Management and publishing of TERENO data from distributed data bases

Ralf Kunkel, Jürgen Sorg
and the Tereno Coordination Team
Data Management

Publication of Tereno data using Digital Object Identifiers (DOI)

Jens Klump

TERENO Workshop

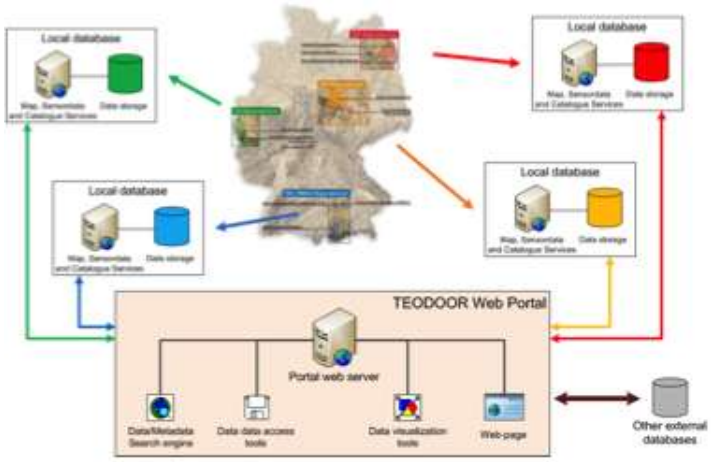
January, 24.-25. 2012, Potsdam

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

TERENO distributed data infrastructure

- Individual data infrastructures for each observatory
- Metadata services providing information on monitoring stations and measured data
- Central data portal (TEODOOR) for information exchange, data search and data access
- Communication between local databases and TEODOOR via OGC-compliant Web-services



TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Status of Tereno station implementation

	Established in Eifel/LRV Observatory	Established in Harz/CGL Observatory	Established in Alps/pre-Alps Observatory	Planned in NE German Lowland Observatory
Meteorological stations Incoming short wave radiation Precipitation Air humidity Air temperature Windspeed/ -direction	**- Schöneseiffen** **- Wüstebach** **- Rollesbroich** **- Selhausen** **- Merzenhausen** **- Tietz** Planned 2011-12: 5 further stations	- **Kreinitz** (no radiation) - **Gimritz** (no radiation) - **Zöberitz** - **Greifenhagen** (no rad) Planned 2011-12: 7 further stations	**- Graswang** **- Rottenbuch** **- Fendt** **- Schechenfilz** **- Garmisch-Partenkirchen** **- Höglwald** **- Bavarian Forest** EC-Stations[1] (as above, at Garmisch in construction)	**- DEMMIN** Planned 2011-12: 20 climate stations **- Müritz Nationalpark Fürstenseer-See** 4 climate stations plus throughfall and stemflow
Hydrological stations Streamflow discharge Water temperature Electrical conductivity pH Redox potential Chlorophyll a Dissolved organic matter	**- Wüstebach** (3 stations) **- Erkensruhr** **- Rollesbroich** Planned 2012: 1 further station (BMBF project "Huminstoffe")	**- Meisdorf** **- Silberhütte** **- Hausneindorf** **- Hadmersleben** **- Sauerbach** **- Athensleben** **- Staßfurt** **- Rappbode Observatory**	Discharge data will be available from local authorities	**- Müritz NLP Fürstenseer-See** **Planned (2011-12)** Water levels in lake and groundwater, temperatures, EC **- Uecker Catchment** Planned
Soil monitoring stations Soil water content Soil temperature Soil suction Soil organic matter	**- Schöneseiffen** **- Wüstebach** **- Rollesbroich** **- Selhausen** **- Merzenhausen** **- Tietz** Planned 2011-12: 5 further stations	- **Kreinitz** (no SOM) - **Gimritz** (no SOM) - **Zöberitz** (no SOM) - **Greifenhagen** (no SOM) Planned 2011-12: 7 further stations	**- Höglwald** (no soil suction) **- Graswang** **- Rottenbuch** **- Fendt**	**- DEMMIN** Planned 2011-12: SoilNet system (design to be defined) **- Müritz NP Fürstenseer-See** Planned (2011-12)

HOLTZ
OCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

FZJ data input from monitoring stations

	Established in Eifel/LRV Observatory
Meteorological stations Incoming short wave radiation Precipitation Air humidity Air temperature Windspeed/ -direction	**- Schöneseiffen** **- Wüstebach** **- Rollesbroich** **- Selhausen** **- Merzenhausen** **- Tietz** Planned 2011-12: 5 further stations
Hydrological stations Streamflow discharge Water temperature Electrical conductivity pH Redox potential Chlorophyll a Dissolved organic matter	**- Wüstebach** (3 stations) **- Erkensruhr** **- Rollesbroich** Planned 2012: 1 further station (BMBF project "Huminstoffe")
Soil monitoring stations Soil water content Soil temperature Soil suction Soil organic matter	**- Schöneseiffen** **- Wüstebach** **- Rollesbroich** **- Selhausen** **- Merzenhausen** **- Tietz** Planned 2011-12: 5 further stations

- Runoff, water quality, soil, slow climate
  - 53 stations
    - Automatic stations (10'-60')
    - Offline data (laboratory)
  - 26400 data values per day (14.9 millions in total)
- Soil, slow climate (SoilNet)
  - 275 automatic stations (10')
  - 686000 data values per day (146.5 millions in total)
- Eddy-Covariance
  - 7 stations (20 Hz-10')
  - 133 million data values per day (14.3 billions in total)
- Weather radar
  - 2 radar devices (5-10')
  - 576 rasters, 4.3 GB raw data per day (47000 rasters, 6.8 TB raw data in total)
- Lysimeters (SoilCan)
  - 126 lysimeters (15')
  - About 300000 data values per day

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Components of local data infrastructure
(Example FZJ)

- Comprehensive and flexible observation data model for
  - Sites
  - Sources and metadata
  - Sensors
  - Data classification, categories, data levels, attributes
  - Data generation, lab methods, sample handling
- Automated data import and preprocessing tools
- Archive, backup and versioning strategies
- Sample management
- Quality control tools
- Data handling and manipulation tools
- OGC-conformal Web services for data publishing:
  - 13 Sensor Observation Services (SOS)
  - 1 newly developed Raster SOS
  - Catalog service (WCAT)
  - 2 Web-Map Services (WMS)

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Current status of local database implementation

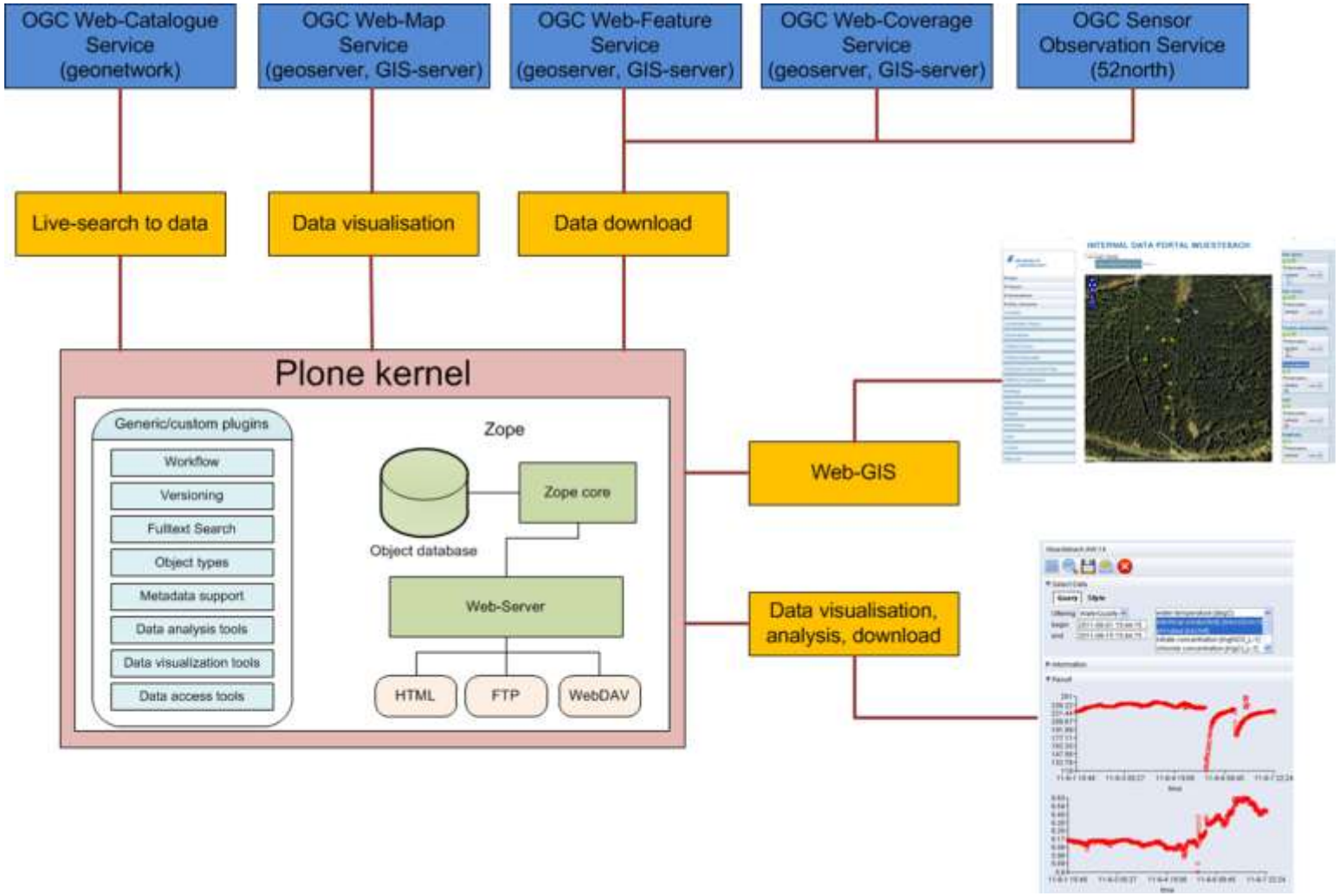
TEODOOR ONLINE DATA PORTAL

- Installation and instrumentation at TERENO stations almost completed for all observatories (except GFZ)
- Data transfer from the stations to the local data bases established
- Automated data processing and visualization via individual Web-pages partly realized
- Interfaces to TEODOOR data portal via SOS established in KIT and FZJ (UFZ very soon)
- Catalogue services in UFZ and FZJ established

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Tereno Portal TEODOOR: Interfaces and functions



TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Hierarchical and spatial data search

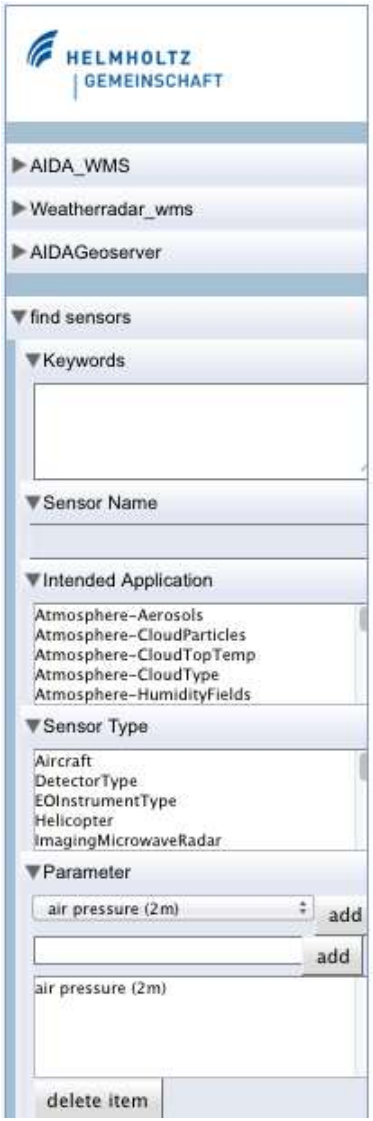
- Search and find data in TEODOOR and in distributed OGC-Catalogue services
- Hierarchical search:
  - Categories e.g.:
    Germany > Rur
    Hydrosphere> SurfaceWater > ConductivityElectrical
  - Display of detailed metadata and storage location information

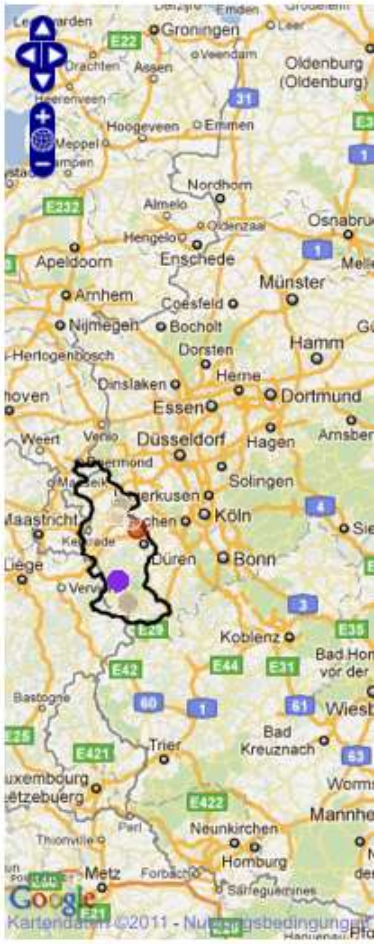
HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Hierarchical and spatial data search

- Search and find data in TEODOOR and in distributed OGC-Catalogue services
- Hierarchical search:
  - Categories e.g.:
    Germany > Rur
    Hydrosphere> SurfaceWater > ConductivityElectrical
  - Display of detailed metadata and storage location information
- Spatial data search: Display of all stations fulfilling search criteria
  - Keywords
  - Sensor names
  - Sensor types
  - Intended applications
  - Parameters





TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Web-Gis functions in TEODOOR

- Implemented using OpenLayers and GWT
- 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

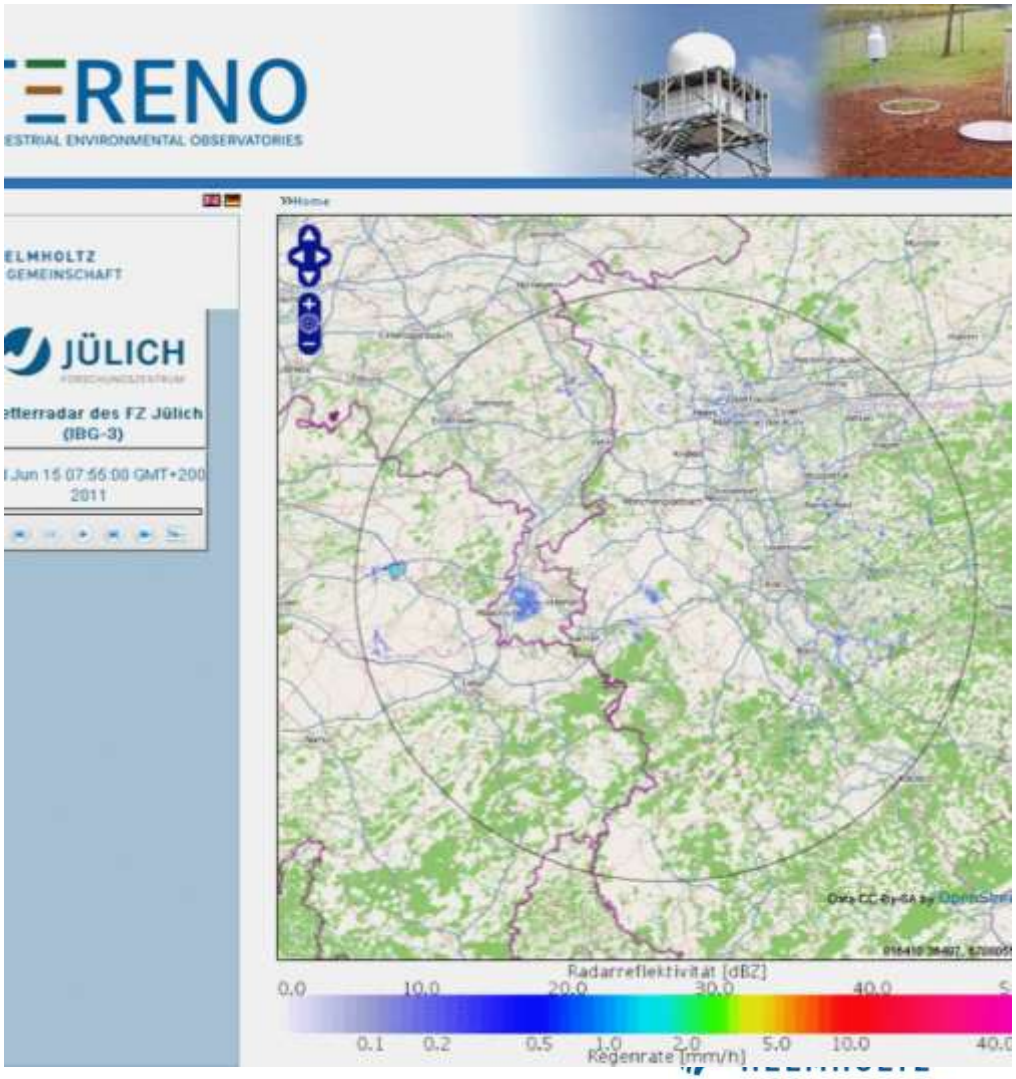
- Connecting to OGC-SOS services
- Graphical selection of stations
- Display of:
  - Station information (sensorML metadata)
  - Latest observations
  - Offerings
  - Available parameters
- Visualisation of station data time series
- Data download (E-Mail)

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

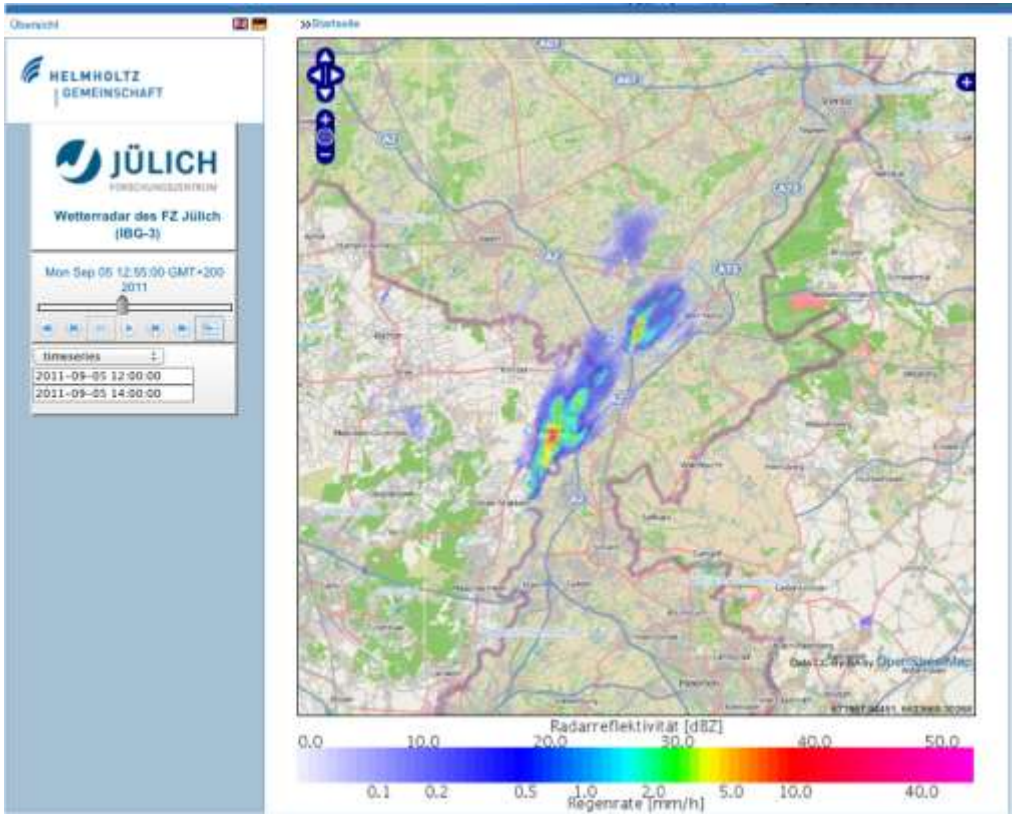


ASSOCIATION

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

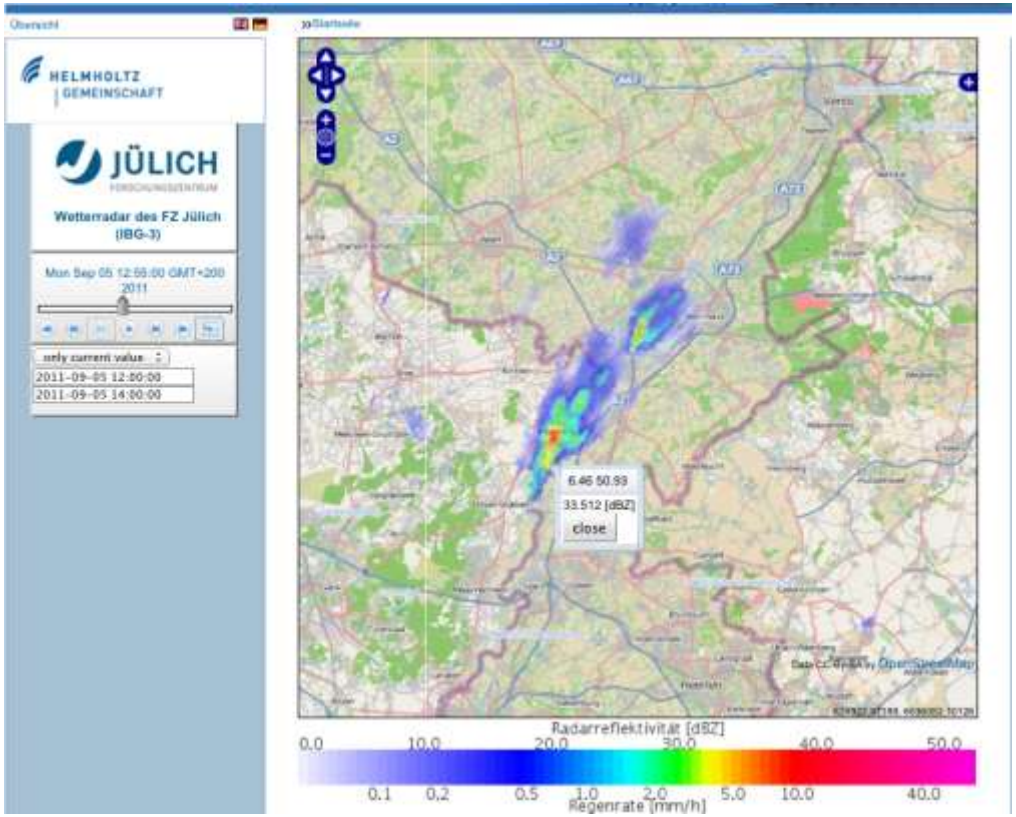


HELMHOLTZ ASSOCIATION

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

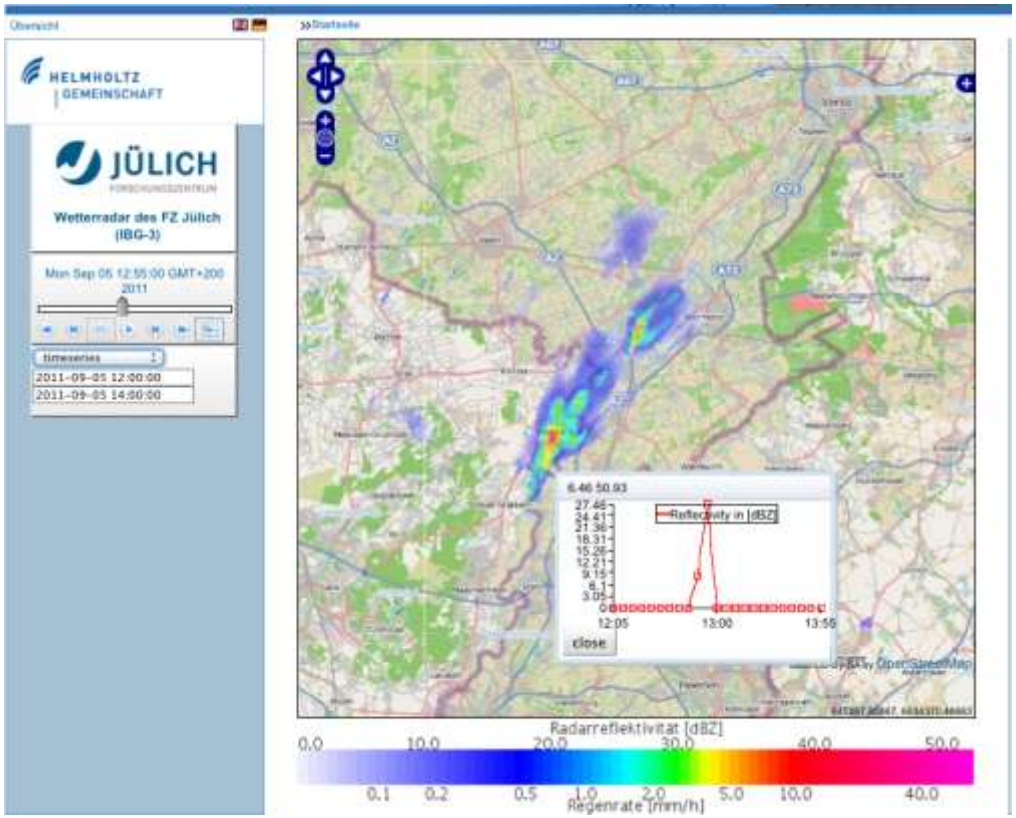


HELMHOLTZ ASSOCIATION

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:
  - Definition and implementation of standards (parameters, Thesauri, Metadata profile ...)
  - Improvement of quality assessment of the primary data
  - Inclusion of data sets with ecological content and spatial data (e.g. from remote sensing)
  - Publication of primary data using persistent Digital Object Identifiers (DOI)

HELMHOLTZ
ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Zusätzliche Dienste des GFZ

Das GFZ bietet eine Reihe von zusätzlichen Diensten für Forschungsdaten an:

- Datenveröffentlichung mit DOI
- Persistente Identifikatoren für Probenmaterial (IGSN)
- Datenmanagementsystem für dateibasierte Daten

- DOI und IGSN können miteinander verknüpft werden.
- Daten können aus dem Datenmanagementsystem heraus mit DOI veröffentlicht werden.

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Forschungsdaten in der Praxis

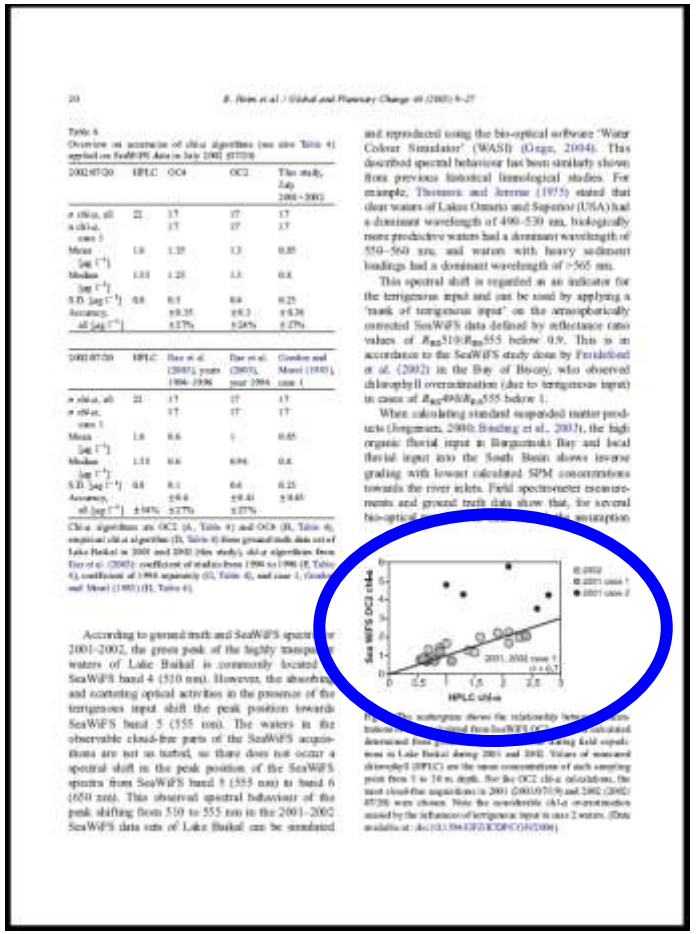


Table 6
Overview on accuracies of chl-a algorithms (see also Table 4) applied on SeaWiFS data in July 2002 (07/20)

2002/07/20	HPLC	OC4	OC2	This study, July 2001–2002
n chl-a, all	22	17	17	17
n chl-a, case 1		17	17	17
Mean [µg l⁻¹]	1.6	1.35	1.3	0.85
Median [µg l⁻¹]	1.55	1.25	1.3	0.8
S.D. [µg l⁻¹]	0.8	0.5	0.6	0.25
Accuracy, all [µg l⁻¹]		±0.35 ±37%	±0.3 ±24%	±0.26 ±27%

2002/07/20	HPLC	Darecki et al. (2005), years 1996–1998	Darecki et al. (2005), year 2004	Gordon and Morel (1983), case 1
n chl-a, all	22	17	17	17
n chl-a, case 1		17	17	17
Mean [µg l⁻¹]	1.6	0.6	1	0.85
Median [µg l⁻¹]	1.55	0.6	0.94	0.8
S.D. [µg l⁻¹]	0.8	0.1	0.6	0.25
Accuracy, all [µg l⁻¹]	±34%	±0.6 ±37%	±0.41 ±37%	±0.45

Chl-a algorithms are OC2 (A, Table 4) and OC4 (B, Table 4), empirical chl-a algorithm (D, Table 4) from ground truth data set of Lake Baikal in 2001 and 2002 (this study), chl-a algorithms from Darecki et al. (2005): coefficient of relation from 1998 to 1996 (E, Table 4), coefficient of 1998 separately (G, Table 4), and case 1, Gordon and Morel (1983) (H, Table 4).

According to ground truth and SeaWiFS spectra for 2001–2002, the green peak of the highly transparent waters of Lake Baikal is commonly located in SeaWiFS band 4 (510 nm). However, the absorbing and scattering optical activities in the presence of the terrigenous input shift the peak position towards SeaWiFS band 5 (555 nm). The waters in the observable cloud-free parts of the SeaWiFS acquisitions are not so turbid, so there does not occur a spectral shift in the peak position of the SeaWiFS spectra from SeaWiFS band 5 (555 nm) to band 6 (650 nm). This observed spectral behaviour of the peak shifting from 510 to 555 nm in the 2001–2002 SeaWiFS data sets of Lake Baikal can be simulated and reproduced using the bio-optical software 'Water Colour Simulator' (WASI) (Gege, 2004). This described spectral behaviour has been similarly shown from previous historical limnological studies. For example, Thomson and Jerome (1955) stated that clear waters of Lakes Ontario and Superior (USA) had a dominant wavelength of 490–530 nm, biologically more productive waters had a dominant wavelength of 550–560 nm, and waters with heavy sediment loadings had a dominant wavelength of >565 nm.

This spectral shift is regarded as an indicator for the terrigenous input and can be used by applying a 'mask of terrigenous input' on the atmospherically corrected SeaWiFS data defined by reflectance ratio values of $R_{RS}510/R_{RS}555$ below 0.9. This is in accordance to the SeaWiFS study done by Fietzek et al. (2002) in the Bay of Biscay, who observed chlorophyll overestimation (due to terrigenous input) in cases of $R_{RS}490/R_{RS}555$ below 1.

When calculating standard suspended matter products (Jorgensen, 2000; Reuter et al., 2003), the high organic fluvial input in Barguzinsky Bay and local fluvial input into the South Basin shows inverse grading with lowest calculated SPM concentrations towards the river inlets. Field spectrometer measurements and ground truth data show that, for several bio-optical [illegible] assumption

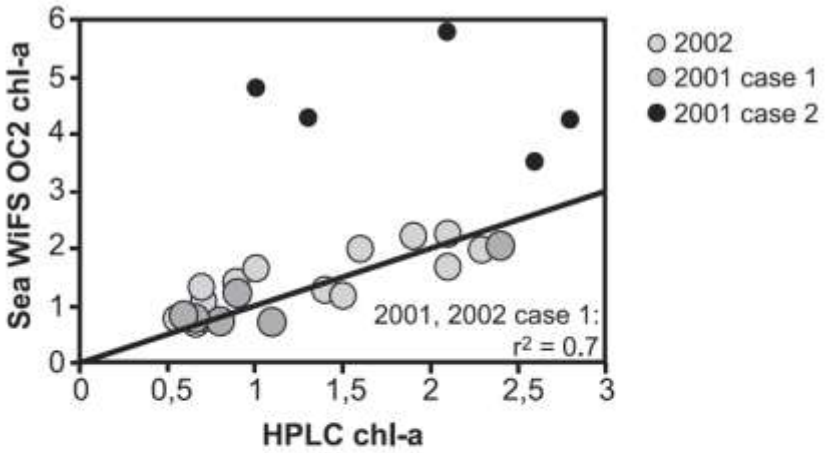
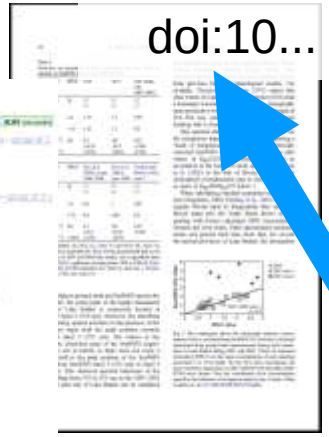


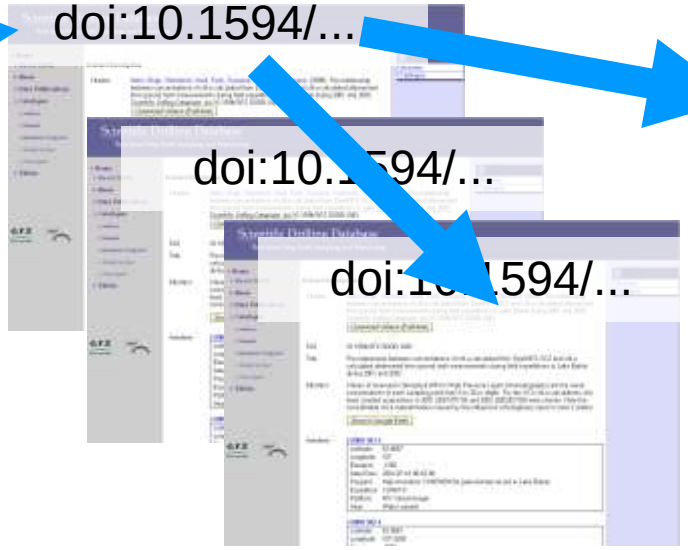
Fig. 2. The scattergram shows the relationship between concentrations of chl-*a* calculated from SeaWiFS OC2 and chl-*a* calculated determined from ground truth measurements during field expeditions in Lake Baikal during 2001 and 2002. Values of measured chlorophyll (HPLC) are the mean concentrations of each sampling point from 5 to 30 m depth. For the OC2 chl-*a* calculations, the most cloud-free acquisitions in 2001 (2001/07/19) and 2002 (2002/07/20) were chosen. Note the considerable chl-*a* overestimation caused by the influences of terrigenous input in case 2 waters.

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Literatur, Daten, Proben





IGSN hdl: ...

HELMHOLTZ
ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

GFZ - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

pmd.gfz-potsdam.de/panmetaworks/showshort.php?id=escidoc:43440

Scroogle SSL German

GFZ

GFZ
Helmholtz Centre
POTSDAM

Helmholtz Centre Potsdam
GFZ GERMAN RESEARCH CENTRE
FOR GEOSCIENCES

HELMHOLTZ
ASSOCIATION

Dataset Description

Cite as Damian Ulbricht(2011): panMetaDocs – A tool for collecting and managing digital objects in a scientific research environment. Deutsches GeoForschungsZentrum. http://dx.doi.org/10.5072/GFZ.PMD.panMetaDocs

Abstract Data management in scientific projects is a challenging task. In many cases projects operate with a limited budget for data management that does not allow the development of customized software for data curation. On an institutional scale research data in the earth sciences are described by a number of different metadata schemata. panMetaWorks , which is the precursor to panMetaDocs , was developed to collect metadata and data in collaborative projects situated at more than one institution. Internet browsers allow easy interaction with panMetaWorks' PHP-based web user interface. Metadata are entered and data objects uploaded through this graphical user interface. A key feature of panMetaWorks is its ability to accommodate any metadata schema. The metadata fields can be filled with static or dynamic default entries to make use of the information implicitly available from the project context. This feature reduces the number of fields that require manual entries to a minimum. The business logic of panMetaWorks is reused in the development of panMetaDocs, except for authentica- tion and data management functions of panMetaWorks, which are delegated into the repository framework eSciDoc. The eSciDoc repository framework is designed as a Service Oriented Architecture and can be controlled via a REST interface that is accessed by panMetaDocs to create eSciDoc repository items. The framework is designed as an institution-wide data archiving infrastructure and can be accessed by more than one application instance. Once data objects are uploaded to the eSciDoc infrastructure it is even possible to drop the software instance that was used to collect the data while the collected data and metadata reside in the eSciDoc infrastructure and are available for use in other applications. This approach of expendable data curation tools allows for a significant reduction in costs for software maintenance. panMetaDocs' intention is to allow easy collaboration within a project, collect and curate experimental and measurement data and transfer data objects from a shared into a persistent data curation domain. To accomplish this, only a subset of the lifecycle of eSciDoc items is used. During the workflow starting from the state "pending", through stage "submitted" to the final status "released", objects will be moved from the shared data curation domain to the persistent domain and become available for publication of their data and metadata through data portals. Review and publication of data is, in the case of GFZ Potsdam, a service of its library and therefore the transfer of items to the status "released" is not part of the initial panMetaDocs development. With a RSS interface for recent datasets the reused business logic of panMetaWorks allows project members to be informed about data contributions by other project members. The implementation of the Open Archives Initiative Protocol for Metadata Harvesting interface (OAI-PMH) [4], which is also part of panMetaWorks, can be used to syndicate data catalogues to research data portals. panMetaWorks and panMetaDocs are optimised to serve the panFMP data portal framework.

Location Latitude: 52.3795 Longitude: 13.0648

Keywords Oceans panMetaDocs Geoscientific Information

Data PanMetaDocs-EGU.pdf 2004749 Bytes

Metadata +

Karte

Imprint

© Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Metadata | File(s)

Preferences:

Entry form type: DataCite-2.0 Change Destination Folder: DOI: Do DOI operation

Identification

Title: panMetaDocs – A tool for collecting and managing digital objects in a scientific research environment

Date: 2011-12-07

Author (1): Damian Ulbricht

Description:

Data management in scientific projects is a challenging task. In many cases projects operate with a limited budget for data management that does not allow the development of customized software for data curation. On an institutional scale research data in the earth sciences are described by a number of different metadata schemata. panMetaWorks , which is the precursor to panMetaDocs , was developed to collect metadata and data in collaborative projects situated at more than one institution. Internet browsers allow easy interaction with panMetaWorks' PHP-based web user interface. Metadata are entered and data

Project Name:

NASA DIF Data

Southernmost Latitude: 52.3795

Northernmost Latitude: 52.3795

Westernmost Longitude: 13.0648

Easternmost Longitude: 13.0648

Category: EARTH SCIENCE

Topic: Oceans

Term/Keyword (1): panMetaDocs

ISO Topic: Geoscientific Information

Data Center Contact: ulbricht@gfz-potsdam.de

Publication

DOI: 10.5072/GFZ.PMD.panMetaDocs

Resource Type: Dataset

Publisher (1): Deutsches GeoForschungsZentrum

Publisher Short: GFZ Potsdam

Publication Year: 2011

language: en

copyright:

File upload (max. 100 Mb per file)

Please choose the file you wish to upload (txt, html usw.):

Durchsuchen... Upload

WARNING: Do not distribute copyrighted material for which you don't own the rights or have permission from the owner.

Dataset versions:

2012-01-02 12:25:30 (current) Change

Dataset Files:

PanMetaDocs-EGU.pdf [2 Mb]