Towards a common data model for semantic annotation of digital media: A new FOSS toolchain

Rossenova, Lozana
lozana.rossenova@tib.eu
TIB – Leibniz Information Centre for Science and Technology, Germany

Sohmen, Lucia
lucia.sohmen@tib.eu
TIB – Leibniz Information Centre for Science and Technology, Germany

Duchesne, Paul
paul.duchesne@tib.eu
TIB – Leibniz Information Centre for Science and Technology, Germany

Günther, Lukas
lukas.guenther@tib.eu
TIB – Leibniz Information Centre for Science and Technology, Germany

Schubert, Zoe
zoe.schubert@tib.eu
TIB – Leibniz Information Centre for Science and Technology, Germany

Bluemel, Ina
ina.bluemel@tib.eu
TIB – Leibniz Information Centre for Science and Technology, Germany

Abstract

Traditionally, the development of metadata standards and collection management systems has focused on the management of the separations between analog objects, their descriptive surrogates and the creation of finding aids (Hurley 2005). But born-digital artifacts, such as various media files and software applications, and their associated metadata records are created, managed, and used in the same environment. Describing the ever-expanding range of significant properties of these artifacts (Laurenson 2014), tracking data provenance (Rossenova et al. 2019), and allowing collaborative research and annotation can prove particularly challenging to traditional relational database systems and their information architecture models (Hurley 2005; Wiedemann 2019), which separate schema from content, depend on fixed vocabularies and categorizations, and tend to remain siloed and closed off to external audiences. In the field of cultural heritage and the digital humanities, 2D- and 3D-digital representations of cultural assets are particularly heterogeneous in formats and structure (Blümel / Wessel 2019), hence standardized access and visualisation tools fail to meet new research objectives and institutional requirements. To bridge the gaps across traditional data management tools and media-rendering environments, at TIB's Open Science Lab we have developed a suite of tools as part of a larger national effort which involves the partnership between several research, library and cultural institutions (Altenhöner et al. 2020). In this poster we present our current development of an integrated FOSS (Free and Open Source Software) toolchain (Rossenova et al. 2022) that consists of three main components: 1) OpenRefine – for data reconciliation and batch upload (Sterner 2019); 2) Wikibase – for linked open data storage (Alípio et al. 2019); and 3) Kompakkt – for rendering and annotating 3D models, and other 2D and AV media files (Eide et al. 2019). All components of the toolchain feature graphical user interfaces aiming to lower the barrier of participation in the semantic web for a wide range of practitioners and researchers. The toolchain is envisioned as a collaborative environment with different levels of read/write access, wherein groups of researchers can work collaboratively on data upload and annotation with clear data provenance. In addition, we have placed particular emphasis on an open, community-oriented and agile development process.

Our toolchain follows FAIR principles, and adopts common standards like PIDs and the W3C annotation model, but the integration of the tools also depends on the application of a common data model that needs to accommodate a range of knowledge domains involved in the study of cultural heritage objects. This model facilitates linking media objects and annotations, and their cultural context (including historical people and places, geo-location and digital-capture-technology metadata), to the broader semantic web and various national and international authority records (GND, Getty's AAT, VIAF and more). The poster presentation will focus on the development of the common data model and how it aims to improve interoperability between datasets of digitised cultural objects. The presentation will be of interest to DH researchers, librarians and digital curators who manage research datasets containing diverse media files, and want to learn more about the possibilities of linked open data, open source software and collaborative annotation workflows.

Bibliography

Alípio, Samantha / Abdulai, Mohammed S. / Burnett, Georgina / Shick, Dan (2021): "Wikibase: the Software for Open Data projects", in: *Wikimedia Tech News*. https://tech-news.wikimedia.de/en/2021/04/14/wikibase-the-software-for-open-data-projects/ [14.01.2023].

Altenhöner, Reinhard / Blümel, Ina / Boehm, Franziska / Bove, Jens / Bicher, Katrin / Bracht, Christian / Ortrun Brand, Dieckmann, Lisa / Effinger, Maria / Hagener, Malte / Hammes, Andrea / Heller, Lambert / Kailus, Angela / Kohle, Hubertus / Ludwig, Jens / Münzmay, Andreas / Pittroff, Sarah / Razum, Matthias / Röwenstrunk, Daniel / Sack, Harald / Simon, Holger / Schmidt, Dörte / Schrade , Torsten / Walzel, Annika-Valeska / Wiermann, Barbara (2020): "NFDI4Culture - Consortium for research data on material and immaterial cultural heritage", in: *Research Ideas and Outcomes* 6: e57036. DOI: 10.3897/rio.6.e57036.

Blümel, Ina / Wessel, Raoul (2019): "DDB goes 3D". *Zenodo*. DOI: 10.5281/zenodo.5579159.

Eide, Øyvind / **Schubert, Zoe** / **Türkoğlu, Enes** / **Wieners, Jan G.** / **Niebes, Kai** (2019): "The intangibility of tangible objects: re-telling artefact stories through spatial multimedia annotations and 3D objects", in: *ICOM Kyoto 2019, 25th ICOM General Conference: Museums as Cultural Hubs: The Future of Tradition, Kyoto*. DOI: 10.5281/zenodo.3878966.

Hurley, Chris (2005): "Parallel Provenance" https://www.descriptionguy.com/images/WEBSITE/parallel-provenance.pdf [14.01.2023].

Laurenson, Pip (2014): "Old Media, New Media? Significant difference and the conservation of software based art", in: Graham, B. (ed.) New Collecting: Exhibiting and Audiences after New Media Art. Farnham, Surrey, England; Burlington, Vermont: Ashgate: 73–96.

Rossenova, Lozana / **de Wild, Karin** / **Espenschied, Dragan** (2019): "Provenance for Internet Art: Using the W3C PROV data model", in: *Proceedings of 16th International Conference on Digital Preservation iPRES 2019*, Sept 16–20, 2019 Amsterdam, The Netherlands.

Rossenova, Lozana / **Schubert, Zoe** / **Vock, Richard** / **Blümel, Ina** (2022): "Beyond the render silo - Semantically annotating 3D data within an integrated knowledge graph and 3D-rendering toolchain", in: *DHd 2022 Kulturen des digitalen Gedächtnisses. 8. Tagung des Verbands "Digital Humanities im deutschsprachigen Raum"*, Potsdam. DOI: 10.5281/zenodo.6328155.

Sterner, Elizabeth (2019): "Cleaning Collections Data Using OpenRefine", in: *Issues in Science and Technology Librarianship* 92. DOI: 10.29173/istl30.

Wiedeman, Gregory (2019): "The Historical Hazards of Finding Aids", in: *University Libraries Faculty Scholarship* 124. https://scholarsarchive.library.albany.edu/ulib_fac_scholar/124 [14.01.2023].