OstData – Building a Research Data Service for Enabling Interdisciplinarity and Regional Collaboration in Central, East, and Southeast European Studies

Frank, Ingo
frank@ios-regensburg.de
Leibniz Institute for East and Southeast European Studies, Regensburg, Germany

Introduction

OstData is a research data service for Central, East, and Southeast European Studies. It is being developed in a joint project funded by the German Research Foundation (DFG) [1]. The poster demonstrates how our metadata schema supports interdisciplinary research by enabling detailed descriptions of what methods were applied to create research data.

Profile and Objectives

OstData supports the publication of research data from multi- and interdisciplinary research on the following countries (see also map in fig. 1): Albania, Belarus, Bosnia-Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Estonia, Finland, Hungary, Kosovo, Latvia, Lithuania, Modern Greece, Moldavia, Montenegro, North Macedonia, Poland, Romania, Russia, Serbia, Slovakia, Slovenia, Ukraine.

Fig. 1. Regional coverage of the OstData research data service

Research on the regions of Central, East, and Southeast Europe is conducted within humanities and social sciences, especially in relation to the following disciplines: Archaeology, Economics, Ethnology/Anthropology, Historical Sciences, Linguistics and Literature, Political Science, Sociology.

The most relevant objectives of OstData as seen from the perspective of research infrastructure are

- to provide a metadata schema for the description of research data, thus enabling these to be searched and re-used,
- to integrate institutional research data repositories [2] by ingesting metadata for its central search platform,
- to enable publication and archiving of research data by providing a reliable infrastructure and the necessary incentives, and
- to operate an accessible, user-friendly, and structured search interface for research data originating from Germany.

Approach and Methods

According to the following definition of interdisciplinarity, interdisciplinary research is not only characterized by applying a method or theory from one discipline in another discipline, but also by reusing data created in another disciplinary setting.

> *Interdisciplinarity (ID)* is typically characterized by integration of information, data, methods, tools, concepts, and/or theories from two or more disciplines or bodies of specialized knowledge. Proactive focusing, blending, and linking of disciplinary inputs foster a more holistic understanding of a question, topic, theme, or problem by individuals or teams. (Klein, 2014, p. 15)

Unfortunately, research data repositories are not very well suited to the information needs of interdisciplinary researchers. Metadata about published research data often lacks methodological information about the creation of research data and the theoretical and disciplinary context in which the data were created. In order to support researchers in finding relevant research output in form of research data to be (re)used in their interdisciplinary research, information system have to provide interdisciplinary perspectives on data and the tools and methods which were used to create the data. [3]

The OstData metadata schema (ODMS) is based on essential elements of the DataCite metadata schema (DataCite Metadata Working Group, 2021) and extends it with additional fields to enable researchers to describe their research data in more detail with a focus on the subject of the data and the methods which were used to create the data. As an application profile (Heery / Patel, 2000) it also reuses properties from the Data Catalog Vocabulary (DCAT) (Albertoni et al., 2020). In order to support interdisciplinary research data management, the ODMS follows a faceted classification approach. To support especially interdisciplinary researchers, the ODMS provides so-called zones and subzones for discipline/domain-specific information, whereat the zone ‘Subject Indexing’ and the zone ‘Discipline-specific Indexing’ with the additional subzones for the social sciences, digital editions, and linguistics are most important.

In combination with the zone ‘Subject Indexing’ the zone ‘Discipline-specific Indexing’ allows a kind of phenomenon-based knowledge organization (Szostak et al., 2016). Different disciplinary classification systems and thesauruses are used in order to grasp various disciplinary perspectives (see table 1 for details).

Identifier for Dataset	
Authorship & Research Information	
Title Information	
Subject Indexing	
geographicClassification	spatial coverage of the data (BSB-DDC)
timePeriodClassification	temporal coverage of the data (BSB-DDC)
subjectClassification	topical coverage of the data (BSB-DDC)
subjects	GND subject classification
subjectNormed/subjectNormedType	AAT, ELSST, Iconclass, LCSH, STW, Wikidata
subjectClassification/subjectClassificationType	CESSDA Topic Classification, JEL Classification
timePeriodDescription	textual description of temporal coverage of the data
timePeriodDates	time interval describing the temporal coverage of the data
geoLocation	location or place of the spatial coverage of the data
Discipline-specific Indexing	
discipline	Destatis classification system for scientific disciplines
method/methodType	humanities: TaDiRAH economics: JEL Classification social sciences: DDI Controlled Vocabulary for Mode of Collection
schoolOfThought	theory, theoretical background or paradigm
Discipline-specific Indexing / Social Sciences	
analysisUnit	entry from DDI Controlled Vocabulary for Analysis Unit
timeMethod	entry from DDI Controlled Vocabulary for Time Method
Discipline-specific Indexing / Digital Editions	
textClass	text genre (from GND authority files)
isEditionOfDocument	text basis of the edited text
Discipline-specific Indexing / Linguistics	
Provenance Information & Dataset Relations	
placeOfCollection	place where the data was collected
dataProvenance	textual description of the provenance of the data
Manifestation Description	
fileFormat	entry from PRONOM registry
fileSize	file size in bytes
structureConformsTo	e. g. CIDOC CRM 7.1.2
resourceDimension	e. g. the scale of a map
Legal Information	
license	e. g. a Creative Commons license
accessRights	concept from COAR Access Rights vocabulary

Table 1. Zones and subzones in the OstData metadata schema (ODMS) with a selection of metadata fields

Results and Outlook

Based on its metadata schema OstData provides a faceted classification system (see fig. 2 showing the faceted navigation system currently in its final development stage) including amongst others the Taxonomy of Digital Research Activities in the Humanities (TaDiRAH) [4] and the DDI Controlled Vocabularies (Jaaskelainen et al., 2010) for granular methodological description of research data.

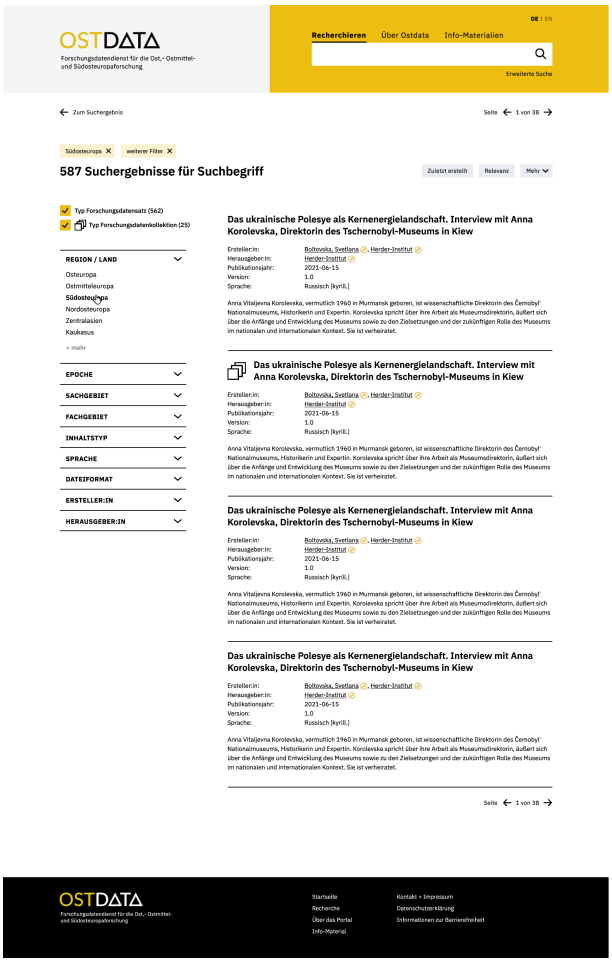

Fig. 2. Web design (sketch) for the faceted navigation system of the OstData discovery system to be deployed on http://www.ostdata.de/

OstData permits metadata curators to use different classification systems and thesauri to describe objects of research and data collection methods. Thereby, researchers can retrieve research data from their disciplinary perspectives. Furthermore, researchers from different disciplinary backgrounds get the opportunity to find and integrate data from other disciplinary contexts into their interdisciplinary research. [5]

In the long run, OstData wants to foster networking of national and international research communities, research infrastructures, funders, etc. During its ongoing development, we work on opportunities for the continuous *FAIR*ification [6] and *FRBR*ization [7] of research data published and provided via the OstData research data service.

Notes

1. The project is a collaboration between the Bavarian State Library (BSB), the Collegium Carolinum – Research Institute for the History of the Czech Lands and Slovakia (CC), the Leibniz Institute for the History and Culture of Eastern Europe (GWZO), the Herder Institute for Historical Research on East Central Europe (HI), and the Leibniz Institute for East and Southeast European Studies (IOS) supported by the DFG grant number 413708228 in the DFG program *Research data and software*

(Scientific Library Services and Information Systems): https://gepris.dfg.de/gepris/projekt/413708228?language=en
2. IOS's LaMBDa research data portal is currently being integrated in OstData's infrastructure: https://lambda.ios-regensburg.de/
3. The FAIR principles R2 and F1 (cf. Wilkinson et al., 2016) are most relevant for that purpose.
4. See an interactive visualization of the taxonomy: https://tadirah.info/pages/Browser.html
5. See our presentation *Inter- und transdisziplinäres Forschungsdatenmanagement in der Ost-, Ostmittel- und Süudosteuropaforschung* (Štanzel et al., 2021) at the FORGE 2021 conference for some more details about the zones and subzones in the metadata schema: https://doi.org/10.5281/zenodo.5519305 (slides)
6. See for example our recent workshop material on making legacy data FAIR by means of RDF Data Cube Vocabulary and SKOS: https://www.osmikon.de/forschungsdaten/digitales-praxislabor-geschichtswissenschaft & https://github.com/ingof-rank/digitales-praxislabor-geschichtswissenschaft
7. See our working paper (Frank / Štanzel, 2022) where we apply the conceptual framework of FRBR to the domain of research data management.

Bibliography

Albertoni, R. et al. (eds) (2020). Data Catalog Vocabulary (DCAT) – Version 2. W3C Recommendation. https://www.w3.org/TR/vocab-dcat-2/.

DataCite Metadata Working Group. (2021). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs. DOI: 10.14454/3w3z-sa82.

Frank, Ingo / **Štanzel, Arnošt** (2022). Forschungsdaten in ihrer Granularität beschreiben. Zenodo. DOI: 10.5281/zenodo.6951770.

Heery, Rachel / **Patel, Manjula** (2000). Application Profiles: Mixing and Matching Metadata Schemas. *Ariadne*, (25). http://www.ariadne.ac.uk/issue/25/app-profiles/.

Jaaskelainen, Taina / **Moschner, Meinhard** / **Wackerow, Joachim** (2010). Controlled Vocabularies for DDI 3: Enhancing Machine-Actionability. *IASSIST Quarterly*, 33(1-2). DOI: 10.29173/iq649.

Klein, Julie Thompson (2014). *Interdisciplining Digital Humanities*. University of Michigan Press. DOI: 10.3998/dh.12869322.0001.001.

Štanzel, Arnošt et al. (2021). Inter- und transdisziplinäres Forschungsdatenmanagement in der Ost-, Ostmittel- und Südosteuropaforschung. In Helling, Patrick / Speer, Andreas / Eide, Øyvind (eds), *FORGE 2021: Forschungsdaten in d en Geisteswissenschaften – Mapping the Landscape – Geisteswissenschaftliches Forschungsdatenmanagement zwischen lokalen und globalen, generischen und spezifischen Lösungen (FORGE 2021)*. Zenodo. DOI: 10.5281/zenodo.5379653.

Szostak, Rick / **Gnoli, Claudio** / **López-Huertas, Maria** (2016). *Interdisciplinary Knowledge Organization*. Springer International Publishing. DOI: 10.1007/978-3-319-30148-8.

Wilkinson, Mark D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(160018). DOI: 10.1038/sdata.2016.18.