



NFDI4Objects

Research Data Infrastructure
for the Material Remains of
Human History

TRAIL 2.3:

Towards interoperability of archaeological and natural history object data

Partner **Lead:** RGZM;

Co-applicants: RGZM, DAI, MPIWG, SPK (SMB)

Participants: Friedrich-Alexander-Universität Erlangen-Nürnberg / Competence Unit for Research Data and Information (CDI-FAU), Museum für Naturkunde Berlin (MFN), RWTH Aachen University / Chair for Architectural History;

External members: NFDI4BioDiversity, NFDI4Earth, NFDI4Culture, NFDI Cross-Cutting Topics

Contact Dr Christian Bölling (MFN) / christian.boelling@mfn.berlin,

Florian Thiery M.Sc. (RGZM) / thiery@rgzm.de

Summary

Interdisciplinary collection-based research is impeded by the fact that object data in collections are created with different domain-specific perspectives. This heterogeneity prevents the linking of object data, even if they have subject-related references to each other, and affects the following phases in the research data lifecycle: describe, discover, integrate. This TRAIL focuses on one aspect of this problem: interoperability of provenance descriptions of objects from archaeological and natural history collections. Based on domain-specific data models, the subject of this TRAIL is the development of a conceptual data model realizing semantic interoperability for collection object provenance data for the different disciplines, its formal description using the Web Ontology Language (OWL) and its reference implementation in WissKI.

Description

The objects held in archeological and natural history (botany, zoology, palaeontology and geology) as part of humankind's cultural and natural history heritage are an important basis for vouchered knowledge about natural and human history, and interactions between them. In this context, the boundaries between the disciplines of natural history and archaeology are fluid, especially in the light of new analytical methods and digital evaluation strategies, and multiple usage perspectives on the data about collection objects. However, concepts and methods of managing collection data and representing documented and analytically derived object data have so far been developed largely independently of each other, making efficient, cross-disciplinary use of data a challenge. An essential prerequisite for interoperability, both on a conceptual and operational level, is a model realizing semantic compatibility as the basis for representing specific objects, their properties and relations to other objects. While there are some detailed approaches to modelling within individual domains – [CIDOC-CRM](#) (incl. relevant extensions CRMarchaeo, CRMsci, CRMgeo, CRMba), [Lightweight Information Describing Objects \(LIDO\)](#), [Access to Biological Collections Data \(ABCD\)](#), [Extension For Geosciences \(EFG\)](#), [Darwin Core](#), [Dublin Core](#) – we still need abstracting models that enable integration of object data from natural history and archaeological contexts. Due to the complexity of this task, this TRAIL focuses on the contexts of origin and ownership, which are factors of central importance and common to all objects regardless of their domain affiliation, and are highly relevant to all areas of contemporary collection-based research. On the basis of existing, conventionally used domain models, the TRAIL will develop a conceptual model that establishes semantic interoperability for collection object provenance data from natural history and archaeological domains and can be used to connect to further domain- and collection-specific semantic models. Model development includes specification, documentation and prototypical implementation. The project is related to other TRAILS that aim to consolidate terminological resources, in particular TRAILS 2.1, 2.4 and 4.2. The model specification in the Web Ontology Language (OWL), the reference implementation in the Software Application Service (SAS) WissKI, and the accompanying documentation result in an Interoperable Dataset Service (IntS). The following Data Services (DaS) from N4O partners are included in this TRAIL: From the RGZM data on archaeological collection objects, particularly from the BMBF-funded project *African Red Slip Ware digital (ARS3D)*, in which data modelling is carried out based on CIDOC-CRM. Data on natural history objects from the Museum für Naturkunde Berlin are used, particularly collection object data from the DFG-funded project *Das Fenster zur Natur und Kunst – eine historisch-kritische Aufarbeitung der Brandenburgisch-Preußischen Kunstkammer (Window on Nature and Art – a historical-critical study of the Brandenburg-Prussia Kunstkammer)* based on a project-specific application ontology for source-supported documentation of origin, ownership and collection embedding. The solutions developed in this TRAIL will, for the first time, enable an interoperable representation of object properties relevant to object provenance across the boundaries of the two collection knowledge domains.

Relevance

The following aspects of the research data lifecycle are addressed: describe, discover, integrate.

Stakeholders at different points in the data lifecycle will benefit from the developed resources, including scientists and data curators, infrastructure providers and system integrators. A semantically interoperable framework for representing object provenances from different disciplines will improve data integration in a wide range of contexts. This includes the standardized representation of contexts of origin (related to TA1) and of statements on object biography derived using analytical methods (related to TA3), also fostering identification of synergies in samples and sample data management. The transferability of this approach to temporal properties of objects will be tested for dating questions in the context of building research (related to TA4).

The TRAIL is particularly relevant for aggregating building blocks in TA6, especially for M6.2 (improving data integration through data harmonisation) and for M6.3 (modelling the interdisciplinary research cycle on objects). Synergies with the topics and goals of TA6 will arise from the FAU's close cooperation with the Germanisches Nationalmuseum Nürnberg and the *Nuremberg TimeMachine* project on object biographies. The Chair of Early Christian Archaeology at FAU will also contribute to M6.4 (Education and training), as the reference implementation of WissKI will be taught and tested in courses there. The resources developed in this TRAIL can be used to improve the maturity of datasets according to the FAIR principles, in particular the indicators RDA-F2-01M, RDA-I1-01M and RDA-I3-03M. Semantic and technical interoperability of research data and data infrastructures is a central goal of the NFDI. This TRAIL thus contributes to a pivotal concern of N4O and the NFDI as a whole for the field of natural history and archaeological collection data and collection-based research. The results from this TRAIL will be useful for cooperation with other NFDI consortia and in the context of the NFDI cross-cutting topics, in particular the priority area Terminologies, Metadata and Provenance, to realize further solutions for semantic and technical interoperability of data.

Deliverables

- *White Paper* on requirements for the data model
- *White Paper* on testing the model using provenance data on objects from the MfN (as an example of a natural history collection) and the RGZM (as an example of an archaeological collection)
- Formal description of the model in OWL (**IntS**)
- Reference documentation of the model (*Blue Paper*)
- Prototype of the model in [WissKI](#) (**SAS**)
- **Commons:** White Paper / Blue Paper / OWL

Work plan

- Year 1, Month 6: Collection of use cases, evaluation of existing domain models and creation of requirements analysis
- Year 1, Month 12: Creation of the basic version of the model
- Year 2, Month 6: Evaluation, formal specification and prototype of the model

Funding / co-applicants' own contribution

Besides in-kind services from the participants, the project requires a total of 1 FTE for 18 months:

- FAU/CDI (0.5 FTE, 18 months)
 - collection of use cases, legacy model evaluation, requirements analysis
 - definition and formal specification of the model
 - provide the [WissKI Distillery](#) and [WissKI Cloud](#)
- MFN (0.5 FTE, 18 months)
 - collection of use cases, requirements analysis
 - definition and formal specification of the model
 - expertise in data modelling of complex natural history objects
 - expertise on standards for the representation and exchange of natural history collection data
- Chair for Architectural History, RWTH Aachen University
 - evaluation of the model in architectural history use cases
- RGZM
 - collection of use cases
 - expertise on specialist systems for archaeological collections
 - ARS3D project data with CIDOC-CRM modelling of African Red Slip Ware Collection data

- DAI
 - expertise in modelling archaeological objects and object biographies (in kind)
- MPIWG
 - expertise in modelling objects and research related questions
- SPK
 - expertise on specialist systems for archaeological collections
- Chair for Early Christian Archaeology, FAU Erlangen
 - seminars and courses on reference implementation of WissKI

*FAIR*¹ F2:RDA-F2-01M; I2:RDA-I2-01M; I3:RDA-I3-03M

TRAILS part of TRAIL 2.1, related with TRAIL 2.2, TRAIL 2.4

¹ Nach Tabelle 1 von Bahim, C., Casorrán-Amilburu, C., Dekkers, M., Herczog, E., Loozen, N., Repanas, K., ... Stall, S. (2020). The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments. *Data Science Journal*, 19(1), 41. DOI: <http://doi.org/10.5334/dsj-2020-041> [cc by 4.0](https://creativecommons.org/licenses/by/4.0/)