



Getting closer to the researcher's desk from acquisition and management to publishing scientific data

Stephan Mäs, Daniel Henzen and Lars Bernard

Most of the available research data infrastructures only support the very last step in the research workflow: the final publication of the research data. The daily work of the researchers that comes prior to the data publication is usually hardly considered. Consequently, the various technical components for data acquisition, transformation, integration, and management are usually not very well connected to data publishing and provisioning services. This has consequences for acquisition of metadata and in particular data provenance, which most researchers experience as a tedious task during the data publication.

The COLABIS project

This submission proposes an integrated software architecture to cover the complete data live cycle. The architecture results from the research project COLABIS (Collaborative Early Warning Information Systems for Urban Infrastructures, https://colabis.de).

It also contains interfaces for interoperable data exchange with external data platforms and applications and introduces a software stack that integrates several open-source products to provide the required functionality. Within COLABIS, a specific focus is laid on the use of standardized interfaces and formats for the exchange of environmental data.



Figure 1: The three pillars of the research data management infrastructure

Researcher's Environment: the Data Management Platform (DMP)

Architecture

The COLABIS architecture is organised as shown in Figure 1. It can be mapped to the regular data life cycle phases (collection / acquisition, storage, curation, qualification, publishing and exploration) and offers a seamless data management environment.

A general aim of the platform is to get as close as possible to data creation and easing and minimizing manual metadata acquisition wherever possible. This is for example done by the consideration of the data driven workflows of the research project (Figure 2). This information is used to structure the folders in the DMP and to automatically annotate the data with provenance information. The provenance information can be explored using a provenance viewer (Figure 3) within the CKAN catalogue.

- Supports the daily work of the researchers prior to data publication
- Scalable, multi-purpose file manager to collect, manage and share data in one common environment
- Supports to enrich and update data, while keeping track of changes
- Generic metadata tagging system
- Data adapters for periodic harvesting of external data sources

2. Research Data Sharing: Comprehensive Knowledge **Archive Network**

- Open-source software for open data catalogues
- Metadata storage, discovery, publishing, indexing and data preview
- Broadly used in many different field of application
- Our extensions include eodata publishing interfaces, lineage metadata and its visualization.

3. Data Services and Applications for Stakeholders

- Publish (qualified) data sources, provides filtering mechanisms, exploration, alerts
- Information and early warning applications
- (Lightweight) APIs for OGC web services and RESTful service interface
- Authorization layer





Figure 2: Workfloweditor of the Data Management Platform

demo_org This is a demo organization read more Social Twitter Facebook License Creative Commons Attribution OPER Date	^C demo_org DATASET USGS SRTM30 ^C demo_org DATASET ^C demo_org DATASET ^C demo_org	MODEL PROMET PROC demo_org	SESS	DATASET PROMET demo_org	
About CKAN CKAN API CKAN Association OFEN DATA Debug 11			Powered by Ckan Language: English	T	

Figure 3: Provenance visualization within the CKAN catalogue



https://colabis.de



Mitglied im Netzwerk von