Placeholder Name of the conference

Placeholder Session title

https://doi.org/10.52825/xxxx..... DOI placeholder (WILL BE FILLED IN BY TIB Open Publishing)

© Authors. This work is licensed under a Creative Commons Attribution 4.0 International License

Published: (WILL BE FILLED IN BY TIB Open Publishing)

Harmonizing Heritage

Towards FAIR Digital Objects in KIK-IRPA's Intervention Dossiers

Wim Fremout[1,*] https://orcid.org/0000-0002-1684-377X, Stephanie Buyle[1] https://orcid.org/0000-0002-1481-8749, and Véronique Van der Stede[3] https://orcid.org/0000-0003-4102-6293

[1]Royal Institute for Cultural Heritage (KIK-IRPA), Belgium

*Correspondence: Wim Fremout, wim.fremout@kikirpa.be

Abstract. The Royal Institute for Cultural Heritage (KIK-IRPA) manages a critical 75-year archive of intervention dossiers, encompassing diverse data from conservation and material science. To overcome the challenges of preserving and ensuring the reuse of this complex, evolving data, we developed a framework for FAIR Digital Objects (FDOs). This involved creating 40 data models based on existing metadata schemas (E-RIHS, IRUG, IGSN, PIDINST) to formally document data provenance and content. The chosen repository, Cordra, utilises the Handle System and ElasticSearch for management and search. This architecture aligns with FDO configuration type 2, employing both metadata-only and combined DOs. This project establishes a vital paradigm shift towards machine-actionable data and will integrate the intervention dossiers into KIK-IRPA's institutional platform, BALaT, ensuring complete FAIRness and contextualisation for the heritage science community.

Keywords: Heritage Science, Data Model, Metadata schema, FAIR Digital Object

1. Introduction

The Royal Institute for Cultural Heritage (KIK-IRPA) serves as a scientific institution dedicated to the preservation of Belgium’s artistic, historical and monumental heritage. In this capacity, KIK-IRPA maintains an invaluable, continuously expanding archive of documentation, or intervention dossiers, compiled during conservation-restoration campaigns, material science investigations and art-historical research. These comprehensive dossiers encompass a diverse range of data, including technical reports, photographic documentation, scientific imaging, material analysis data and physical samples (e.g. cross-sections). As indispensable resources for future preservation and scholarly pursuits, these dossiers form an intrinsic part of a heritage object's legacy. However, the archive’s 75-year evolution, spanning from entirely paper-based records to born-digital data, presents a significant and unique challenge for effective documentation and accessibility.

In a series of sequential initiatives over recent years, we have explored novel strategies to formally document the diverse data types within the intervention dossiers and to facilitate the meaningful correlation and contextualisation of these data. The overarching goal is to effectively publish, share, and ensure the long-term archiving and reusability of the data within their original scientific context.

In a series of sequential initiatives over the past few years, we have explored ways to document the various data types in the intervention dossiers and to facilitate the correlation of data, in order to effectively publish, share and archive the data within their context.

2. Data Modelling and Implementation Strategy

To formalise the documentation, we have developed data models for over 40 distinct data types identified within the dossiers. This work is informed by and contributes to analogous efforts within the European Research Infrastructure for Heritage Science (E-RIHS) and the European Cloud for Heritage OpEn Science (ECHOES) project. Crucially, these models draw on a range of established schemas:

- Cross-domain metadata schemas, e.g. DataCite's Metadata Schema [1].
- Domain-specific models, e.g. the metadata requirements from the Infrared and Raman Users' Group (IRUG) for spectral data [2].
- Resource-specific persistent identifiers (PIDs) metadata models, e.g. International Generic Sample Numbers (IGSN) [3] and the PIDINST schema for instruments [4].

Each data schema is formally defined using JSON Schema specifications (currently Draft-07, for practical reasons). These schemas employ internal references to establish clear relations between entities, as conceptually outlined in Figure 1. Entities marked with an asterisk in the figure represent a group of similar data types (e.g. *Optical Microscopy*, *Scanning Electron Microscopy, X-ray Fluorescence Spectrometry*, etc. are represented as *DataAcquisition* type).

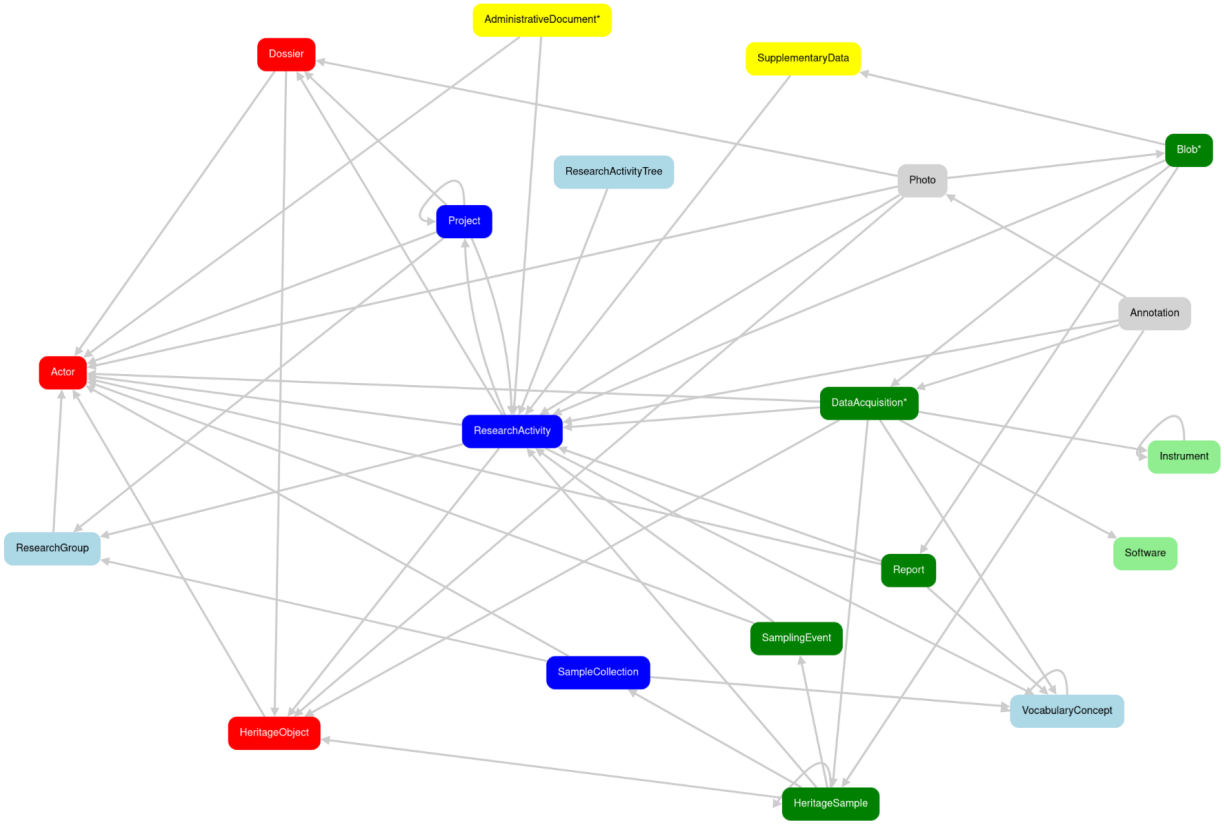

Figure 1. *Relations between the different data types that are part of or related to intervention dossiers.*

Cordra has been selected as the repository solution for managing and publishing the resulting data and metadata. Its native support for the Handle System and straightforward integration with ElasticSearch are significant operational benefits, positioning the repository for full compliance with FAIR Digital Object (FDO) requirements in the future.

The implementation of this system represents a substantial paradigm shift from existing archival workflows, necessitating the development and phased rollout of new software tools. These tools are designed to assist researchers in the creation of individual Digital Objects (DOs) for newly generated data. However, the complete, item-by-item conversion of the historical archive is constrained by time and budget. Consequently, a strategic decision was made to capture and document existing historical intervention dossiers through a limited, aggregated set of DOs, while ensuring the necessary granularity is maintained to allow for future, on-demand segmentation into finer-grained FDOs.

3. Towards a FAIR Digital Objects Repository

While the initial impetus was the development of a comprehensive solution for documenting, publishing, and archiving intervention dossiers, the resulting Digital Object-based system already addresses several key requirements of the FDO Forum standards [5]. This includes the assignment of globally unique Persistent Identifiers (PIDs) to each DO and the use of the well-defined, community-informed metadata schemas detailed above.

The applied methodology, which incorporates both metadata-only DOs and combined metadata/data DOs, aligns with the conceptual framework of FDO configuration type 2 in which the metadata, bit sequence and PID record are contained within the same object [6], thus supporting the full machine-actionability required by the FAIR principles.

The ultimate objective is the seamless integration of the intervention dossiers into KIK-IRPA's institutional knowledge platform, BALaT, which is currently undergoing a comprehensive reconstruction. This integration is anticipated by the end of 2026. BALaT already provides unified access to diverse data sources, including KIK-IRPA's photographic collections (via IIIF), the heritage objects database, and the institute's library catalogue. The unified interface, powerfully backed by ElasticSearch's robust indexing and search capabilities, will guarantee the complete FAIRness and crucial contextualisation of KIK-IRPA's intervention dossiers, marking a significant and timely step towards a harmonised and accessible heritage science knowledge repository.

Data availability statement

KIK-IRPA's Cordra instance is available online on https://cordra.kikirpa.be. Metadata schemas as digital objects within Cordra, and can be consulted on https://cordra.kikirpa.be/#objects/?query=type:Schema.

Author contributions

W.F.: Conceptualization, Methodology, Writing – original draft, Supervision, Funding Acquisition. S.B. and V.V.D.S: Methodology, Investigation.

Competing interests

The authors declare that they have no competing interests.

Funding

This work discusses part of the outcomes of the H-SEARCH project, funded by the Belgian Science Policy PPS (BELSPO) under grant number RT/23/H-SEARCH.

References

[1] DataCite Metadata Working Group, “DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs v4.6”, Dec. 2024, doi: https://doi.org/10.14454/MZV1-5B55.

[2] B. A. Price, B. Pretzel, S. Q. Lomax, C. Davis, and J. H. Carlson, “Revised JCAMP-DX Spectral File Format for Submissions to the Infrared & Raman Users Group (IRUG) Spectral Database”. http://irug.org/uploads/documentation/irug-jcamp-dx-revised-white-paper-text-only-with-2b-version-1-26-sept-2013.pdf (2025-11-01)

[3] IGSN–DataCite Crosswalk and Metadata Management Working Group, “IGSN ID Metadata Recommendations”. DataCite Support. https://support.datacite.org/docs/igsn-id-metadata-recommendations (2025-11-01)

[4] R. Krahl *et al.*, “Metadata Schema for the Persistent Identification of Instruments”, 2021, doi: https://doi.org/10.15497/RDA00070.

[5] A. Ivonne *et al.*, “FDO Forum FDO Requirement Specifications”, Jan. 2023, doi: https://doi.org/10.5281/zenodo.7782262.

[6] L. Larry *et al.*, “FDO Forum FDO Configuration Types”, Oct. 2022, doi: https://doi.org/10.5281/zenodo.7825703.