

Bio CRM: A Data Model for Representing Biographical Data for Prosopographical Research

Jouni Tuominen^{1,2}, Eero Hyvönen^{1,2}, and Petri Leskinen¹

¹ Semantic Computing Research Group (SeCo), Aalto University, Finland, and

² HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Finland

<http://seco.cs.aalto.fi>, <http://heldig.fi>

`firstname.lastname@aalto.fi`

Keywords: Linked Data, Data models, Biographical representation, Event-based modeling, Role-centric modeling, Prosopography

Type of submission: original unpublished work

Biographies make a promising application case of Linked Data: they can be used, e.g., as a basis for Digital Humanities research in prosopography and as a key data and linking resource in semantic Cultural Heritage portals. In both use cases, a semantic data model for harmonizing and interlinking heterogeneous data from different sources is needed. We present such a data model, Bio CRM [1], with the following key ideas: 1) The model is a domain specific extension of CIDOC CRM, making it applicable to not only biographical data but to other Cultural Heritage data, too. 2) The model makes a distinction between enduring unary roles of actors, their enduring binary relationships, and perduring events, where the participants can take different roles modeled as a role concept hierarchy. 3) The model can be used as a basis for semantic data validation and enrichment by reasoning. 4) The enriched data conforming to Bio CRM is targeted to be used by SPARQL queries in flexible ways using a hierarchy of roles in which participants can be involved in events.

Bio CRM provides the general data model for biographical datasets. The individual datasets concerning different cultures, time periods, or collected by different researchers may introduce extensions for defining additional event and role types. The Linked Data approach enables connecting the biographies to contextualizing information, such as the space and time of biographical events, related persons, historical events, publications, and paintings. Use cases for data represented using Bio CRM include prosopographical information retrieval, network analysis, knowledge discovery, and dynamic analysis.

The development of Bio CRM was started in the EU COST project "Reassembling the Republic of Letters" [2] and it is being piloted in the case of enriching and publishing the printed register of over 10 000 alumni of the Finnish Norssi high school as Linked Data [3].

[1] <http://seco.cs.aalto.fi/projects/biographies/>

[2] <http://www.republicofletters.net>

[3] Eero Hyvönen, Petri Leskinen, Erkki Heino, Jouni Tuominen and Laura Sirola:

[Reassembling and Enriching the Life Stories in Printed Biographical Registers: Norssi High School Alumni on the Semantic Web](#), Proceedings, Language, Technology and Knowledge 2017. June 19-20, Galway, Ireland, Springer-Verlag, 2017.