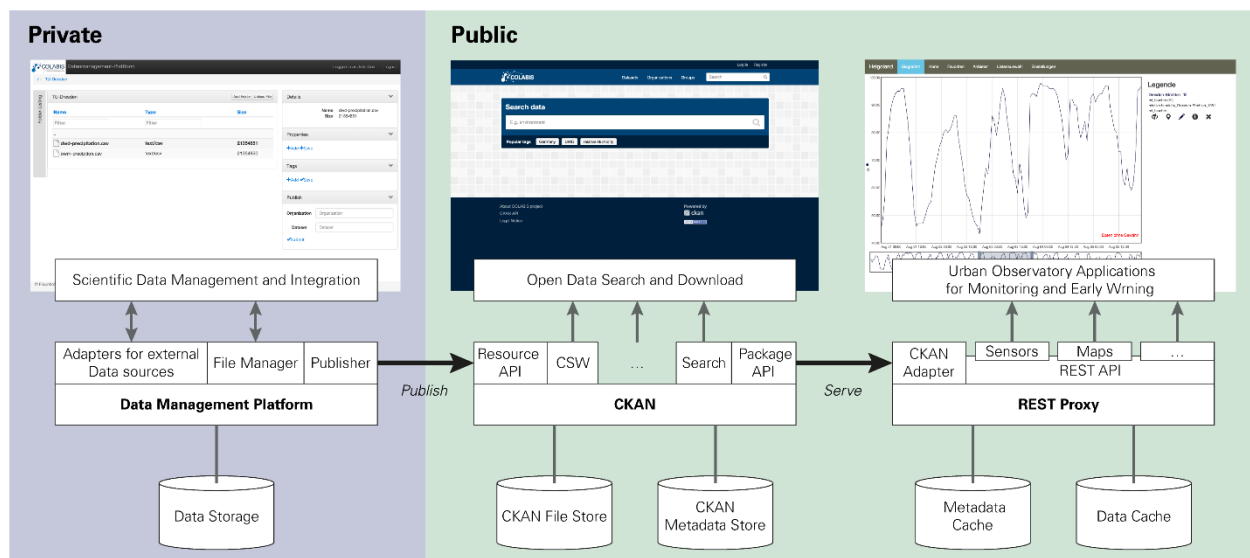


## Getting closer to the researcher's desk - from acquisition and management to publishing scientific data

Stephan Mäs, Daniel Henzen, Lars Bernard  
Professur für Geoinformatik, TU Dresden

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.

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.



The presented architecture and its pilot implementations showcase a seamless, simple to use and secure solution, trying to get as close as possible to data creation and easing and minimizing manual metadata acquisition wherever possible.

The architecture is organised in three main pillars. 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.

As a first pillar, the Data Management Platform supports the daily work of the researchers prior to the publication of the data. It plays a significant role as a data collection and management

system for all data being in use in a research or planning process by an individual or within a small working group. The access to the data can be restricted to team members. The Data Management Platform allows to search and integrate previous data from distributed sources and simplifies the inclusion of the external data sources within projects. It can consume various kinds of (semi-)structured and (non-)semantic data and the integrated metadata enrichment processes eases the follow-up publishing process. All metadata are collected once and continuously updated during the data life cycle. The platform captures also information about the specific research workflows and the steps in which data are collected or changed. This information is used to automatically enrich the metadata with provenance or versioning information.

The second pillar supports the publishing of sufficiently qualified data in the CKAN component (<http://ckan.org>), a software solution for open data management and dissemination. Within COLABIS the CKAN has been extended with data offerings as interoperable, OGC-compliant services to provide spatial data and map visualisations.

The third pillar provides light weight interfaces to allow application development of different information applications. Depending on the application, these can be for example graph visualizations of time series data or maps for decision support and early warning. All software developments of COLABIS are published as Open Source or integrated into the respective open source frameworks.