Management and publishing of TERENO data from distributed data bases

Tereno Coordination Team
Data Management

TERENO Advisory Board Meeting

26. September 2011, Blankenburg

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

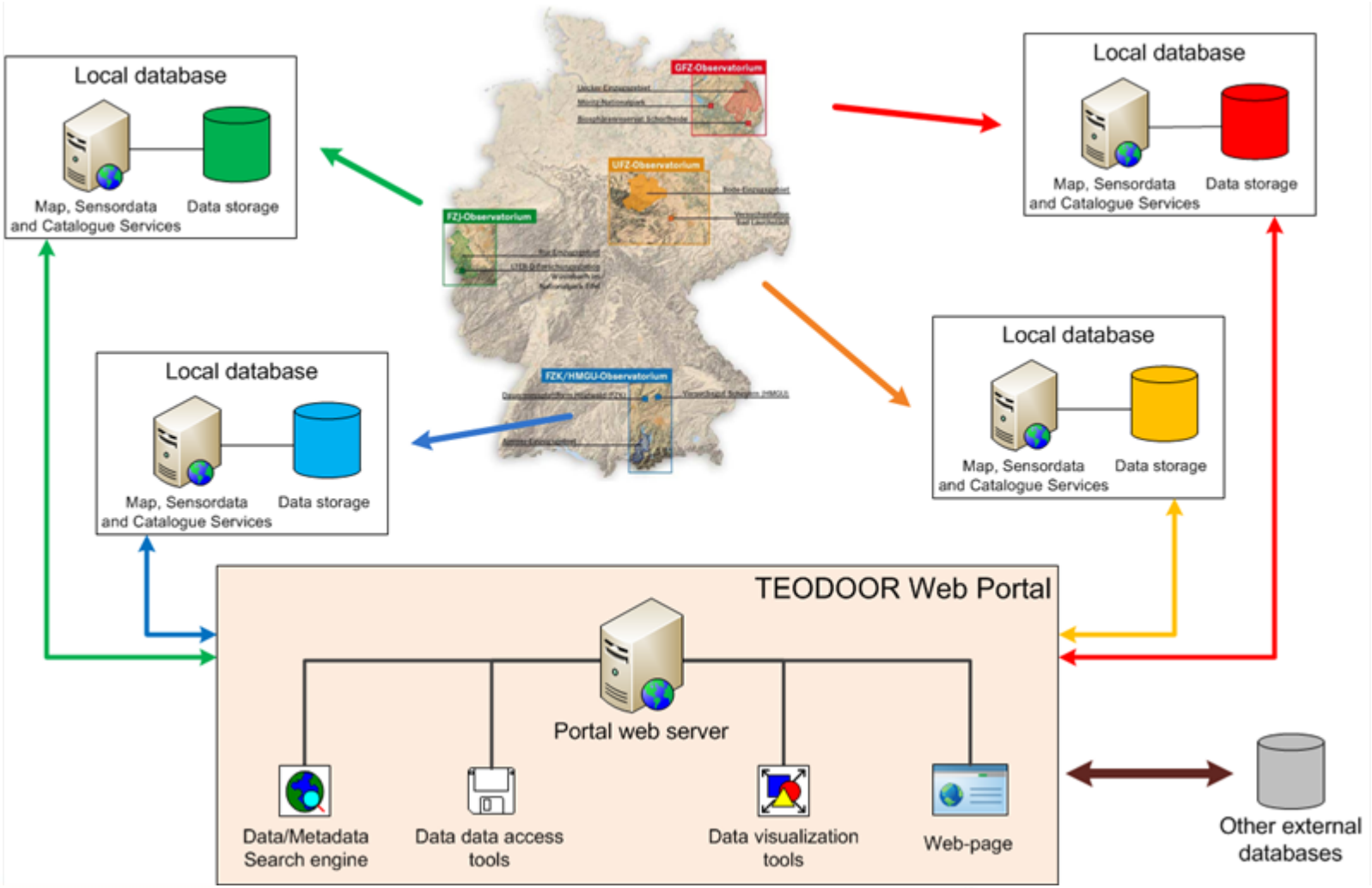
Current status at last advisory board meeting

- Status 2009:
  - Creation of a data management plan; query of data management issues in the different observatories
  - Begin of infrastructure implementation
    - Local databases
    - TERENO data portal
- Work progress in 2010
  - Data policy
  - Continuation of infrastructure implementation
    - Implementation of interfaces (Web-Services)
    - Importing and publishing of existing data
    - Publishing data from local databases (weather radar)
    - Coupling of TEODOOR and local databases, e.g. metadata query

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

TERENO distributed data infrastructure design

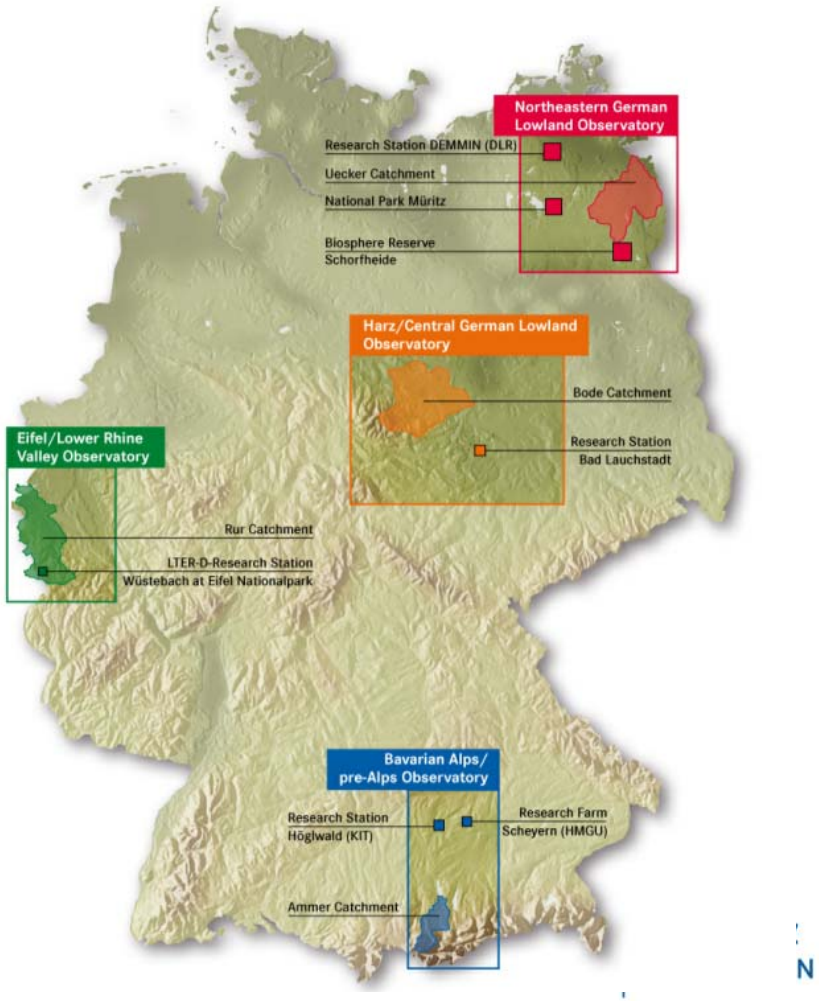


TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Local Tereno data being processed

Example: FZJ

- Documents and other file based data: 10 GB a^{-1}
- GIS- and remote sensing data : 100 GB a^{-1} (500 GB in total)
- Data from automatic stations:
  - SoilNet: 4750 st.*par. (15‘$^{-1}$)
  - SoilCan: 2205 st.*par. (10‘$^{-1}$)
  - Other: 120 st.*par. (10‘$^{-1}$)
- Eddy-Covariance data:
  - 100 st.*par. (20“$^{-1}$)
  - 100 st.*par. (30‘$^{-1}$)
- Weather radar: 600 MB h^{-1} (two data sets each 5‘)



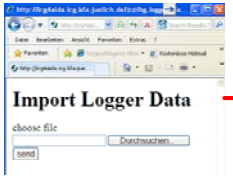
TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Automated import and processing tool for site data

Automatic sampling stations

E-Mail attachments

E-Mail parser



Manual upload

upload folder

Data import processing tool

Data overdue check

Technical warner

E-Mail notification

Data range check

Content warner

E-Mail notification

Data calculator

Raw data archive

Water samples

Local database

Data converter

PostGIS database

Raw data (Level 0)

Processed data (Level 1)

Quality checked data (Level 2)

Set quality checked data (Web-processing service)

Sensor Observation Service (SOS)

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

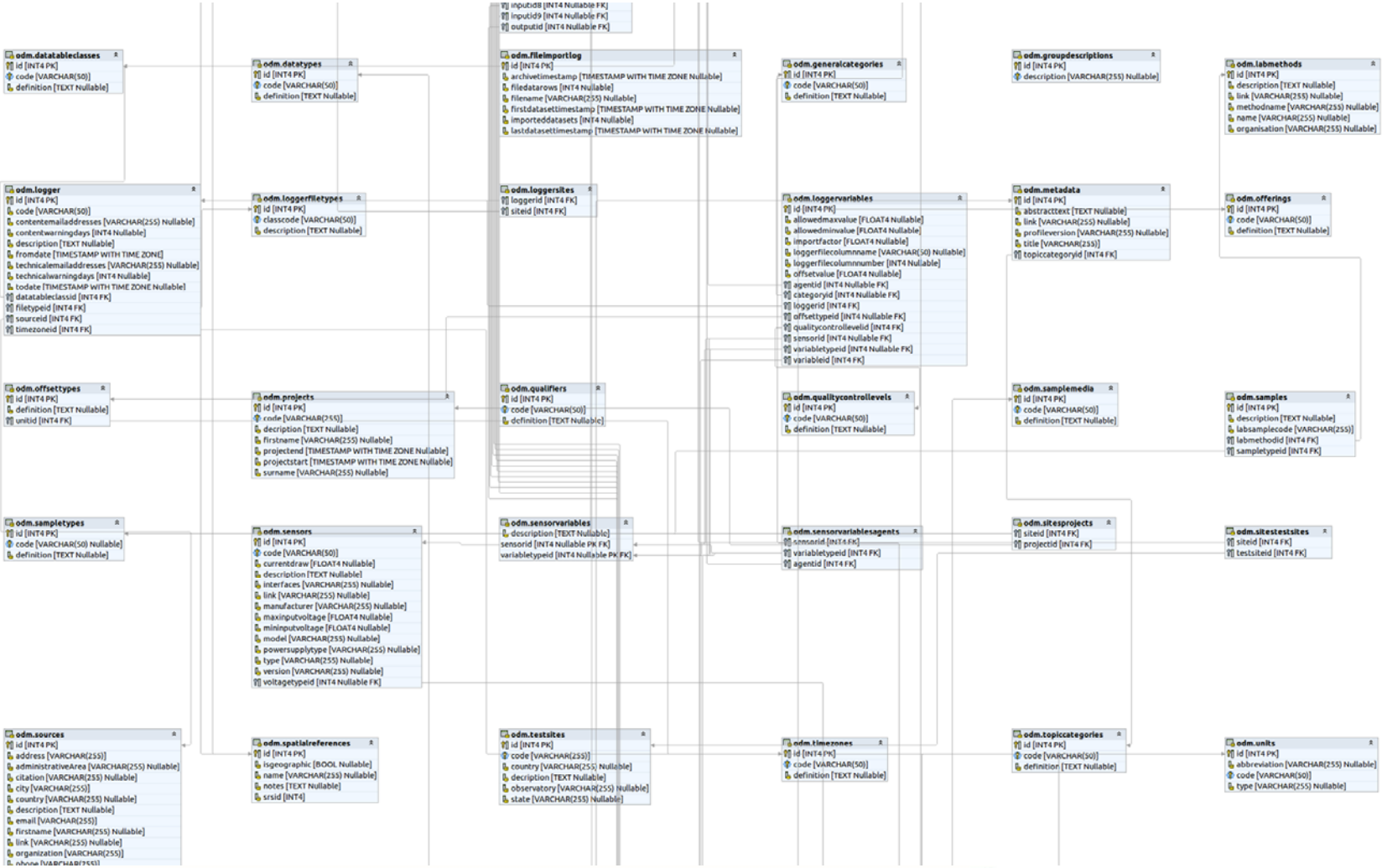
Data model and data publishing

- Comprehensive data model based on the CUAHSI Observation Data Model (http://his.cuahsi.org/odmdatabases.html):
  - Sites
  - Sources and metadata
  - Sensors
  - Data classification, categories, data level, attributes
  - Data generation, lab methods, sample handling
- Extension of the model to
  - specify individual sensors and and data import by logger files
  - store all relevant information in one relational data base that can be assessed by the users
- Implementation in JAVA using Hibernate3 for
  - Data base independent processing
  - Automated table generation and management
  - Versioning

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Data model used at FZJ-IBG



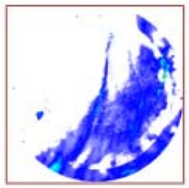
TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Managing and publishing weather radar data

Local data preprocessing

Data processing and reflectivity calculation

Raw data storage



Creation of precipitation images

Creation of metadata from templates

Teodoor (Tereno Data Portal)

Precipitation raster data download

Animated precipitation raster player/viewer

Metadata search engine

Metadata viewer and editor

Web-Map Service (geoserver)

Catalog Service (geonetwork)

Internet

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

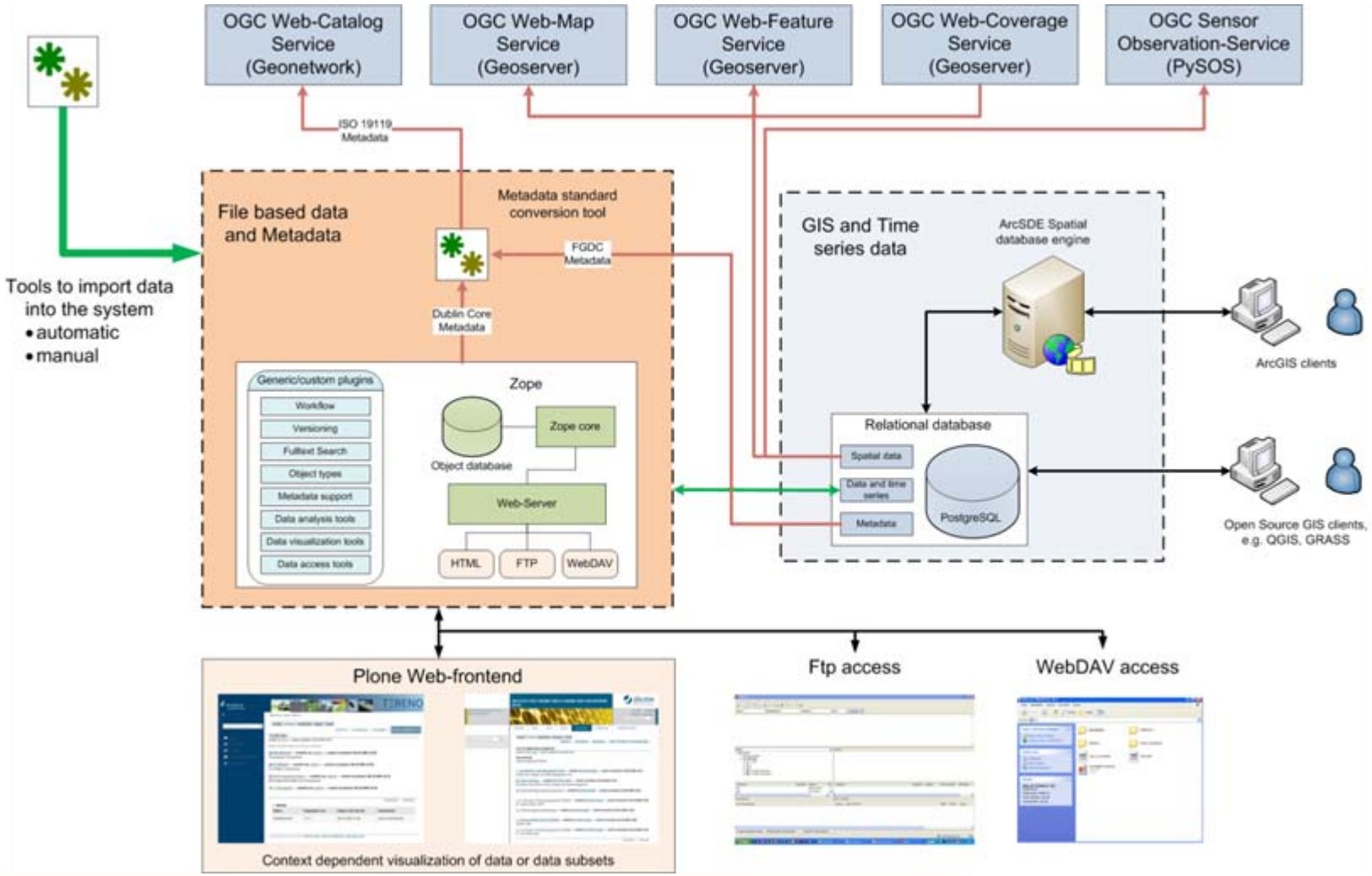
Sensor Observations Service (SOS)

- Most important web-service to provide access to time series observations from sensors in a standardized way
- Widely used for point data
- Although mentioned in OGC-SOS specification, no existing SOS implementation is able to deliver raster data time series (only point data)
- SOS extension implementation (Master Thesis J. Sorg):
  - Data storage in PostgreSQL data base
  - Time series output of rasters or subrasters (spatial filters) as
    - WMS or WCS layer references
    - O&M discrete coverages (geometries and attributes)
  - Output of time series of individual locations within as raster
  - Zonal detection of special events (e.g. rain storms)

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Backend: FZJ/IBG-3 data infrastructure (AIDA, in operation since 2009)



TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

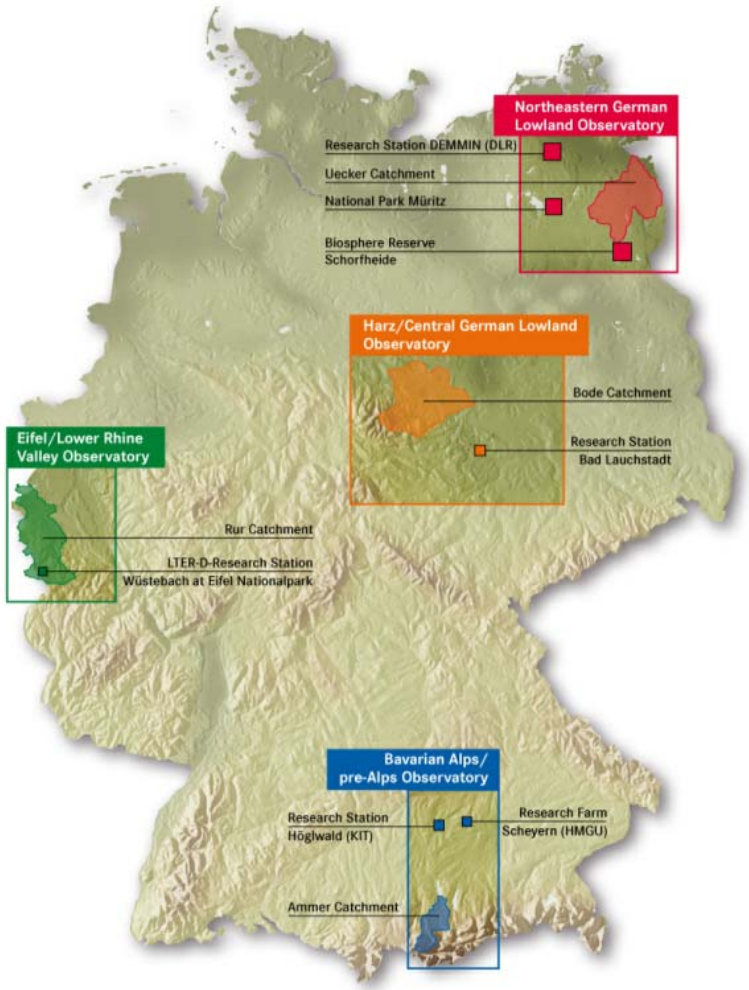
Summary: Data publishing infrastructure (FZJ)

- Catalog service running, supporting both ISO 19115 and DublinCore, coupled to the AIDA content management system
- At present, data from more than 200 stations are published online
  - 6 SOS-Services for public data (quality checked)
  - 5 SOS-Services for internal data (raw data)
  - Testsites: Rollesbroich, Wuestebach, Selhausen, SoilNet, Rur-catchment
  - Parameter classes: Climate, Discharge, Soil, RawData (only internal)
- New developed Raster-SOS running containing more than 20000 individual raster layers
- 2 WMS-Services, ArcGIS Server, PostGIS database ...

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Current status of local database implementation



- Installation and instrumentation at TERENO stations almost completed for all observatories (except GFZ, which just started installation and instrumentation)
- Data link from the stations to the local data bases established
- Automated data transfer from the stations to the local data bases established
- Automated data processing and visualization via individual Web-pages partly realized
- Interfaces between local data bases and TEODOOR data portal via SOS in progress

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

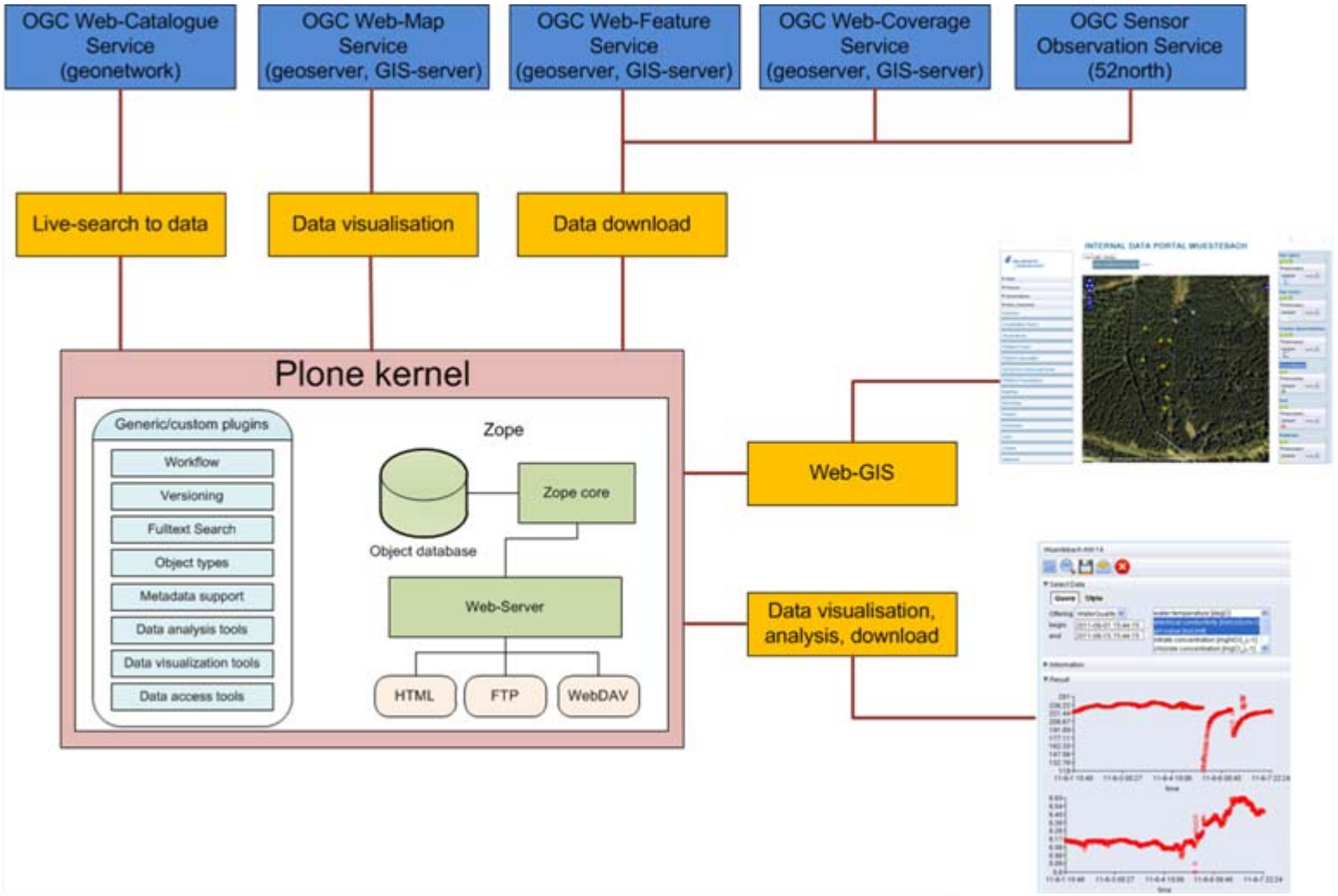
TEODOOR: The TERENO Data Portal

http.//www.tereno.net

- Implemented in Plone
- Contains practically no own data
- Communicates to local databases via OGC-compliant Web-services
- Internal and external live search to data
- Included Web-GIS functions

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Portal backend: Interfaces and functions



TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Tereno Metadata profile

- Draft document distributed to the CT DM members
- Compliant to common standards (e.g. ISO 19115/19139, INSPIRE, GDI-DE)
- Contains
  - Data description
  - Data location
  - Contact information

Theme	Element	Multi-plicity	ISO 19115 core	
	Resource title	1	Dataset title	M
	Resource abstract	1	Abstract describing the dataset	M
Identification	Resource type	1		
	Resource locator	0..*	Online resource	O
	Unique resource locator	1..*		
	Resource language	0..*	Dataset language	M
Classification	Topic category	1..*	Dataset topic category	M
Keywords	Keyword value	1..*		
	Originating controlled vocabulary	0..1		
Geographic location	Geographic bounding box	1..*	Geographic location of the dataset	C
Temporal reference	Temporal extent	0..*	Additional extent information for the dataset	O
	Date of publication	1..*	Dataset reference date	M
	Date of last revision			
	Date of creation			
Resolution and validity	Lineage	1	Lineage	O
	Spatial resolution	0..*	Spatial resolution of the dataset	O
Access constraints	Conditions applying to access and use	1..*		
	Limitations on public access	1..*		
Responsible party	Responsible party	1..*	Dataset responsible party	O
	Responsible party role			
Metadata metadata	Metadata point of contact	1..*	Metadata point of contact	M
	Metadata date	1	Metadata date stamp	M
	Metadata language	1	Metadata language	C

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Hierarchical Search

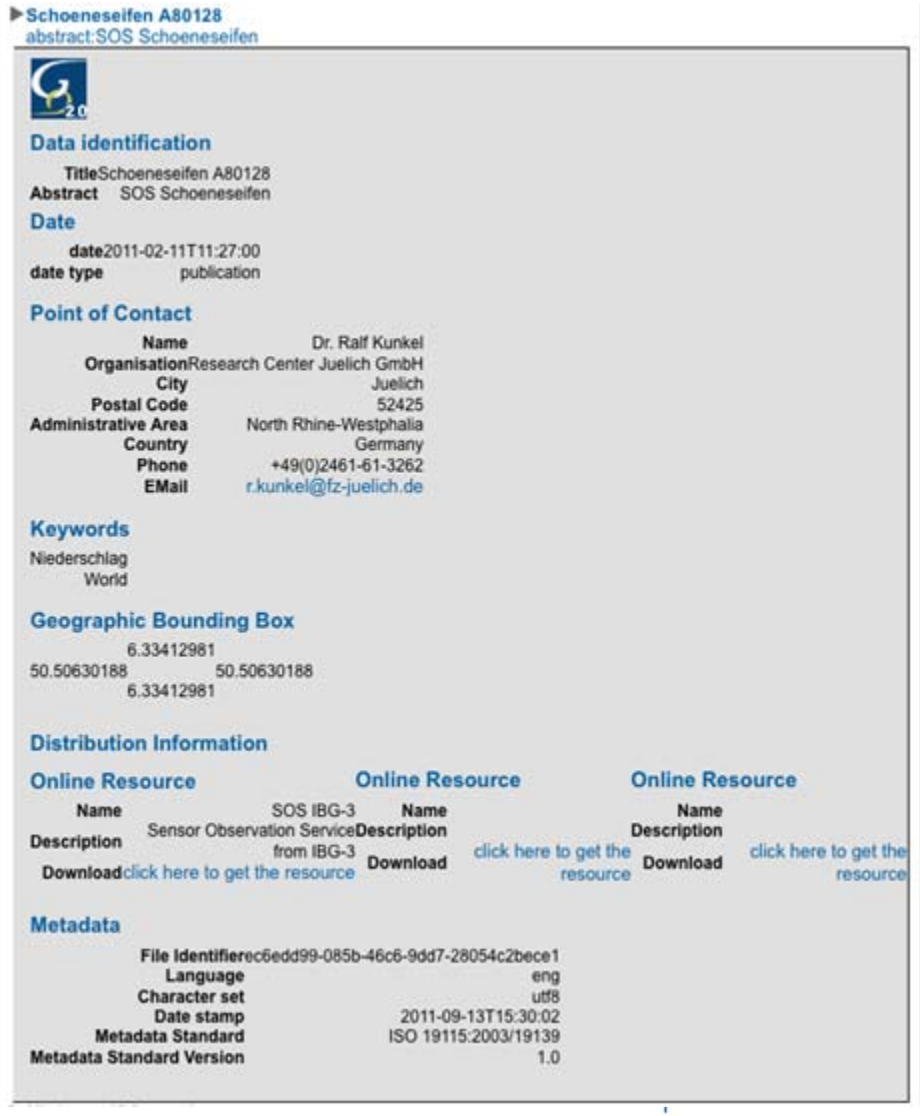
- Search and find data in TEODOOR and in distributed OGC-Catalogue services
- Supports hierarchical search in Metadata:
  - Germany > Rur
  - Hydrosphere> SurfaceWater > ConductivityElectrical
- Thesaurus:
  - Common vocabulary, keyword description
  - Defined hierarchy
  - Basis: open source EU-GEMET thesaurus (http://www.eionet.europa.eu/gemet)
  - Additional Tereno Thesaurus

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Hierarchical Search

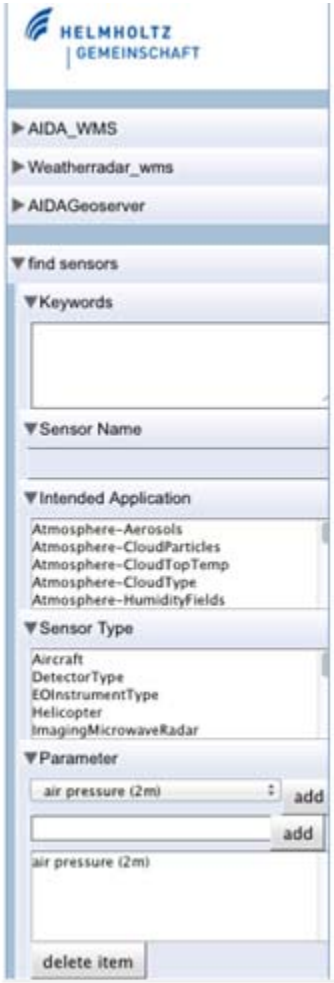
- Search and find data in TEODOOR and in distributed OGC-Catalogue services
- Supports hierarchical search in Metadata:
  - Germany > Rur
  - Hydrosphere> SurfaceWater > ConductivityElectrical
- Display of detailed metadata and storage location information



TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Spatial search

- Search and find data in TEODOOR by Web-GIS:
  - Keywords
  - Sensor names
  - Sensor types
  - Intended applications
  - parameters
- Display all stations fulfilling search criteria
- Display station information
- Data visualization



TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Web-Gis functions in TEODOOR

- Implemented using OpenLayers
- Supports multiple WMS and SOS
- Customized
  - Default content
  - Default region
  - Visible WMS
  - Visible SOS
- Plone workflow support for adjusted data views and access

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Data visualisation in TEODOOR

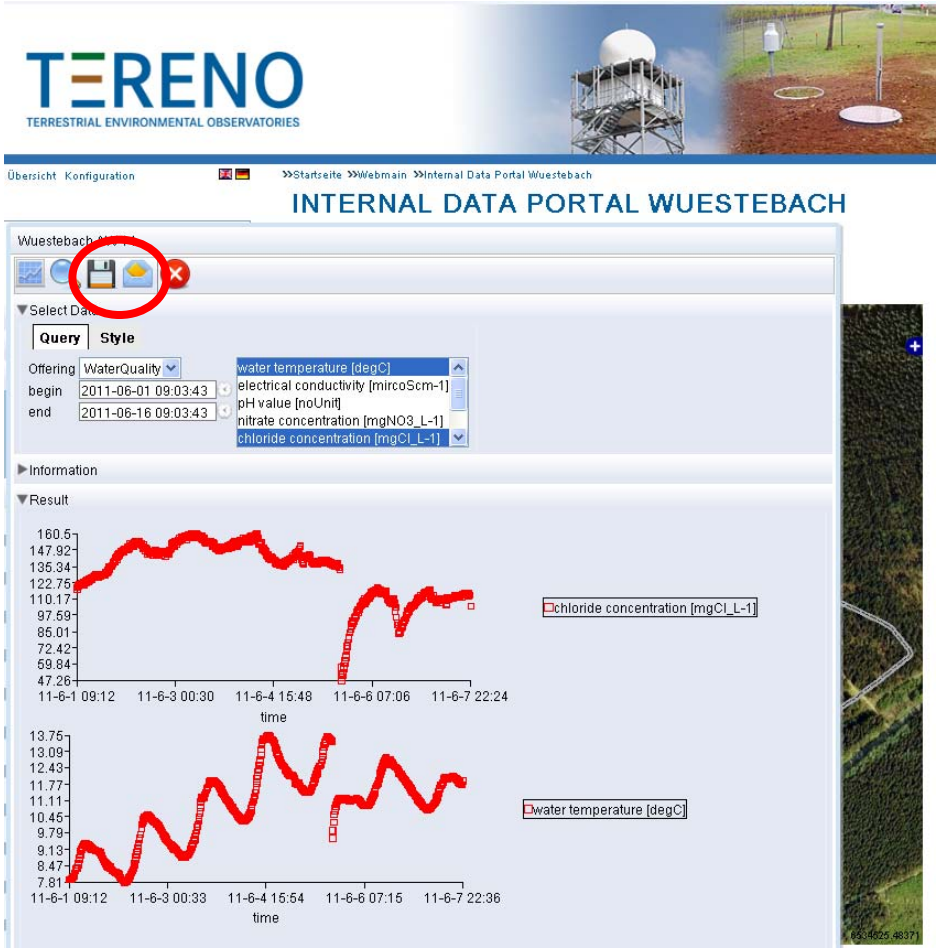
- Connecting to OGC-SOS services
- Graphical selection of stations
- Display of:
  - Station information (sensorML metadata)
  - Latest observations
  - Offerings
  - Available parameters

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Data visualisation in TEODOOR

- Connecting to OGC-SOS services
- Graphical selection of stations
- Display of:
  - Station information (sensorML metadata)
  - Latest observations
  - Offerings
  - Available parameters
- Visualisation of station data time series
- Data download (E-Mail)

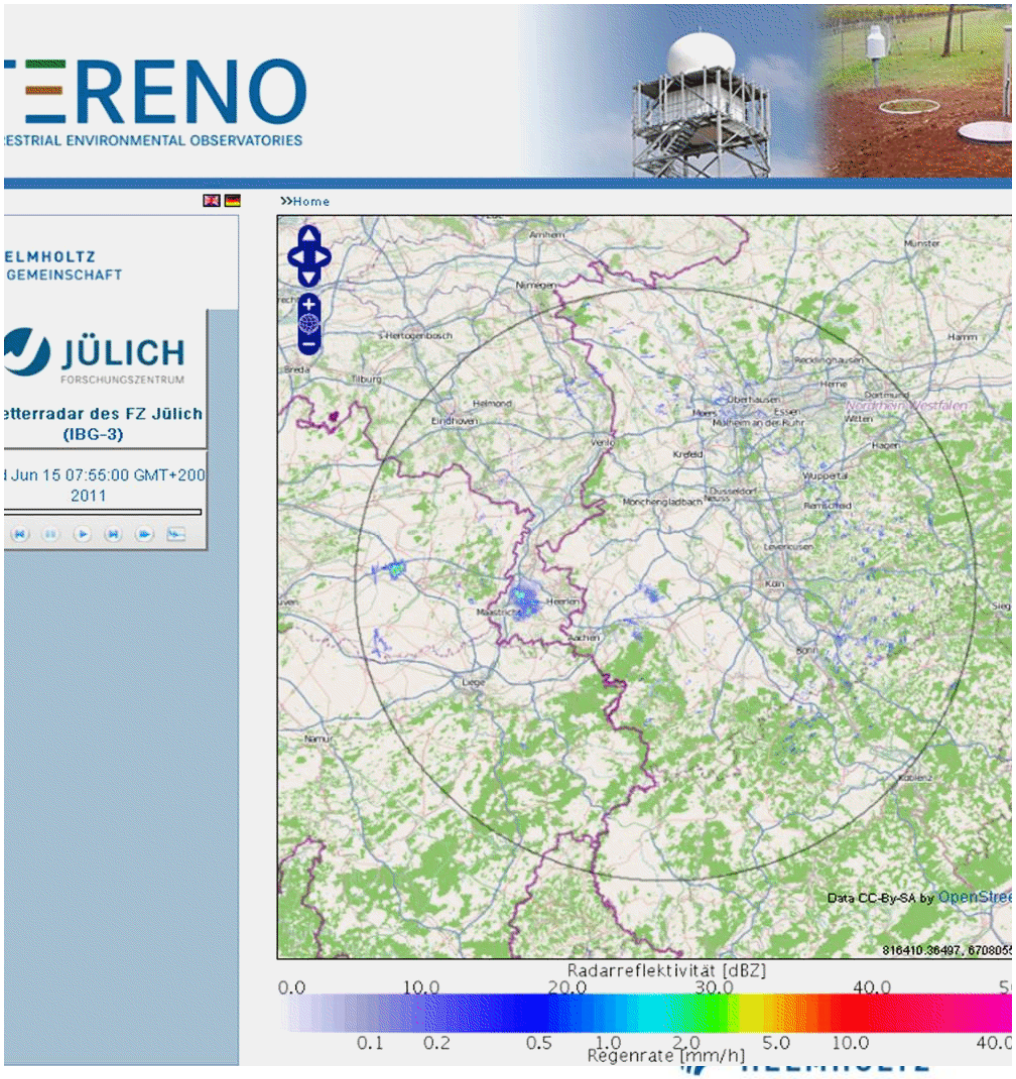


HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Weather radar data visualization

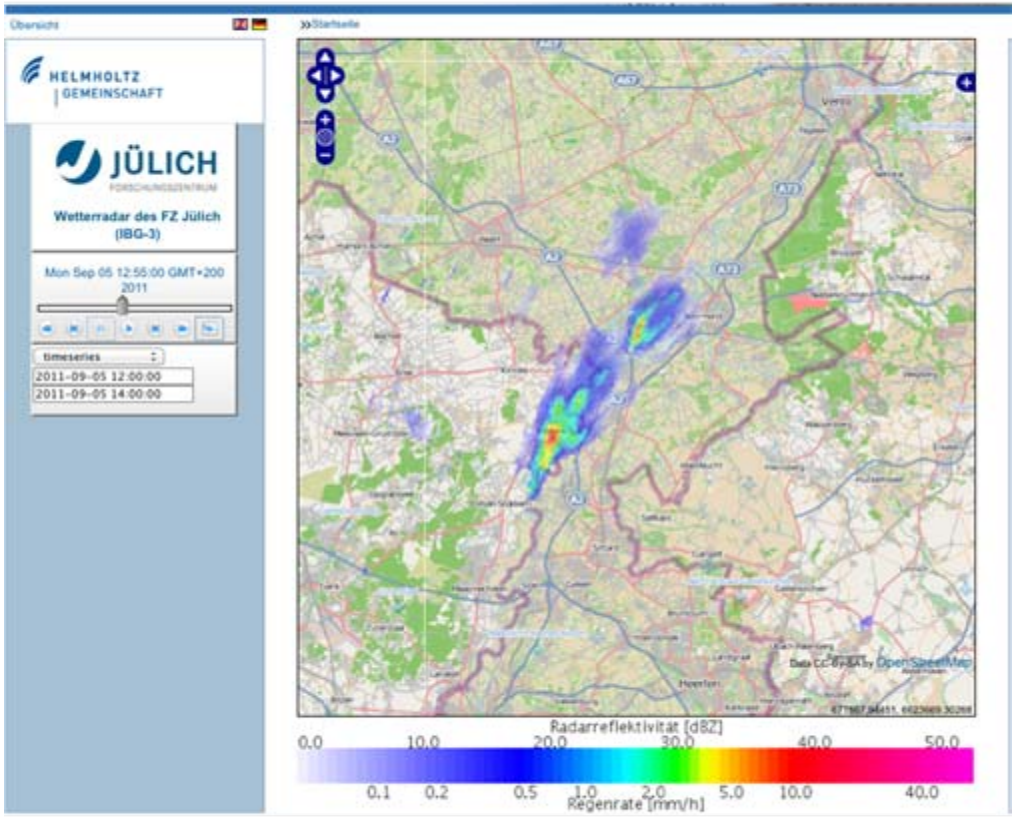
- Data visualization using distributed OGC-Raster SOS and WMS
- Raster data animation for custom
  - time periods



TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Weather radar data visualization

- Data visualization using distributed OGC-Raster SOS and WMS
- Raster data animation for custom
  - time periods
  - regions of interest
- Reflectivity/precipitation display for a given raster point

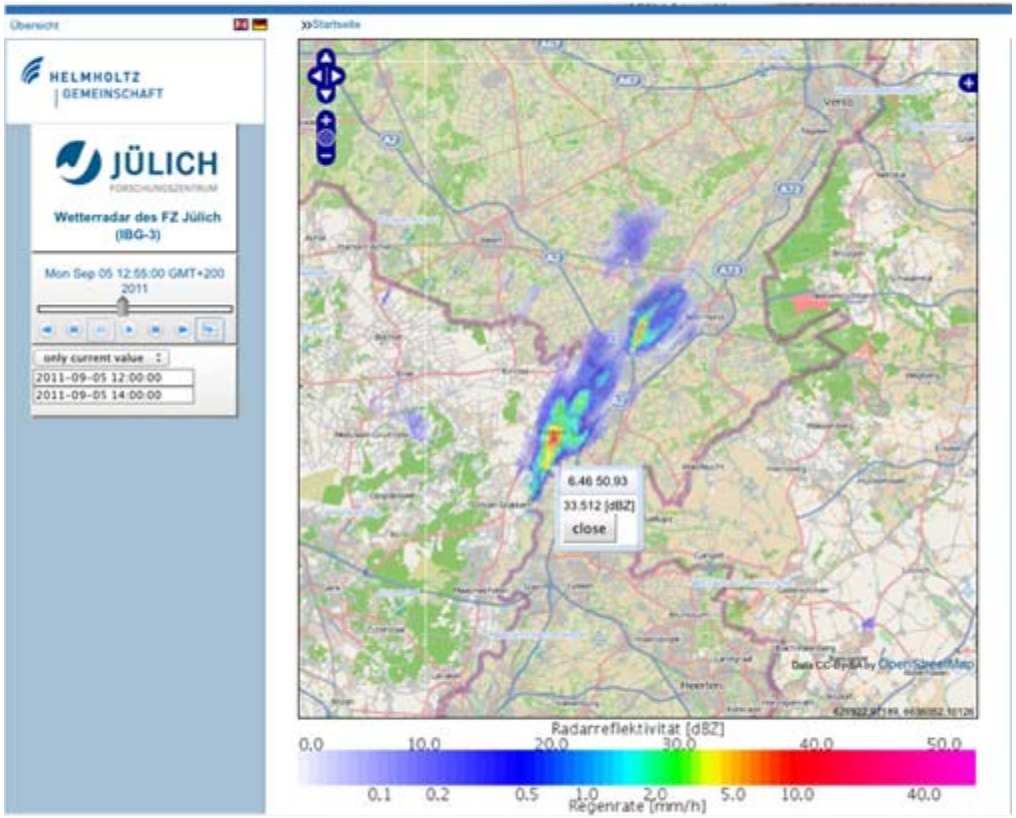


HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Weather radar data visualization

- Data visualization using distributed OGC-Raster SOS and WMS
- Raster data animation for custom
  - time periods
  - regions of interest
- Reflectivity/precipitation display for a given raster point

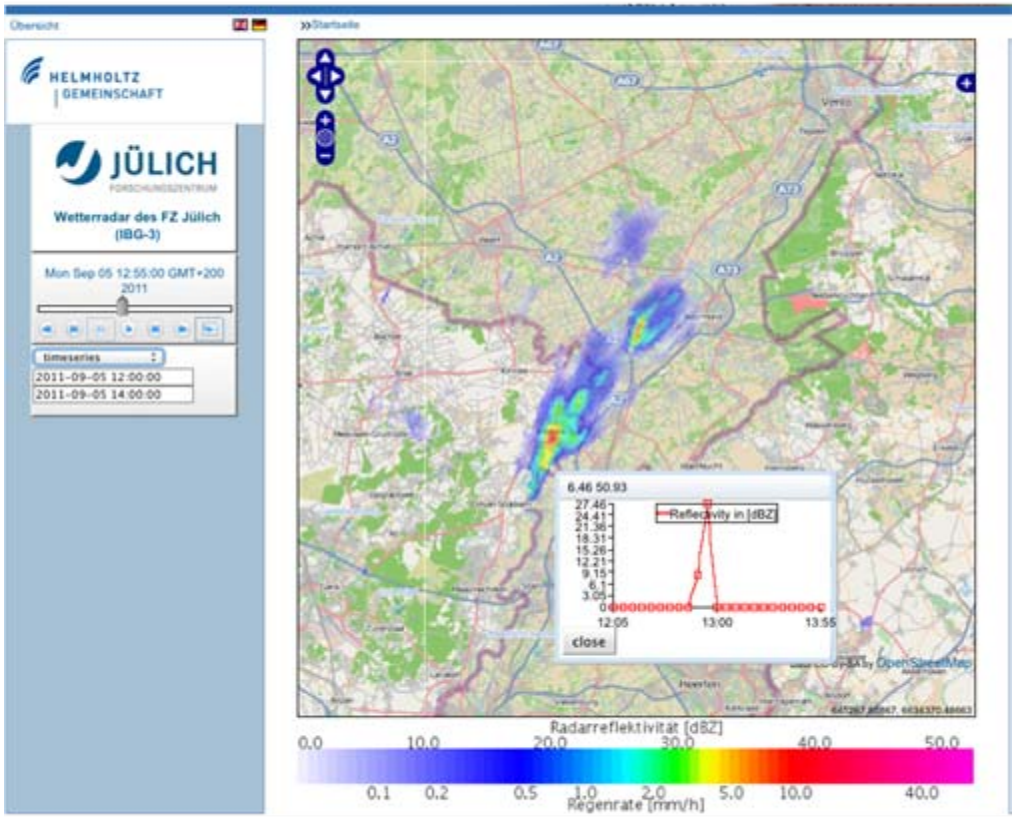


HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Weather radar data visualization

- Data visualization using distributed OGC-Raster SOS and WMS
- Raster data animation for custom
  - time periods
  - regions of interest
- Reflectivity/precipitation display for a given raster point
- Reflectivity/precipitation time series graphs for a given raster point



HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Conclusions and outlook

- Current status:
  - Local databases in place for all observatories (except GFZ)
  - Internal data import, storage, processing and vizualization mostly running
  - Interfaces for data exchange partially in work, partially in progress
  - Catalogue services partially online, currently adapted to Tereno Metadata profile
  - TEODOOR data portal online, coupling to local databases working
- Outlook:
  - Publish primary data using persistent Digital Object Identifiers (DOI)
  - Improvement of quality control of the primary data and the descriptive metadata (see also UFZ poster)
  - Ensuring long-term availability of the published data in online repositories
  - Include data sets with ecological content and spatial data (e.g. from remote sensing)

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Roadmap for Tereno Data management infrastructure implementation

- For each observatory
  - List of stations to be published — end of 10/11
  - Catalogue Service set up — FZJ+UFZ done, others until 01/12
  - Sensor Observation Services set up — partially done, rest starting 01/12, finished for all implemented stations until 07/12 (GFZ later)
  - Other Web-services set up — depending on demand
- Standards definition and implementation
  - Metadata — INSPIRE standard
  - Thesaurii — GEMET + own, in development, Version 1.0 until 01/12
- TEODOOR
  - INSPIRE support — End of October 2011
  - additional services and tools — under discussion, if required
- Others
  - DOI referenced datasets — pilot phase starting 1/12, operational until 12/12

HELMHOLTZ ASSOCIATION