types and participant roles, to support also other

kinds of prosopographical datasets, concerning different cultures and time periods.

- Support for principal query types for prosopographical research: finding a set of people who share selected characteristics, and

extracting networks of people based on some criterion for further analysis, e.g., with external visualization toolkits.

- Distinguish between unary roles (e.g., professions), binary relationships (e.g., family relations), and events (e.g., baptism) of

person's biography for intuitive information representation and query writing, with a shared role--centric modeling approach.

- Represent the roles of participants of events as instances of OWL classes, enabling reasoning and validation of the integrity of

prosopographical data created based on the model, ensuring the quality of the data (e.g., a baptism may involve only participants in

the roles of baptismal candidate, officiant, etc.).

Bio CRM facilitates the integration of different prosopographical datasets represented with their own classification schemes for

events, actors, roles, etc., in the spirit of CIDOC CRM, which aims to harmonize datasets on the field of cultural heritage.

More information and a working paper on Bio CRM:

stsm-abstract-for-cost-website-JT
<http://seco.cs.aalto.fi/projects/biographies/>