

Rebuilding vaults together: How the Corpus Vitrearum International interconnects digital resources that describe stained glass

Jonatan Jalle Steller (ORCID 0000-0002-5101-5275)
Academy of Sciences and Literature Mainz, Germany

Isabelle Lecocq
Royal Institute for Cultural Heritage, Belgium

Conference

Breaking Silos, Connecting Data: Advancing Integration and Collaboration in Digital Humanities

DH Benelux 2024

5-7 June 2024, Leuven

Abstract

Topic. The [Corpus Vitrearum International](#) (CVI) is an international coalition of projects engaged in the preservation, documentation, and study of stained glass. Based on a long tradition of producing printed volumes, participating projects also make data on stained-glass objects (or windows, in particular) available on the web. In this paper, we want to respond to a central question: how can a diverse and federated set of projects overcome the information silos that we built to improve access to their digital resources without centralisation? We present and discuss joint efforts across these projects to make their work more equitable starting from a position where, for example, rare books under contractual agreements with publishers, largely unstandardised object and image data, and national web repositories leave accessibility to be desired – in the sense of technological, language, and knowledge divides. After a brief look into the CVI's past, the paper focuses on four measures to interconnect the data that participating projects provide and to make it more FAIR in the process: harmonising metadata by developing shared recommendations, investing into Linked Open Data (LOD) and connectable data points, engaging with the Culture Knowledge Graph project to interconnect fine-grained cultural-heritage data along with search and SPARQL interfaces, and modernising the repository software used to manage and publish data. This work is being done in the multidisciplinary environment of the CVI, whose members include art historians, DH, and other humanities scholars as well as restoration practitioners (cf. Frodl-Kraft *passim*).

Background. Existing structures of the CVI were both beneficial and challenging for the initial data harmonisation effort. Coordinated efforts to preserve and document stained glass as cultural heritage originally emerged after WWII to document mediaeval artworks. The project, placed under the patronage of the Union Académique Internationale (UAI), was officially

created in 1952. Currently, fifteen countries contribute to it: Austria, Belgium, Canada, France, Germany, Italy, Luxembourg, the Netherlands, Poland, Portugal, Russia, Spain, Switzerland, the UK, and the USA. Since 2016, the Corpus Vitrearum has shifted its efforts to include more recent and modern artworks as well. The national committees of the CVI have successfully produced approximately 150 printed volumes, which adhere to shared publication guidelines (cf. Comité international d'histoire de l'art and Union Académique Internationale *passim*). The existence of such guidelines along with their periodical revision led to widespread acceptance of the idea to harmonise digital efforts as well. While the CVI encourages efforts to retrodigitize printed volumes, to produce open-access databases, and to experiment with multimodal digital publications, however, each national project works under vastly different legal, financial, and technological conditions. Several national projects have a strategy to pass along data to aggregators such as Europeana, but their strict licensing requirements, their focus on object data (rather than object groups and spaces), and their limited interconnections (cf. Europeana 32, 60f.)¹ mean that established players can only be one part of the CVI's use case. The idea of a European collaborative cloud for cultural heritage is similarly intriguing, but it currently remains a funding strategy rather than a technical solution (cf. Directorate-General for Research and Innovation 1-2).

Harmonisation. An initial attempt to align the data produced across CVI member projects during a meeting in 2013 brought about a rough structure of what types of data were required based on the existing publication guidelines, but national databases remained disparate and their data hard to compare. A renewed willingness to collaborate led to the establishment of a preliminary Scientific Unit of Digital Resources in the Corpus Vitrearum (Digital Unit, DU) in 2022. With the backing of practitioners and scholars in preserving and researching stained glass, the DU pooled knowledge of previous and current digital efforts into a working group with monthly video calls. Its first measures were to compare the often crumbling software platforms of member projects to each other and to build a table comparing the image and object properties that were being used so far. As a result, the group produced and published recommendations for existing and future projects collecting data on stained glass, including a set of up to 49 shared metadata fields as well as guidance on available software solutions, including the pros and cons of various technical choices (cf. Steller *et al. passim*). The federal and often volunteer-based nature of the CVI means that the DU can ultimately only recommend software instead of prescribing it, and that the metadata harmonisation recommendations are worded to highlight suitable connections over a single, shared ontology.

LOD. The DU cannot unify individual project conditions. The Swiss project, for example, needs a fully functional object management solution with a strong focus on art history because its public database is a joint initiative of the Vitrocentre and the Vitromusée Romont. The German project, on the other hand, develops its own software for multimodal cultural-heritage editions because it is guided by DH departments in the German academies programme. The Belgian data is included in a national cultural-heritage database. And the British project is in the process of adapting an older data set for inclusion in an

¹ The EDM's type system only focuses on text, image, sound, video, and 3D files, and most properties are intended for literals which do not work well for interconnections. The main connecting tissue between Europeana data are *edm:isRelatedTo* and its sub-properties.

archaeology-centric database. Hence the DU sees Linked Open Data as a shared vision and a central requirement for the software we use. The recommendations highlight which metadata properties need to link to authority files such as Wikidata, GeoNames, VIAF, GND, or Iconclass, and which need to be available in standardised formats like ISO 8601 dates and date times or CVI-specific window identifiers. We also acknowledge, however, that LOD can be seen as a spectrum ranging from the inclusion of URIs to offering RDF serialisations and content negotiation all the way to SPARQL endpoints (cf. Tietz *et al.* 1-2). Hence 5-star LOD functions as an ideal, but most member projects are more likely to reach the stage of structured data (3- to 4-star LOD). Due to technical challenges and a wealth of suitable ontologies, a single SPARQL entry point into harmonised data realistically requires an external provider and thus a centralisation that runs counter to the federated vision of LOD. Since the data the CVI provides largely revolves around images and objects, [IIIF](#) and [LIDO](#) are also part of the formats the DU recommends because both have vivid and supportive communities in the cultural-heritage domain.

Culture Knowledge Graph. To actually interconnect the structured data that CVI member projects are able to provide, the Corpus Vitrearum Germany currently participates in NFDI4Culture, a consortium of German institutions and projects to build common research infrastructures for architecture, art history, musicology, and performance and media studies. In particular, we contribute to the Culture Knowledge Graph: an open-access RDF database implementing the NFDI Core Ontology ([nfdicore](#)) and its NFDI4Culture Ontology ([cto](#)) module to collect and connect research data organised around creative works, events, people, organisations, and locations of any type with a limited set of properties that benefit data retrieval via SPARQL queries or fuzzy searches. The goal is not to produce a single data silo, but to allow scholars to identify data on, for example, a building, an artwork, or a motif they require, with detailed information (or, for example, IIIF manifests) available from the original data provider. The Culture Knowledge Graph already includes data from the German Corpus Vitrearum along with that of other participating projects (cf. NFDI4Culture *passim*). We are currently working to prepare, harvest, and (if successful) include data from other CVI member projects. Two challenges currently hinder this effort: the design of a custom ontology that is valid beyond the humanities, and the necessity to build friendlier interfaces around the SPARQL queries that the graph enables. These are common challenges of RDF-based research data management that require further work, compared to the more traditional databases many of the CVI's existing infrastructures employ.

Platforms. Apart from harmonised metadata, LOD, and a common ontology, another central aspect of breaking down walls around CVI data silos is to adapt the software we use. While many member projects are constrained in this effort by existing software solutions to be used as they are, the Corpus Vitrearum Germany is in the process of modernising its ageing software stack, increasing its likelihood of being reused, and making its visual interfaces and data as accessible and LOD-friendly as possible, including at least English and German language versions. The main body of work in this effort is already done in the form of the Cultural Heritage Framework 2 (cf. Steller *passim*), a new set of components for the content management system TYPO3 that provide coherent but modular, semantic data models to build web apps for multimodal cultural-heritage editions. The web pages it produces contain embedded metadata for simple harvesting, and common serialisations such as LIDO for

object data are generated automatically. The accessibility of the data-management interface is handled by TYPO3 itself while the frontend uses its own set of UI elements with as little JavaScript as possible. Conservative frontend technologies with up-to-date underpinnings are a central design decision to enable the software's use in semantic portals that follow the Sampo Model principles (cf. Hyvönen 730-735). This focus on RDF-compatibility and accessibility, however, also runs counter to the reactive interfaces that the private sector currently focuses on.

Bibliography

- Comité international d'histoire de l'art and Union Académique Internationale. "Corpus Vitrearum: Guidelines." *Corpus Vitrearum International*, 4th edition, Troyes, 2016, <https://cvi.cvma-freiburg.de/documents/CVGuidelines.pdf>.
- Directorate-General for Research and Innovation (European Commission). "European collaborative cloud for cultural heritage." Publications Office of the European Union, 29. Jul. 2022, DOI: [10.2777/83783](https://doi.org/10.2777/83783).
- Europeana. "Definition of the Europeana Data Model." *Europeana Pro*, v. 5.2.8, 6 Oct. 2017, https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Definition_v5.2.8_102017.pdf.
- Frodl-Kraft, Eva. "Der Corpus Vitrearum Medii Aevi: Ein Rückblick." *Glas – Malerei – Forschung: Internationale Studien zu Ehren von Rüdiger Becksmann*, edited by Hartmut Scholz, Ivo Rauch, and Daniel Hess, Deutscher Verlag für Kunstwissenschaft, 2004, pp. 13-21.
- Hyvönen, Eero. "Digital Humanities on the Semantic Web: Sampo Model and Portal Series." *Semantic Web*, vol. 14, no. 4, 2023, pp. 729-744, DOI: [10.3233/SW-223034](https://doi.org/10.3233/SW-223034).
- NFDI4Culture. "Public SPARQL Endpoint." *NFDI4Culture*, 2024, <https://nfdi4culture.de/resources/knowledge-graph.html>.
- Steller, Jonatan Jalle. "Cultural Heritage Framework." *GitHub*, 2024, <https://github.com/digicademy-chf>.
- Steller, Jonatan Jalle, Sarah Pittroff, Matthias Knorr, Isabelle Lecocq, Armanda Rodrigues, Joseph Spooner, and Karin von Wartburg. "Matching Metadata: Recommendations for the Digital Resources of the Corpus Vitrearum International." *Zenodo*, v. 1.0.0, DOI: [10.5281/zenodo.10829493](https://doi.org/10.5281/zenodo.10829493).
- Tietz, Tabea, Oleksandra Bruns, Linnaea Charlotte Söhn, Julia Tolksdorf, Etienne Posthumus, Jonatan Jalle Steller, Heike Fliegl, Ebrahim Norouzi, Jörg Waitelonis, Torsten Schrade, and Harald Sack. "From Floppy Disks to 5-Star LOD: FAIR Research Infrastructure for NFDI4Culture." *Workshop on Metadata and Research (Objects) Management for Linked Open Science 3*, 2023, pp. 1-12, DOI: [10.4126/FRL01-006444986](https://doi.org/10.4126/FRL01-006444986).