An integrated software architecture covering the complete data live cycle:

From acquisition and management to publishing INSPIRE services

Stephan Mäs, Daniel Henzen, Lars Bernard
Setting: Case Study COLABIS

- Improve understanding & modelling of pollution processes in urban wastewater systems (e.g. from traffic)
- Provide stakeholders with (early) information to support mitigation (e.g. to schedule street-cleanings or control water discharges)
Setting: Case Study COLABIS

- Data exchange across different research disciplines, workflows, institutions,…
- How to tackle the related research data management?
Current Research Data Management...

- (Generic) Research data exchange publishing platforms do exist or emerge (EUDAT, PANGEA, Data Journals, ...)

http://eudat.eu/

http://doi.pangaea.de/10.1594/PANGAEA.51464
Current Research Data Management…

- Lack of support of the whole research workflow (data publication is mostly done once at the end…)
- Still *too* many tedious metadata tasks
- Lack of data management plans
- Lack of research data qualification mechanisms
- Not *the* solution in place….
Our current approach → https://colabis.de/

Researchers’ data management

Scientific Data Management and Integration

Adapters for external Data sources
File Manager
Publisher

Data Management Platform

Data Storage

Research data sharing

Open Data Search and Download

Resource API
CSW
Search
Package API

CKAN

CKAN File Store
CKAN Metadata Store

Provide Applications to various stakeholders

Urban Observatory Applications for Monitoring and Early Warning

RestProxy

CKAN Adapter
Sensors
Maps

Metadata Cache
Data Cache

Serve

Publish
Researchers’ Environment  
**Data Management Platform**  

- Central part & initial component for all subsequent tasks  
- Scalable, multi-purpose file manager to collect, manage & share data in *one common environment*  
- Supports to enrich & update data, while keeping track of changes  
- (Generic) tagging system (→ *metadata*)  
- Includes data adapters for periodic harvesting of external data sources
Research data sharing:
**Comprehensive Knowledge Archive Network (CKAN)**

- Open Source Software for open data catalogues
- Metadata storage, discovery, publishing, indexing and data preview
- **Broadly used – also non spatial**
- Handles geodata, lineage, tagging
- Extensions for GeoServer publishing, user/role management,…
- Powerful API
Data Services & Applications for Stakeholders

- Publish (qualified) data sources, provides filtering mechanisms, exploration, alerts…
- Information and early warning applications
- Directly links to CKAN
- (Lightweight) APIs
  - OGC web services (GeoServer)
  - RESTful service interface…
- Authorization layer
Advantages for the Researchers

- Closer to the researcher’s daily work: Ideally an *easy-to-use, common* data management platforms is part of the researchers desk
- Support of all research related tasks and the whole research data management & publication process
- Less (but still too many) tedious metadata tasks
- Support semantic enrichment, transformation, qualification & publishing
- Ease of research data sharing/publication
- Quick to provide such a platform

Try out and read more: [https://colabis.de/](https://colabis.de/)
Parallels to the INSPIRE Publication Process

- technical components for data acquisition, integration and management are usually not very well connected to data publication process required for INSPIRE
- a seamless data management environment is often missing
- An integrated metadata enrichment processes eases the follow-up publishing process
- Versioning
- Track of changes and data provenance
- simplifies the inclusion of the external data sources
Thank you!

This work has received funding from the German Federal Ministry of Education and Research, programme Geotechnologien, under grant agreement no. 03G0852A.