NFDI4Objects
Research Data Infrastructure for the Material Remains of Human History

TRAIL 5.1:

Rich Metadata Discovery Service. Prototype: GeoNumismatics

Partner **Lead**: Verbundzentrale des GBV (VZG: Steilen)

Co-applicants: Landesamt für Denkmalpflege Baden-Württemberg, University of Bonn, RGZM, SPK, GDKE, German Archaeological Institute,

Participants: Thüringisches Landesamt für Denkmalpflege und Archäologie (Schlapke), University of Heidelberg (Börner), University of Frankfurt (Tolle), Römisch-Germanische-Komission (Wigg-Wolf), FIZ Karlsruhe (Razum), Numismatic Commission (Weisser)

External members: N4O Linked Open Data Cluster, NFDI4Culture, Text+, Wikimedia Deutschland (Team Wissenschaft), nomisma.org

Contact Gerald Steilen / steilen@gbv.de;

Susanne Börner / susanne.boerner@zaw.uni-heidelberg.de;

Mario Schlapke / SchlapkeM@tlda.thueringen.de

Summary

The aim of the TRAIL is, very early on, to provide a discovery service that includes geodata on the discovery site. During the TRAIL, this service will primarily contain data on numismatic finds, for which all necessary prerequisites exist (FAIR, location of finds, standardisation). Just nine months after the project began, the discovery service is being used as Open Educational Resources (OER) in teaching at the Universität Heidelberg. Due to ongoing evaluation and use of the service, we can optimise it according to the requirements of the professional community at an early stage.

The Rich Metadata Discovery Service with the prototype GeoNumismatics (data on numismatic finds) is based on close networking of data from five TAs within N4O.

From the outset, the prototype will fully meet all 20 essential criteria of the FAIR Data Maturity (FDM) Model[1] plus six more (see above). Of course, the prototype is not limited to numismatic finds. Rather, it demonstrates the potential of the service. The permanent operation of the Rich Metadata Discovery Service is ensured by the VZG and furthermore fulfils all necessary requirements for the establishment and operation of a knowledge graph[2] for NFDI4Objects. For this purpose, disjoint data sources from distributed repositories must first be integrated into the system published interoperably on a platform and subsequently offered for analysis in a user-friendly manner. According to the FAIR Data Maturity Model, a Rich Metadata Discovery Service[3] is offered as a free web service for research via an API (application programming interface). Using the front end, laypersons can search the data with manifold queries. The prototype is the first interface in which data from heterogeneous systems can be searched for complex questions. Integration into the numismatic e-learning platform NumiScience.de and university teaching guarantees knowledge transfer into the specialist community. Extensive practical tests are complemented by linking with literature holdings, incl. digital copies on the basis of the GVK and ZENOM IDs.

Description

Existing services such as BoDeOn, KuLaDig and denkmal.nrw from LVR can be integrated. In addition, coli-conc (VZG), is used as Interoperable Data Set Services (IntS). The technical basis for the infrastructure is the search engine K10plus-Zentral[4] and Lukida[5] is used for the user interface. Within the TRAIL, standard datasets on coin finds are used. These are supplied by Antike Fundmünzen in Europa (Ancient Coins Found in Europe, AFE[6], incl. the Heidelberg collection AFE4HD[7]) for Antiquity and the Fundmünzkatalog (Coin Find Catalogue)[8] of the Numismatischen Kommission der Länder (Numismatic Commission of the German Federal States, NK)[9] for the Middle Ages and Modern Era. By means of a graph database, standardised APIs, and the numismatic ontology the aforementioned datasets are made accessible. A GUI as well as the integration into the e-learning platform NumiScience and university teaching guarantee continuous evaluation and transfer into the scientific community. In particular, the following are used and taken into account:

a. Infrastructure for Spatial Information in the European Community (INSPIRE)[10]

b. Representational State Transfer (REST) web service for distributed systems

[1] https://doi.org/10.15497/rda00050
[2] https://en.wikipedia.org/wiki/Knowledge_graph
[3] see F2 in Table 1: https://datascience.codata.org/articles/10.5334/dsj-2020-041/
[4] https://verbundwiki.gbv.de/display/VZG/K10plus-Zentral
[5] https://www.lukida.org
[6] http://afe.dainst.org
[7] afe.dainst.org
[8] https://kenom.gbv.de/fundkomplexe/
[9] http://www.numismatische-kommission.de
[10] https://inspire.ec.europa.eu/

c. Shibboleth[11] (a distributed authentication and authorisation procedure for web applications) to map legal usage restrictions relevant for N4O

d. N4O standards for metadata sets formulated in TA1 and TA2 in dialogue with TA5

e. N4O standards for geodata referencing (TA4)[12]

Resources to be developed:

f. Standards for using authority files

g. Standardised, project-relevant vocabularies (e.g. Nomisma)

h. Usage scenarios of standardised exchange formats

Challenges include different data sources and genesis, and the need for continuous data transfer into the discovery service. This is the technical prerequisite for both vector space-based search methods and graph databases, which will be used here. The discovery service works with a knowledge graph to show initially non-obvious relationships in the data. For this purpose, proven technical concepts and solutions are applied to new subject domains. For the first time, this enables uniform query and display of supra-regional and cross-epochal complexes of coin finds based on their subject-related, chronological and geographical metadata. Such queries are not easy or user-friendly with table-based data systems.

Relevance

The meta and research data are integrated into the discovery service, where they can be analysed via query modes to discover new relationships in the data. This is published formally via the user interface and NumiScience.de containing necessary authentication and authorisation mechanisms (protected data). The user base is extraordinarily broad. Including the numismatic research community, museum curators (scientists), heritage management authorities, university staff, volunteers and interested citizens.

Therefore, the system is made available for re-use in external infrastructures as well as in developed, generalizable data transfer routines. If a cross-consortium Knowledge graph basic service is developed, this TRAIL will provide it as a core service.

The Discovery TRAIL itself can be extended also as a cross-consortium Knowledge graph basic service. Experience of maintaining knowledge graphs and handling of research data will be exchanged with NFDI4Memory and NFDI4Culture. At regular

[11] https://www.shibboleth.net/
[12] Other relevant data sources, especially from the Federal State Heritage Offices, will be integrated via the INSPIRE-compliant services of TA4.

meetings of (approved and co-applicant) Göttingen NFDI consortia, technical topics for the (re-)use of data are already being discussed.

The discovery service makes existing data much more findable, as it is merged into a new service, searchable together and redistributed. Accessibility is achieved through integration into the e-learning platform and university teaching. Interoperability is ensured through the integration of diverse data sources in one system. An application programming interface (API) guarantees use in external systems.

As an example of best practice, the discovery service provides a time, space and object reference for numismatic finds. Based on this TRAIL, other data sources (e.g. TRAIL 4.3. planned in TA4) can be integrated in the future. In addition, the TRAIL fulfils all requirements for setting up and operating a knowledge graph for NFDI4Objects.

Deliverables

The graph database guarantees a flexible basis for pre-processing metadata for the vector space-based search engine SolrCloud, which is the backbone of VZG's discovery service, K10plus-Zentral. The Lukida software provides a flexibly customisable web front end as well as API for integrating the research platform into third-party systems.

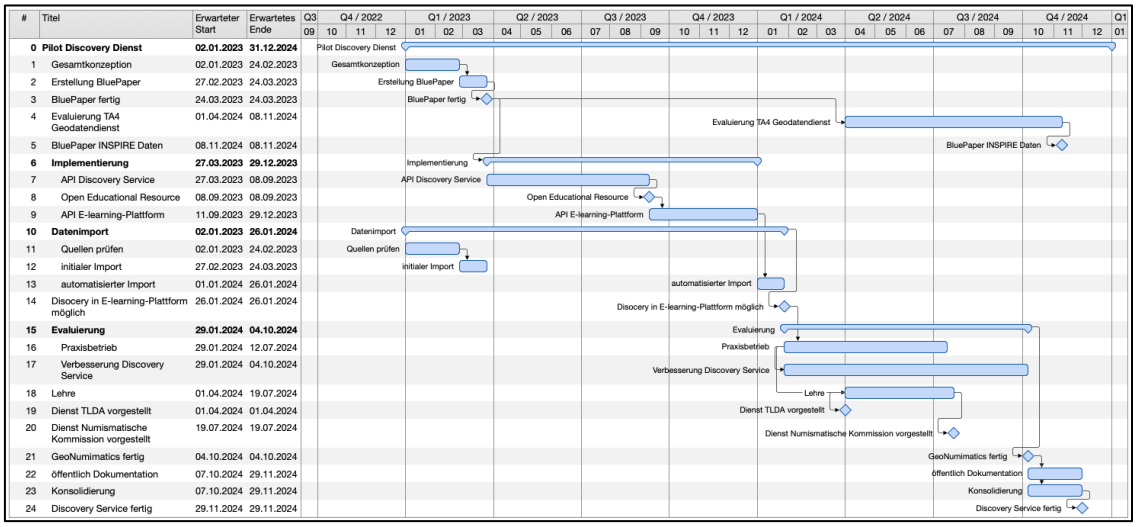
As part of the trail, 2 blue papers will be produced: The first blue paper will will describe the use of machine-readable metadata via an API. Technical specifications will describe both the data flow into the system itself and the use of this discovery service in third-party systems. Naturally, the requirements for decentralised rights management are taken into account. NumiScience serves here as a prototype for the integration of this service into third-party systems. A second blue paper will analyse the possibilities of integrating the TA4 geoservice, which will be completed by the fourth quarter of 2024.

The discovery service, which complies with the FAIR principles, will be integrated into the numismatic e-learning platform NumiScience.de[13] via REST API. In addition, "tandem courses" with representatives from IT and numismatics are offered to strengthen the Digital Humanities. This ensures parallel evaluation and closely interlinked transfer of the TRAIL to the subject community and student body.

[13] The cross-epoch platform NumiScience.de of the Numismatische Verbund in Baden-Wuerttemberg (NV BW) is aimed at pupils, students, interested laypersons, collectors, museums, criminologists from crime prevention and cultural heritage preservation (such as customs investigators, police), as well as detectorists authorised by state heritage offices and, of course, specialist scientists. The numismatic content, consisting of texts, images, digitised sources, films, podcasts, interactive identification exercises, quiz options, etc., is labeld and freely available for long-term use free of charge.

Work plan & requested funding

Duration: 2 years



#	Titel	Erwarteter Start	Erwartetes Ende
0	**Pilot Discovery Dienst**	**02.01.2023**	**31.12.2024**
1	Gesamtkonzeption	02.01.2023	24.02.2023
2	Erstellung BluePaper	27.02.2023	24.03.2023
3	BluePaper fertig	24.03.2023	24.03.2023
4	Evaluierung TA4 Geodatendienst	01.04.2024	08.11.2024
5	BluePaper INSPIRE Daten	08.11.2024	08.11.2024
6	**Implementierung**	**27.03.2023**	**29.12.2023**
7	API Discovery Service	27.03.2023	08.09.2023
8	Open Educational Resource	08.09.2023	08.09.2023
9	API E-learning-Plattform	11.09.2023	29.12.2023
10	**Datenimport**	**02.01.2023**	**26.01.2024**
11	Quellen prüfen	02.01.2023	24.02.2023
12	initialer Import	27.02.2023	24.03.2023
13	automatisierter Import	01.01.2024	26.01.2024
14	Disocery in E-learning-Plattform möglich	26.01.2024	26.01.2024
15	**Evalulierung**	**29.01.2024**	**04.10.2024**
16	Praxisbetrieb	29.01.2024	12.07.2024
17	Verbesserung Discovery Service	29.01.2024	04.10.2024
18	Lehre	01.04.2024	19.07.2024
19	Dienst TLDA vorgestellt	01.04.2024	01.04.2024
20	Dienst Numismatische Kommission vorgestellt	19.07.2024	19.07.2024
21	GeoNumimatics fertig	04.10.2024	04.10.2024
22	öffentlich Dokumentation	07.10.2024	29.11.2024
23	Konsolidierung	07.10.2024	29.11.2024
24	Discovery Service fertig	29.11.2024	29.11.2024

Work plan: (according to Gantt Chart above)

Milestone	Title	Expected launch	Expected end
	Discovery service	02 January 2023	31 December 2024
	Overall concept (including data flows)	02 January 2023	24 February 2023
	Blue paper on data flows	27 February 2023	24 March 2023
1	Blue paper on data flows completed	24 March 2023	24 March 2023
	Evaluation of TA4 Geodata Service	01 April 2024	08 November 2024
2	Blue paper on geoservice	08 November 2024	08 November 2024
	Implementation	27 March 2023	29 December 2023
	Discovery service API	27 March 2023	08 September 2023

3	OER	08 September 2023	08 September 2023
	API for e-learning platform	11 September 2023	29 December 2023
	Data import	02 January 2023	26 January 2024
	Check sources	02 January 2023	24 February 2023
	Initial import	27 February 2023	24 March 2023
	Automated import	01 January 2024	26 January 2024
4	Discoverable in e-learning platform	26 January 2024	26 January 2024
	Evaluation	29 January 2024	04 October 2024
	Practical application	29 January 2024	12 July 2024
	Improvement	29 January 2024	04 October 2024
	Teaching	01 April 2024	19 July 2024
5	TLDA service presented	01 April 2024	01 April 2024
6	Numismatic Commission Service presented	19 July 2024	19 July 2024
7	GeoNumimatics completed	04 October 2024	04 October 2024
	Public documentation	07 October 2024	29 November 2024
	Consolidation	07 October 2024	29 November 2024
8	Discovery service completed	29 November 2024	29 November 2024

Funding (for two years):

1. Co-applicants' own contribution
    - VZG: 12 months Full-time equivalent (FTE)
        - Knowledge
        - Services: Cocoda[14], coli-conc[15], DANTE[16], BARTOC[17], GOKb[18]
        - Software as a Service (SaaS): Lukida[19], K10plus-Zentral[20]
        - Server: Linux High performance and high availability cluster
        - Operation: Monitoring, backup, maintenance, fault acceptance (hot-line, ticket system)
    - Universität Heidelberg:
        - Knowledge
        - Editing, coordinating and documenting the e-learning platform Nu-miScience.de of the Numismatische Verbund in Baden-Württemberg
        - University teaching involving the TRAIL
        - AFE4HD
    - TLDA:
        - Knowledge
        - Technical supervision of the NK's coin find catalogue
        - Provision of Thuringia data for the coin find catalogue
        - Communication of project goals and statuses in the specialist community
    - NK:
        - Knowledge
        - Online coin find catalogue
        - Communication of project goals and results in the specialist community
    - RGK:
        - Knowledge
        - AFE
    - FIZ:
        - Knowledge

[14] https://coli-conc.gbv.de/cocoda/
[15] https://coli-conc.gbv.de
[16] https://api.dante.gbv.de
[17] https://bartoc.org
[18] https://gokb.org
[19] https://www.lukida.org
[20] https://www.lukida.org/k10plus-zentral/

2. Financial support:

    - VZG: 36 months FTE for 2 years (2023 and 2024)

      Taking into account 30 days of leave of an FTE per year and public holidays, the 36 months of FTE work must be financed for a period of two years. This would be, for example, the FTE of two persons with a 0.75 position.

FAIR[21]	*F1:RDA-F1-01M; F1:RDA-F1-01D; F1:RDA-F1-02M; F1:RDA-F1-02D; F2:RDA-F2-01M; F3:RDA-F3-01M; F4:RDA-F4-01M; A1:RDA-A1-02M; A1:RDA-A1-02D; A1:RDA-A1-03M; A1:RDA-A1-03D; A1:RDA-A1-04M; A1:RDA-A1-04D; A1:RDA-A1-05D; A1.1:RDA-A1.1-01M; A1.1:RDA-A1.1-01D; A1:2: RDA-A1.2-02D; A2:RDA-A2-01M; I1:RDA-I1-02M: I3:RDA-I3-01M; I3:RDA-I3-01D; I3: RDA-I3-03M; R1.1:RDA-R1.1-01M; R1:RDA-R1.1-02M; R1.1:RDA-R1.1-03M; R1.3:RDA-R1.3-01M; R1.3:RDA-R1.3-01D*
TRAILS	keine

[21] 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