Semantic Annotation of Heterogeneous, Multimedia Cultural Research Data

A FOSS Toolchain for the Digital Humanities

Rossenova, Lorenza
lozana.rossenova[at]tib.eu
TIB – Leibniz Information Centre for Science and Technology, Deutschland
ORCID-iD: 0000-0002-5190-1867

Sohmen, Lucia
lucia.sohmen[at]tib.eu
TIB – Leibniz Information Centre for Science and Technology, Deutschland

Duchesne, Paul
paul.duchesne[at]tib.eu
TIB – Leibniz Information Centre for Science and Technology, Deutschland

Günther, Lukas
lukas.guenther[at]tib.eu
TIB – Leibniz Information Centre for Science and Technology, Deutschland

Schubert, Zoe
zoe.schubert[at]tib.eu
TIB – Leibniz Information Centre for Science and Technology, Deutschland

Blumel, Ina
ina.blumel[at]tib.eu
TIB – Leibniz Information Centre for Science and Technology, Deutschland
ORCID-iD: 0000-0002-3075-7640

Summary. This workshop will present a workflow for structuring and annotating multimedia datasets within a collaborative, linked open data environment. Participants will be able to take part in practical demonstrations and provide feedback on the Semantic Kompakkt toolchain that connects three existing open source software tools: 1) OpenRefine – for data reconciliation and batch upload; 2) Wikibase – for linked open data storage; and 3) Kompakkt – for rendering and annotating 3D models, and other 2D and AV media files. This toolchain was developed in the context of NFDI4Culture with a particular focus on increasing interoperability and data reuse across different domains of cultural research.

1 Extended Abstract

Traditionally, the development of descriptive 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.[1] But in the context of digital humanities research, born-digital artifacts, such as various media files, 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,[2] tracking data provenance,[3] and allowing collaborative research and annotation[4] can prove particularly challenging to traditional relational database systems and their information architecture models.[5,6] The latter tend to separate schema from content, depend on fixed vocabularies and categorizations, and 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[7], hence standardized access and visualisation tools fail to meet the objectives of critical digital humanities research as well as related research funding requirements (e.g. for FAIR data). To bridge the gaps across traditional research 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

---

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

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

[3] Lozana Rossenova, Karin de Wild, and Dragan Espenschied, "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 (2019).

[4] Øyvind Eide, Zoe Schubert, Enes Türkoğlu, Jan G. Wieners, and Kai Niebes, "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*. (2019). DOI: 10.5281/zenodo.3878966.

[5] Hurley, "Parallel Provenance."

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

[7] Ina Blümel, and Raoul Wessel, "DDB goes 3D," *Zenodo*, July 2, 2019. https://doi.org/10.5281/zenodo.5579159.

involves the partnership between research, library and cultural institutions, namely the NFDI4Culture consortium.[8]

In this workshop, researchers, digital curators and data managers will learn how to make datasets including 3D models and other media files available as linked open data within Semantic Kompakkt, the integrated FOSS (Free and Open Source Software) toolchain developed at TIB.[9] The toolchain consists of three main components (Fig. 1): 1) OpenRefine – for data reconciliation and batch upload;[10] 2) Wikibase – for linked open data storage;[11] and 3) Kompakkt – for rendering and annotating 3D models, and other 2D and AV media files.[12] 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 (Fig. 2). The workshop will feature practical demonstrations of the collaborative environment with different levels of read/write access, wherein researchers can try out the data upload and annotation functionalities for themselves.

[8] Altenhöner, Reinhard, Ina Blümel, Franziska Boehm, et al., "NFDI4Culture - Consortium for research data on material and immaterial cultural heritage," *Research Ideas and Outcomes 6:* e57036 (July 2020). DOI: 10.3897/rio.6.e57036.

[9] Lozana Rossenova, Zoe Schubert, Richard Vock, and Ina Blümel, "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* (2022). DOI: 10.5281/zenodo.6328155.

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

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

[12] Eide, et al, "The intangibility of tangible objects."

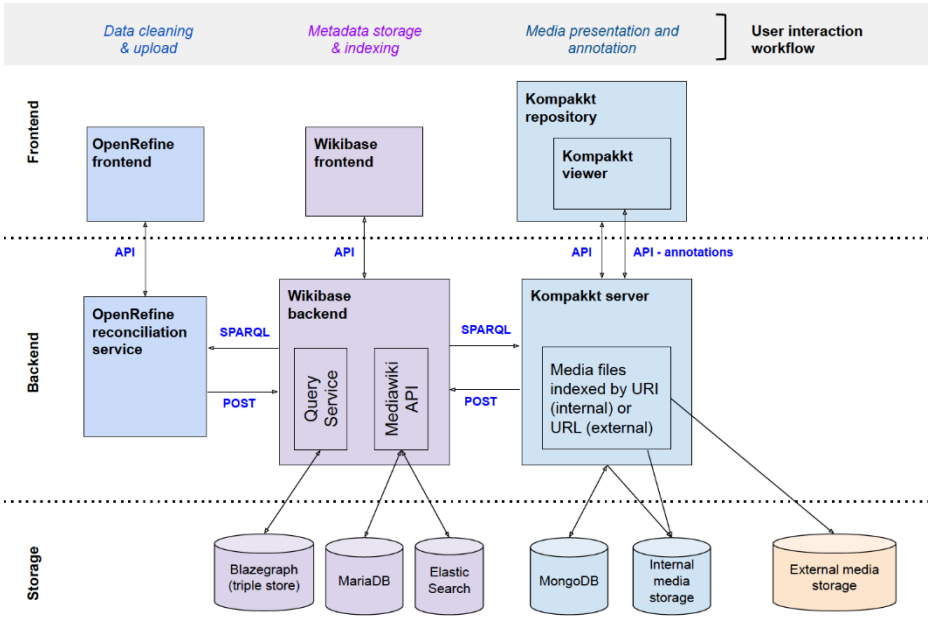

Fig. 1. Diagrammatic representation of the toolchain architecture.

Furthermore, the workshop will highlight how the toolchain follows FAIR principles and adopts a common data model with mappings to standard ontology terms (including CIDOC-CRM, FRBR, CRMdig, and DC terms) resulting in increased interoperability and data reuse across datasets from different research domains. In this way, media objects and annotations, as well as their cultural context (including historical people and places, geo-location and digital-capture-technology metadata), can be linked to the broader semantic web and various national and international authority records (GND, Getty's AAT, VIAF and more). The data model also takes into account the need for clear data provenance across heterogeneous data formats and data sources, and is compliant with the W3C Web Annotation Standard for the annotation of digital media artefacts.

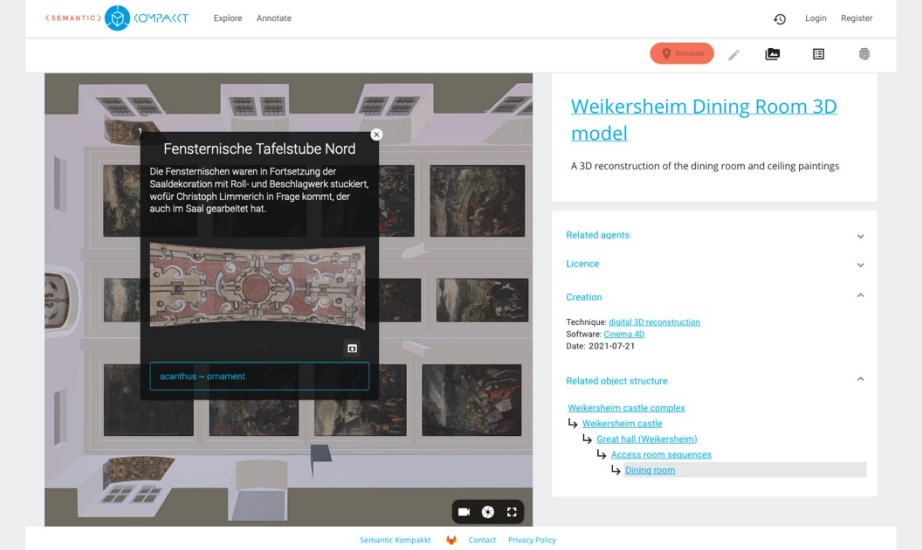

Fig. 2. Image of the Semantic Kompakkt interface showing items from the Corpus der barocken Deckenmalerei in Deutschland (CbDD) project.[13]

The toolchain has so far been developed iteratively thanks to continuous testing with real-world use cases from the NFDI4Culture community.[14] Following this agile approach, after the hands-on demonstrations, the workshop will also include requirements gathering exercises. These will seek to gather feedback and critical perspectives from participants concerning additional features or further areas of development that could benefit their specific research domains. The workshop will be of interest to researchers, digital curators and information science professionals who work with datasets containing 3D media, and want to explore the possibilities of linked open data, open source software and collaborative annotation workflows.

[13] "Corpus der barocken Deckenmalerei in Deutschland (CbDD)," Bayerische Akademie der Wissenschaften, accessed January 18 2022, https://deckenmalerei.badw.de/.

[14] Lozana Rossenova, Zoe Schubert, Richard Vock, Lucia Sohmen, Lukas Günther, Paul Duchesne, and Ina Blümel, "Collaborative annotation and semantic enrichment of 3D media: a FOSS toolchain,." in *Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries (JCDL '22).* Association for Computing Machinery, New York, NY, USA, Article 40, 1–5. (2022). DOI: 10.1145/3529372.3533289.

Bibliography

Alípio, Samantha, Mohammed S. Abdulai, Georgina Burnett, and Dan Shick. “Wikibase: the Software for Open Data projects,” *Wikimedia Tech News,* April 14, 2021. https://tech-news.wikimedia.de/en/2021/04/14/wikibase-the-software-for-open-data-projects/.

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

Bayerische Akademie der Wissenschaften. “Corpus der barocken Deckenmalerei in Deutschland (CbDD).” Accessed January 18, 2022. https://deckenmalerei.badw.de/.

Blümel, Ina, and Raoul Wessel. “DDB goes 3D”. Zenodo, July 2, 2019. https://doi.org/10.5281/zenodo.5579159.

Eide, Øyvind, Zoe Schubert, Enes Türkoğlu, Jan G. Wieners, and Kai Niebes. “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*. (2019). DOI: 10.5281/zenodo.3878966.

Hurley, Chris. “Parallel Provenance.” (2005). https://www.descriptionguy.com/images/WEBSITE/parallel-provenance.pdf.

Laurenson, Pip. “Old Media, New Media? Significant difference and the conservation of software based art.” In *New Collecting: Exhibiting and Audiences after New Media Art*, edited by Beryl Graham, 73–96. Farnham, Surrey, England; Burlington, Vermont: Ashgate, 2014.

Rossenova, Lozana, Karin de Wild, and Dragan Espenschied. "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 (2019).

Rossenova, Lozana, Zoe Schubert, Richard Vock, and Ina Blümel. "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 (2022). DOI: 10.5281/zenodo.6328155.

Rossenova, Lozana, Zoe Schubert, Richard Vock, Lucia Sohmen, Lukas Günther, Paul Duchesne, and Ina Blümel. "Collaborative annotation and semantic enrichment of 3D media: a FOSS toolchain." In *Proceedings of the 22nd ACM/IEEE Joint Conference on Digital Libraries (JCDL '22).* Association for Computing Machinery, New York, NY, USA, Article 40, 1–5. (2022). DOI: 10.1145/3529372.3533289

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

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