Data Modeling as a High-Wire Act. Balancing Requirements, Juggling Vocabularies, and not Falling (Short of Established Best Practice)

Oberreither, Bernhard

bernhard.oberreither@oeaw.ac.at
ACDH-CH, ÖAW, Austria

The *SemanticKraus* project aims to provide a resource for shared register and other data from three projects in the field of digital Karl Kraus research: the online editions of *Die Fackel*, [1] the *Karl Kraus Legal Papers*, [2] and *Dritte Walpurgisnacht*. [3] The final product is to be an RDF dataset created according to FAIR principles [4] which combines the biographical and bibliographical data of these projects and makes these data equally accessible to human exploration and machine query via a GUI as well as a SPARQL endpoint. In addition, the technical foundations will be laid for future expansion of the dataset through the participation of other projects from this or neighboring research areas. The vision of the project is to connect Kraus research to the Linked Data cloud, laying the groundwork for future data sharing and data integration among Kraus researchers. This way, the project wants to account for the frequently raised call for open, sustainable, and reusable edition data (see, e.g., Kamzelak 2016; Wettlaufer 2018).

However, until that point is reached, a number of challenges need to be faced: such as the common problem of having to integrate (in this case: three) different data sources with different database models and data fields. The central challenge, however, is to create an appropriate data model for this diverse data. On the one hand, the *SemanticKraus* data model must do justice to a manifold and very complex subject: The core data set of the project, for example, is a text index of *Die Fackel online*. The structure of the individual issues of this journal is characterized by nested texts, comments on as well as extensive citations of texts, and other complications. In addition, the structure of the journal changed profoundly several times in the course of its existence (1899–1936) due to multiple thematic reorientations (see, e.g., Krolop 2006, Stieg 2022); this alone poses a challenge to any representation of its contents. While one wants to do justice to this matter, at the same time, however, the model must maintain a balance between specificity and standardization – all with regard to the future usability of the data and the further applicability of the data model. This dichotomy needs to be reconciled: On the one hand, Semantic Web ontologies are seen as a possibility to represent considerations and concepts that are regarded as specifically humanities-related (see, e.g., Ciotti 2014, Tomasi 2018) – thus theoretically making the most complex matter representable at the smallest granularity. On the other hand, reuse of widely shared, necessarily less specific standards is considered best practice in the Semantic Web, and with good reason; this, after all, is why the *SemanticKraus* data model as much as possible adopts common and proven vocabularies (such as CIDOC crm, [5] FRBRoo, [6] PRESSoo, [7] etc.).

In the course of evaluating said standards, another question arises: How far can/should one welcome the fact that looking through the lens of an ontology may open a new, possibly fruitful and pragmatic, but also: possibly misleading perspective on our data?

The poster proposed here for DH2023 will try to answer these questions by highlighting central waypoints in the course of creating the *SemanticKraus* data model – such as decisions on the implementation of different ontologies, on the treatment of specific data, on basic modeling principles – and convey the balancing act involved. The presentation of the poster will provide an insight into the development of the data model, cast the lessons learned into theses and by doing so stimulate discussion of issues central to the field of data modeling in the Semantic Web, especially regarding edition data. Furthermore, it aims to highlight future opportunities for participation and exchange for researchers and projects of a similar focus.

Notes

1. fackel.oeaw.ac.at
2. kraus.wienbibliothek.at
3. kraus1933.ace.oeaw.ac.at
4. go-fair.org/fair-principles
5. cidoc-crm.org
6. cidoc-crm.org/frbroo
7. cidoc-crm.org/pressoo

Bibliography

Ciotti, Fabio (2014): "Digital Literary and Cultural Studies: State of the Art and Perspectives", in: Between 4, 8. DOI: 10.13125/2039-6597/1392

Kamzelak, Roland S. (2016): "Digitale Editionen im Semantic Web. Chancen und Grenzen von Normdaten, FRBR und RDF", in: Richts, Kristina / Stadler, Peter (eds.): *"Ei, dem alten Herrn zoll' ich Achtung gern"*. Festschrift für Joachim Veit zum 60. Geburtstag. München: Allitera 423–435. https://musiconn.qucosa.de/en/api/qucosa%3A23339/attachment/ATT-0/ [26.04.2023]

Krolop, Kurt (2006): "Vom "Kampfblatt" zur "Kriegsfackel". Die Werdejahre des 'Antimediums'", in: Lunzer, Heinz / Lunzer-Talós, Victoria / Patka, Markus G. (eds.): *"Was wir umbringen"*. 'Die Fackel' von Karl Kraus. Wien: Mandelbaum 9–31.

Stieg, Gerald (2022): "Die Fackel", in: Prager, Katharina / Ganahl, Simon (eds.): *Karl Kraus-Handbuch*. Leben – Werk – Wirkung. Berlin: Metzler 103–122.

Tomasi, Francesca (2018): "Modelling in the Digital Humanities: Conceptual Data Models and Knowledge Organization in the Cultural Heritage Domain", in: *Historical Social Research, Supplement* 31, 170-179. DOI: 10.12759/hsr.suppl.31.2018.170-179

Wettlaufer, Jörg (2018): "Der nächste Schritt? Semantic Web und digitale Editionen", in: Kamzelak, Roland S., and Steyer, Timo (eds.): *Digitale Metamorphose: Digital Humanities und Editionswissenschaft* (*Zeitschrift für digitale Geisteswissenschaften* special edition 2). DOI: 10.17175/sb03