TERENO data management

CT Data Management

Ralf Kunkel, Martin Abbrent, Anusuriya Devaraju, Rainer Gasche, Bernd Herrmann, Mark Frenzel, Jens Klump, Olaf Kolditz, Frank Neidl, Marcel Pflug, Karsten Rink, Thomas Schnicke, Matthias Schroeder, Jürgen Sorg, Vivien Stender, Ute Wollschläger

TERENO Advisory Board Meeting

September, 16-17 2013, Klink

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Outline

- Accessing and querying data from TEODOOR
  - Data visualization and access
  - New search application
- Current status
  - Processed and published data,
  - Data access through TEODOOR
- Quality assessment of sensor data
  - Modelling data quality
  - Quality assessment workflow
  - Quality flagging tool
- Conclusions and outlook

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

TEODOOR – access to data

- Central data portal accesses the individual services and allows information exchange, data search and data access
- All observatories are connected to TEODOOR
- Redesigned data query and access
  - Predefined views to individual observatories
  - Access to weather radar device data
  - Spatial and thematic search to data

PUBLIC TERENO DATA PORTAL

One task TERENO is to provide long-term statistical series of system variables for the analysis and prognosis of Global Change consequences using integrated model systems, which will be used to derive efficient prevention, mitigation and adaptation strategies. Within the distributed Spatial Data Infrastructure TEODOOR each institution responsible for an individual observatory sets up its own local data infrastructure. This portal applicaton brings together together the data collected and published via OGC Web-services from the individual observatories and provides access of the data to the public. Therefore, it serves as a database node to provide scientists and decision makers with reliable and well accessible data and data products.

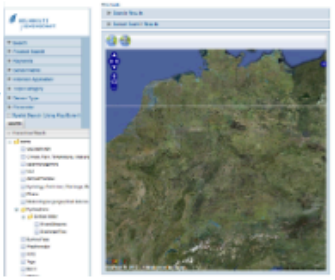
To provide an easy access to the data collected within TERENO, several vizualization and query tools are available:

You can get an overview of the stations and the collected data from all or individual observatories either by clicking into the map of Germany or into one of the observatories or by selecting one of:

- Eifel/Lower Rhine Valley Observatory
- Bavarian Alps / pre-Alps Observatory
- German Lowland Observatory
- Harz/Central German Lowland Observatory
- All four TERENO Observatories

Data from the weather radar devices currently available for the Eifel and the Bavarian Alps observatories can be found here:

- Eifel/Lower Rhine Valley Observatory
  - Jülicher Weatherradar Sophienhöhe
  - Rainscanner Wüstebach
- Bavarian Alps / pre-Alps Observatory (coming soon)



Searching to specific data from all observatories, i.e.

- sensor data with respect e.g. to observed phenomenons, sensor types, station names etc.
- filebased data according to its metadata

can be performed by clicking into the screen shot in the picture to the right

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Data being processed at individual TERENO observatories

	Eifel/Lower Rhine	Harz/Central Lowland	Northeastern Lowland	Bavarian Alps and Prealps	
				HMGU	IMK/IFU
Climate, soil, water	589 stations 980000 obs/d	75 stations 125000 obs/d	179 stations 320000 obs/d	95 stations 848000 obs/d	8 stations 52128 obs/d
Eddy-Covariance	7 stations 133000000 obs/d	3 stations 57000000 obs/d	3 stations 57000000 obs/d	1 station 1900000 obs/d	4 stations 76000000 obs/d
Weather radar	2 devices 576 rasters/d	1 device 288 rasters/d			1 device 288 rasters/d
SoilCan	36 lysimeters 285000 obs/d	30 lysimeters 238000 obs/d	12 lysimeters 95000 obs/d	6 lysimeters 47500 obs/d	42 lysimeters 333000 obs/d

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Data being published via TEODOOR

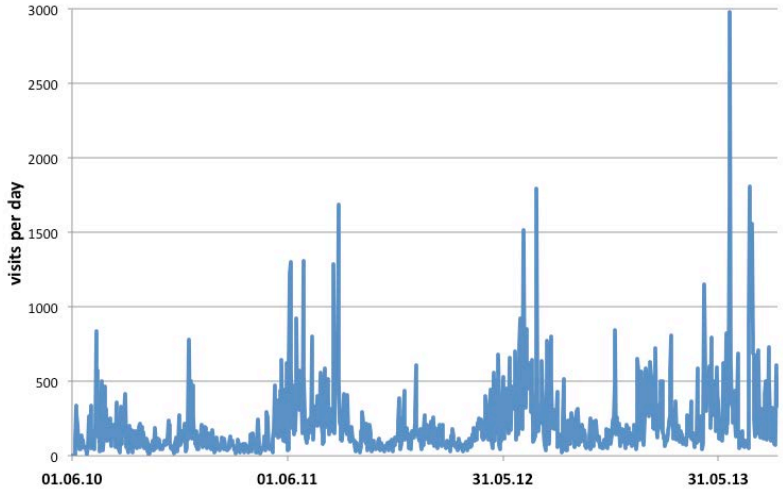
	Eifel/Lower Rhine	Harz/Central Lowland	Northea-stern Lowland	Bavarian Alps and Prealps	
				HMGU	IMK/IFU
Climate, soil, water	589 stations	(75 stations)	20 stations	In preparation	-
Eddy-Covariance	7 stations (climate data)	-	-	In preparation	2 stations (climate data)
Weather radar	2 devices	-			In preparation
SoilCan	In preparation	-	-	-	-

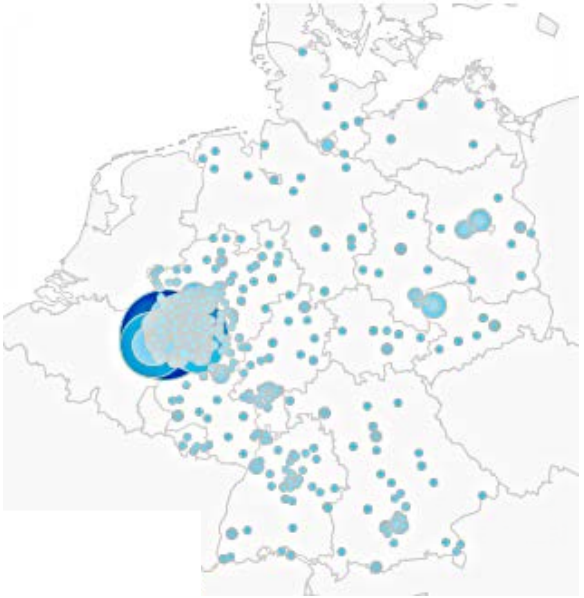
HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

TEODOOR access statistics 1.6.2012 – 31.8.2013

- Visitors: 92 800
- Visits: 241 369
  - 136 countries
  - Germany: 94 %
  - Belgium, Netherlands, USA: 1%
- Page calls: 486.985
  - Weather radar: 291 912
  - Observatories: 26 226
  - Data portal: 9 232
  - Other: 159 615
- Access:
  - Direct: 50.7 %
  - Google: 19.5 %
  - FZJ ref: 2.9 %
  - UFZ ref: 0.8 %
  - GFZ ref: 0.3 %
  - HMGU, IMK ref: < 0.01 %

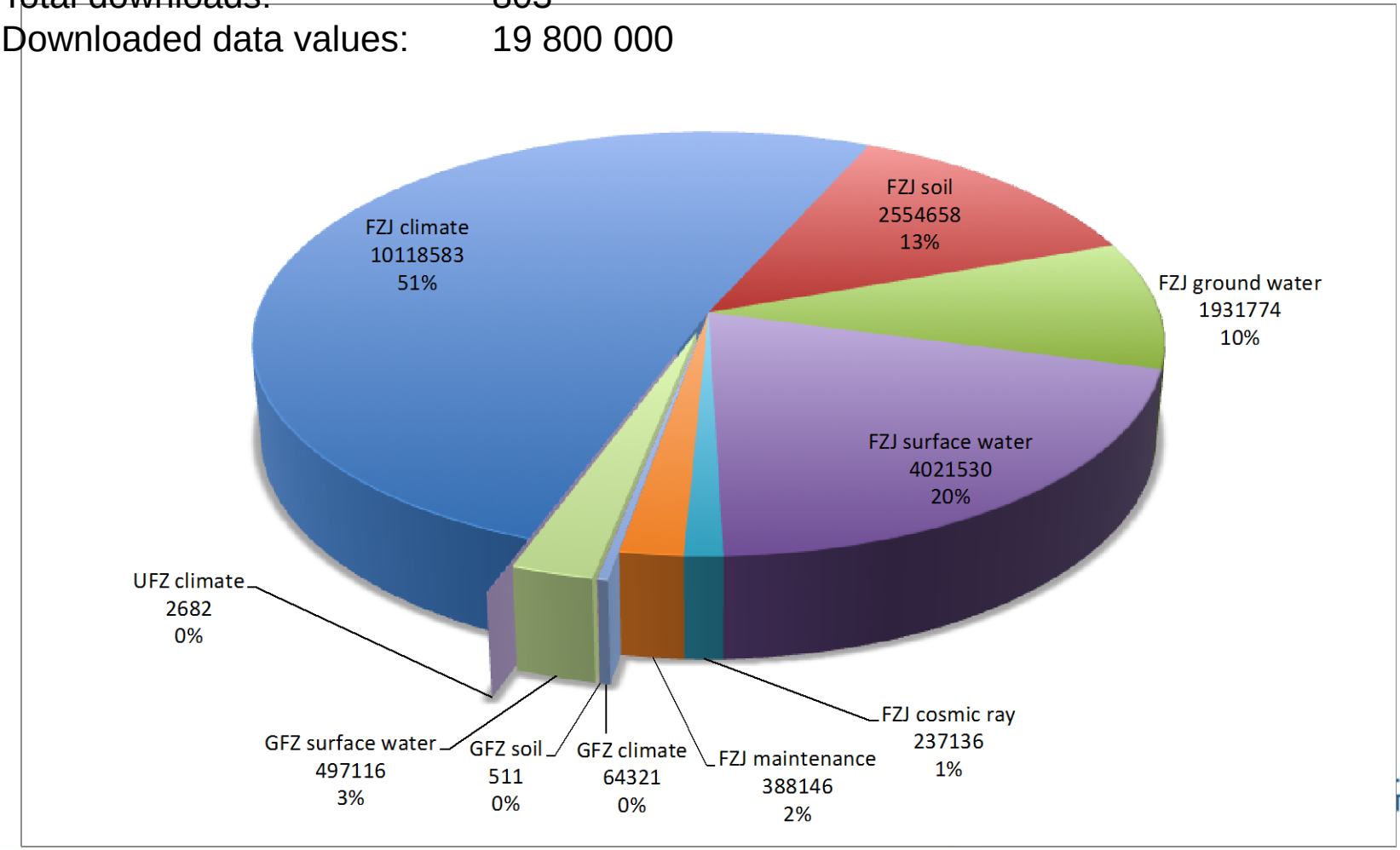




TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Downloaded sensor data from TEODOOR

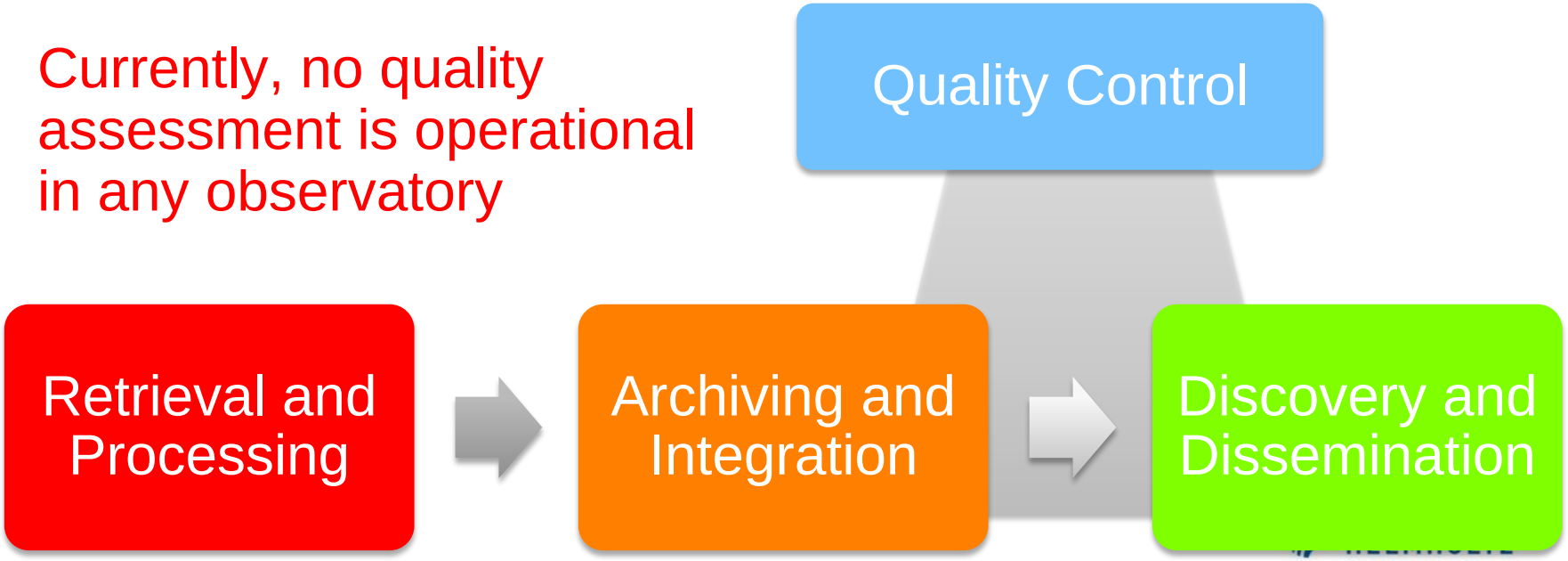
Time period: 12/2012-08/2013
Total downloads: 803
Downloaded data values: 19 800 000



TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Quality Contol of sensor data

- In each observatory up to 2,000,000 data values are collected each day
- Data quality is a primary concern when building trust in shared data.
- Currently, no quality assessment is operational in any observatory



ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Implementing Quality Control into TERENO

- Modelling Quality Control (QC) information of TERENO observations
  - Data types and quality control methods
  - Data processing status/level
  - Data quality flag convention
  - Quality control workflow
- Incorporating QC information into the TERENO infrastructure
  - Observational data model
  - Observational service including request and response messages
  - Quality flagging tool
  - Administration of users and flagging classification

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Digital Object Identifiers (DOI)

- Identification of content objects (e.g. data sets) in the digital environment sets by unique identifiers

  Example: http://dx.doi.org/10.1594/TERENO/FZJ_1

- For a dataset to be assigned a DOI, it must be provided in good condition, with appropriate metadata and of a suitable level of technical quality.

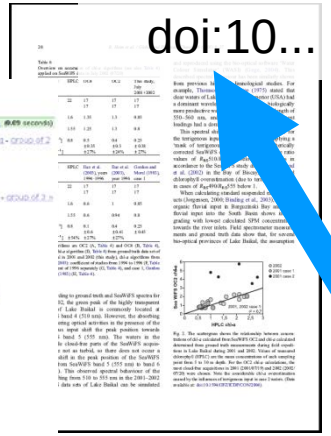
- Within TERENO GFZ may act as a certification authority to assign unique identifiers to
  - Data and documents (DOI)
  - Samples (IGSN)

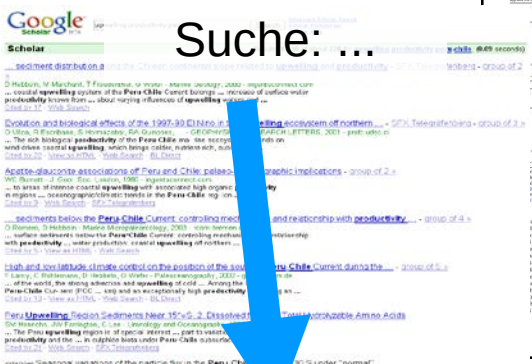
- Unique identification allows the data to be cited

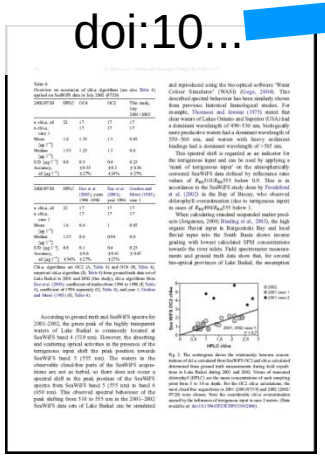
HELMHOLTZ
ASSOCIATION

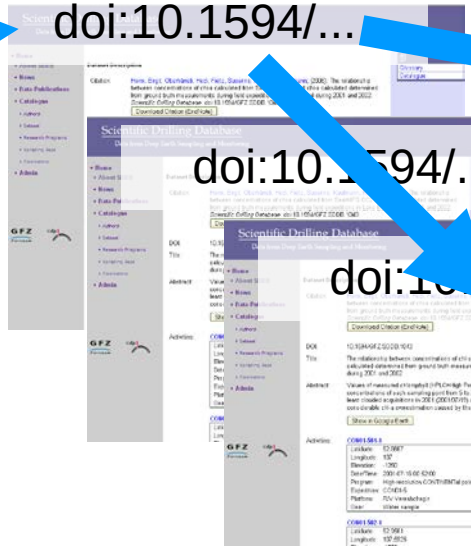
TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Literature, data, samples









IGSN hdl: ...

HELMHOLTZ
ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Conclusions and outlook

- Current status:
  - Local data infrastructures operational and connected to TEODOOR (except HMGU)
  - Data quality modelled, workflows for quality assessment developed and implemented in the data model.
  - Quality flagging tool based on developed workflow implemented
  - Common standards (parameters, Thesauri, Metadata profile ...) defined, implemented and published via TEODOOR
- Outlook:
  - Draft Data Management Plan
  - Operationalizing quality assessment for all observatories
  - 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

Presentations and Publications concerning TERENO data management (2012-2013)

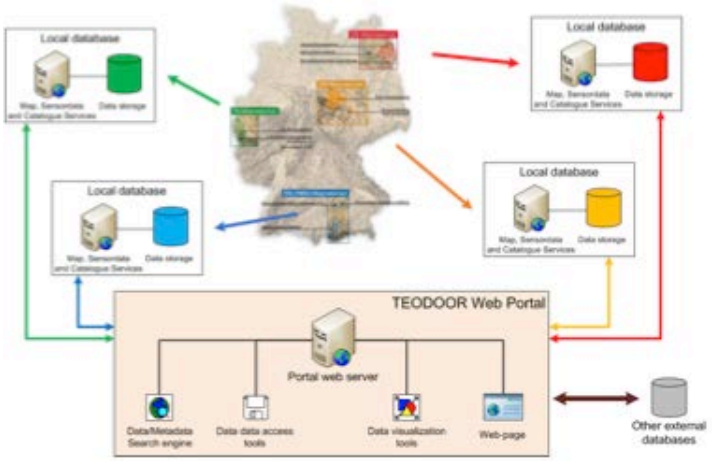
- Publications:
  - Bogena, H., Kunkel, R., Puetz, T., Vereecken, H., Krueger, E., Zacharias, S., Dietrich, P., Wollschlaeger, U., Kunstmann, H., Papen, H., Schmid, H.P., Munch, J.C., Priesack, E., Schwank, M., Bens, O., Brauer, A., Borg, E., Hajnsek, I. (2012): TERENO - Long-term monitoring network for terrestrial environmental research. **Hydrologie und Wasserbewirtschaftung** (3): 138-143
  - Kolditz, O., Rink, K., Shao, H., Kalbacher, T., Kunkel, R., Zacharias, S., Dietrich, P. (2012): Data modelling platforms in environmental Earth sciences**. Environmental Earth Sciences** (4): 1279-1284, doi: 10.1007/s12665-012-1661-8.
  - Kolditz, O., Rink, K., Shao, H., Kalbacher, T., Kunkel, R., Zacharias, S., Dietrich, P. (2012): International viewpoint and news: data and modelling platforms in environmental Earth sciences (vol 66, pg 1279, 2012)**. Environmental Earth Science**s (6): 1859-1859.
  - Kunkel, R., Sorg, J., Eckardt, R., Kolditz, O., Rink, K., Vereecken, H. (2013): TEODOOR - a distributed geodata infrastructure for terrestrial observation data. **Environmental Earth Sciences** (Secial Issue on Catchment Research, May 2013 issue 69/2.), doi: 10.1007/s12665-013-2370-7.
  - Sorg J. (2012) Entwurf, Implementierung und Anwendung eines OGC-konformen Sensor-Observation Service für flächenbezogene Rasterzeitreihendaten. Master Thesis, Fernuniversität Hagen.
  - Sorg, J., Kirschke, T., Kunkel, R. (2012): Entwicklung eines Sensor Observation Service für flächenbezogene Rasterzeitreihendaten, **GIS.Science,** June 2013
- Conference proceedings
  - **TEODOOR: Geodateninfrastruktur zur Verwaltung und Veröffentlichung von terrestrischen Beobachtungsdaten der HGF Infrastrukturmaßnahme TERENO aus verteilten Quellen,**
    WissKom2012 - Vernetztes Wissen – Daten, Menschen, Systeme. Forschungszentrum Jülich GmbH, Jülich, Germany.
  - **TEODOOR – a Spatial Data Infrastructure for terrestrial observation data.**
    Presentation at the IEEE International Conference on Networking, Sensing and Control (ICNSC13), Paris-Evry, France, 2013-04-10-12.
  - **The Design of Monitoring and Data Infrastructures – Applying a forward-thinking Reference Architecture.**
    Presenatation af the IEEE International Conference on Networking, Sensing and Control (ICNSC13), Paris-Evry, France, 2013-04-10-12.
  - **Incorporating Quality Control Information in the SensorWeb.**
    Poster at European Geosciences Union General Assembly (EGU 2013), Vienna, Austria, 2013-04-07-12.
  - **The architecture design of a monitoring infrastructure for the subobservatory Tereno North-East.**
    Poster at European Geosciences Union General Assembly (EGU 2013), Vienna, Austria, 2013-04-07-12.
  - **User Guided Generation of Hydrological Models: Interface Design, Workflows and Concepts for an Extensible Software Framework.** 6th International Congress on Environmental Modelling and Software (iEMSs), Leipzig, Germany, 01.-05. July 2012.

HELMHOLTZ ASSOCIATION

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

TEODOOR - the TERENO data infrastructure

- Individual data infrastructures for each observatory
- OGC-compliant Web Services operated for each observatory provide access to data
- OGC-compliant metadata services provide descriptive information on monitoring stations and measured data
- Central data portal accesses the individual services and allows information exchange, data search and data access

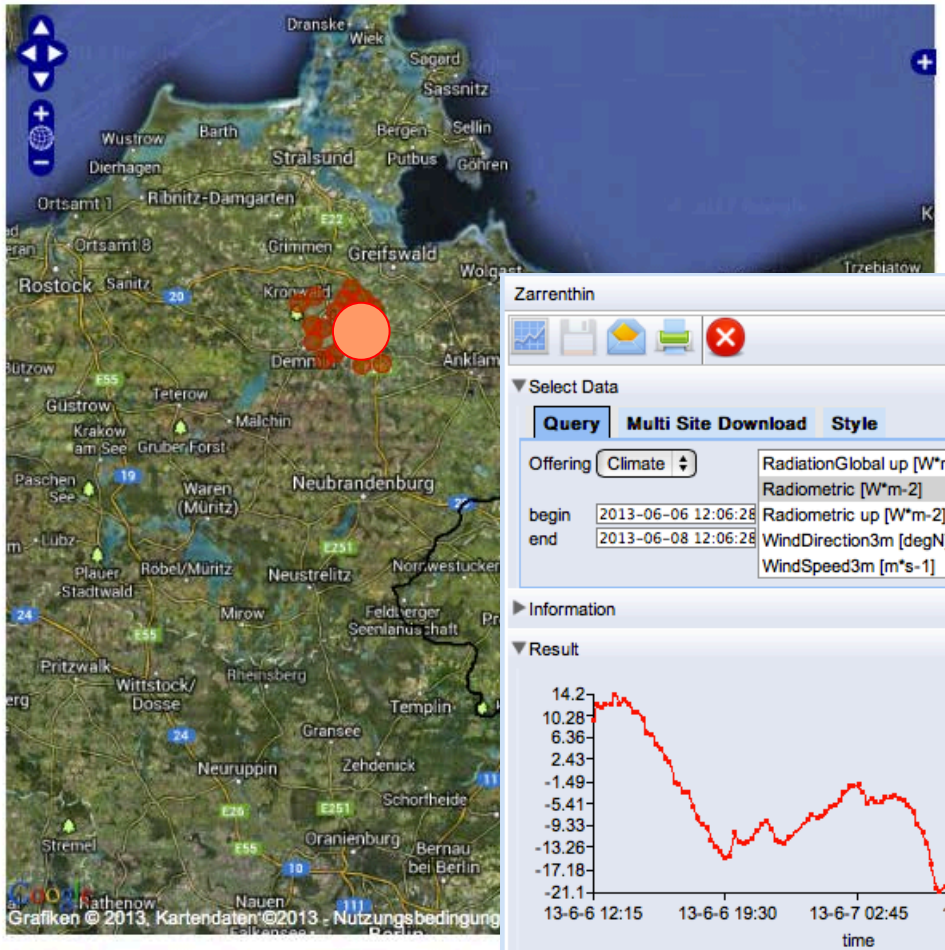


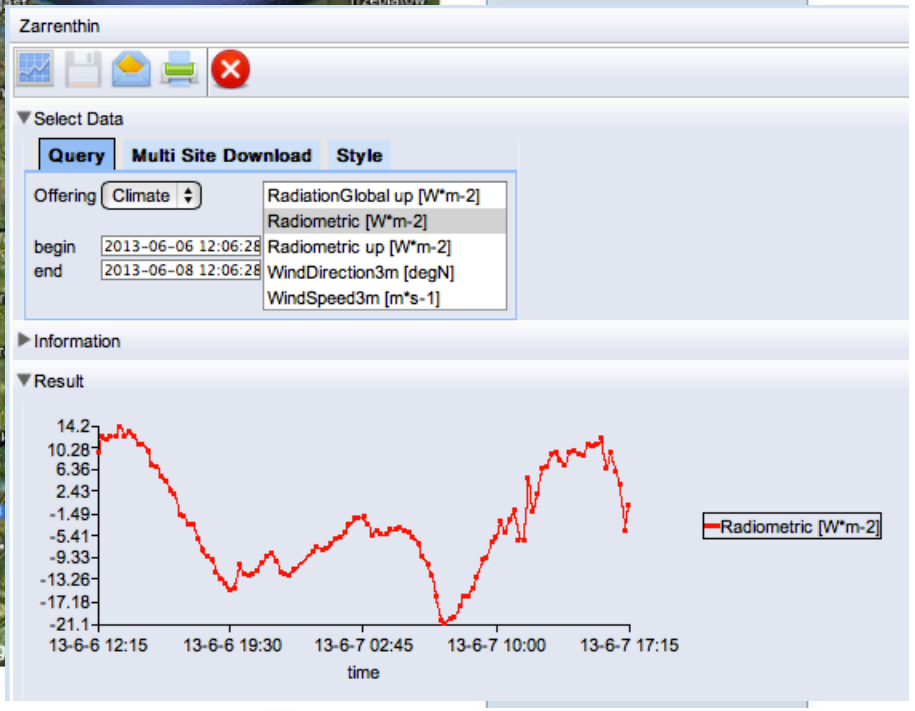
TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Data visualization in TEODOOR

HELMHOLTZ GEMEINSCHAFT

- BaseData
- AIDAGeoserver
- Search
  - Freetext Search
  - Keywords
  - Sensor Name
  - Intended Application
  - Topic Category
  - Sensor Type
  - Parameter
  - Spatial Search (Using Map Extent)
  - search
- Überblick
- Koordinationsteams
- Observatorien
- TERENO Forum
- TERENO Online Data Portal
- TERENO Newsletter
- Meetings

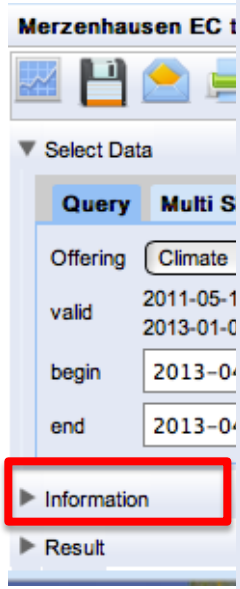


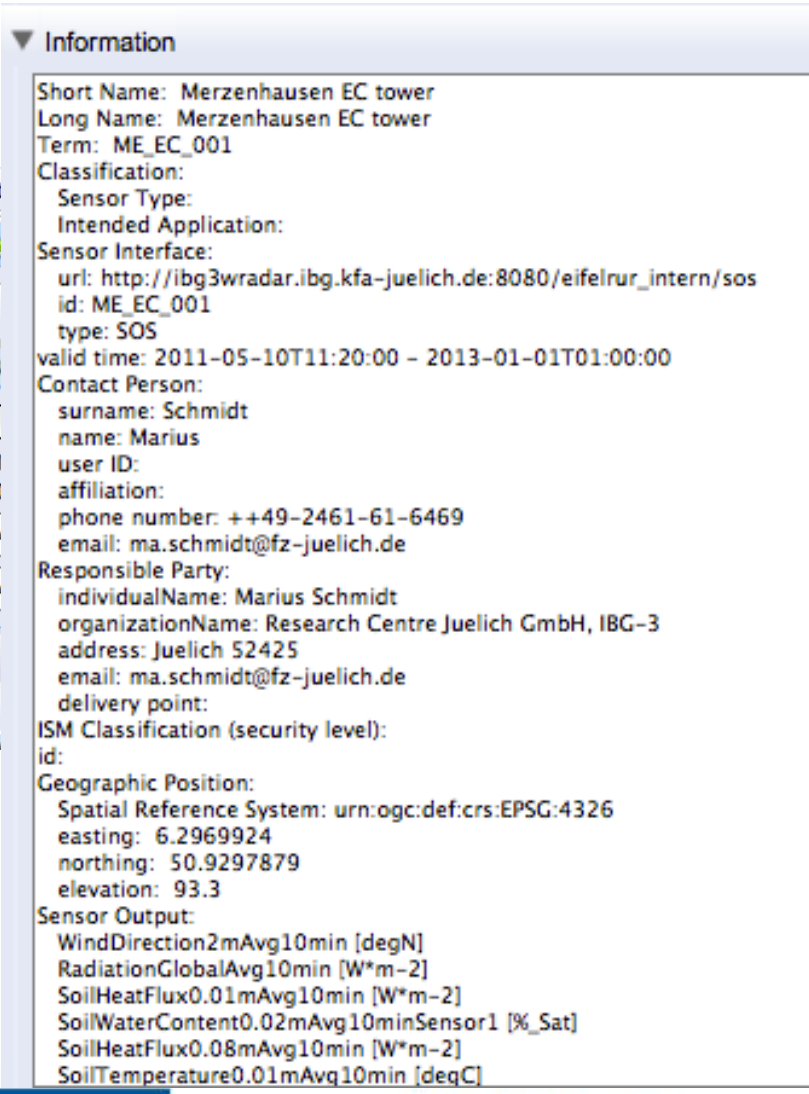


TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Visualising and accessing time series data

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

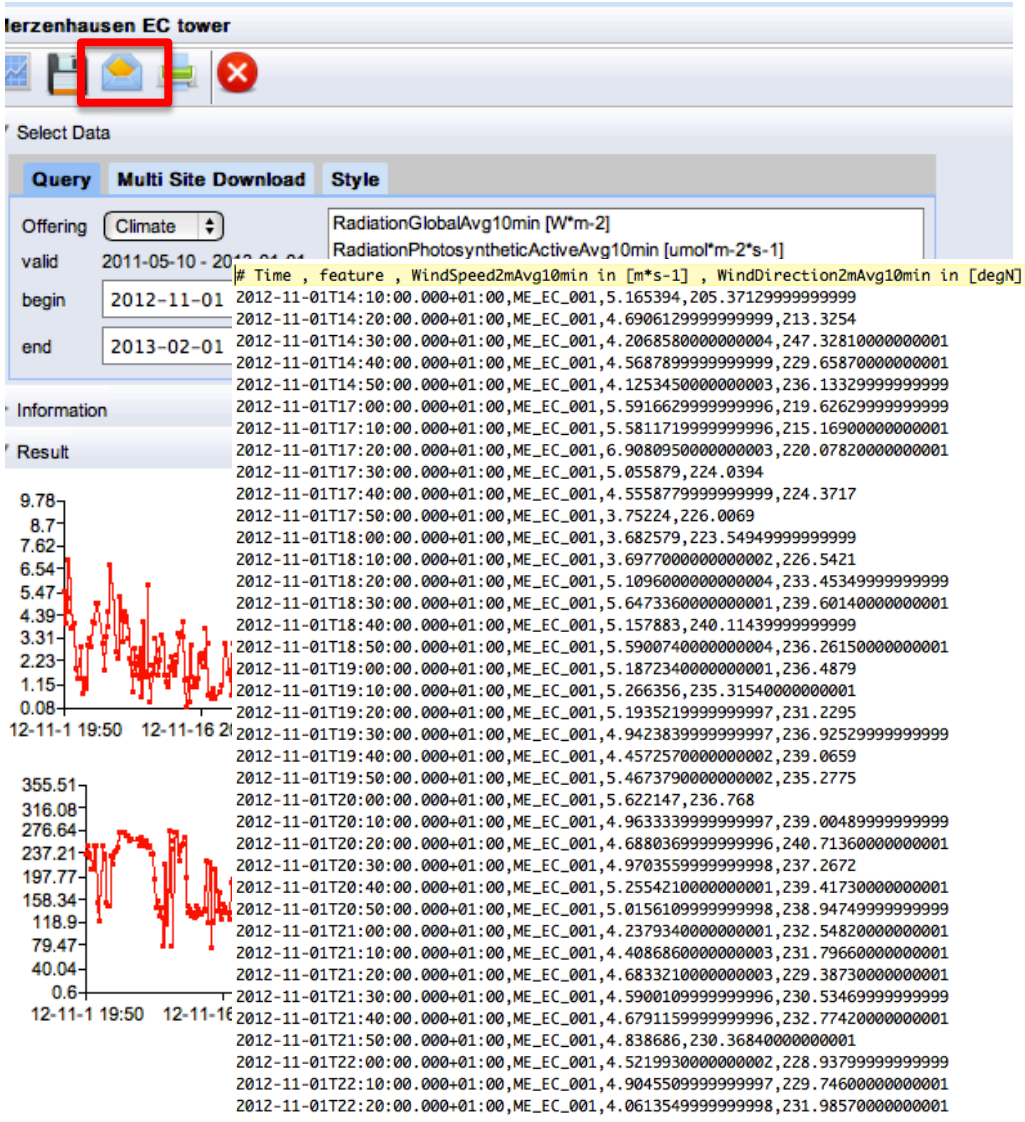




TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

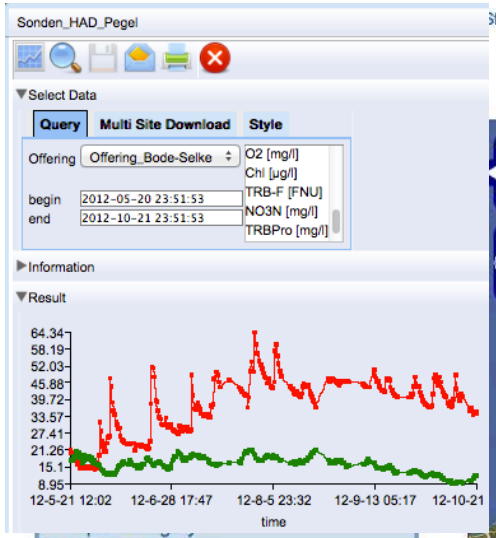
Visualising and assessing time series data

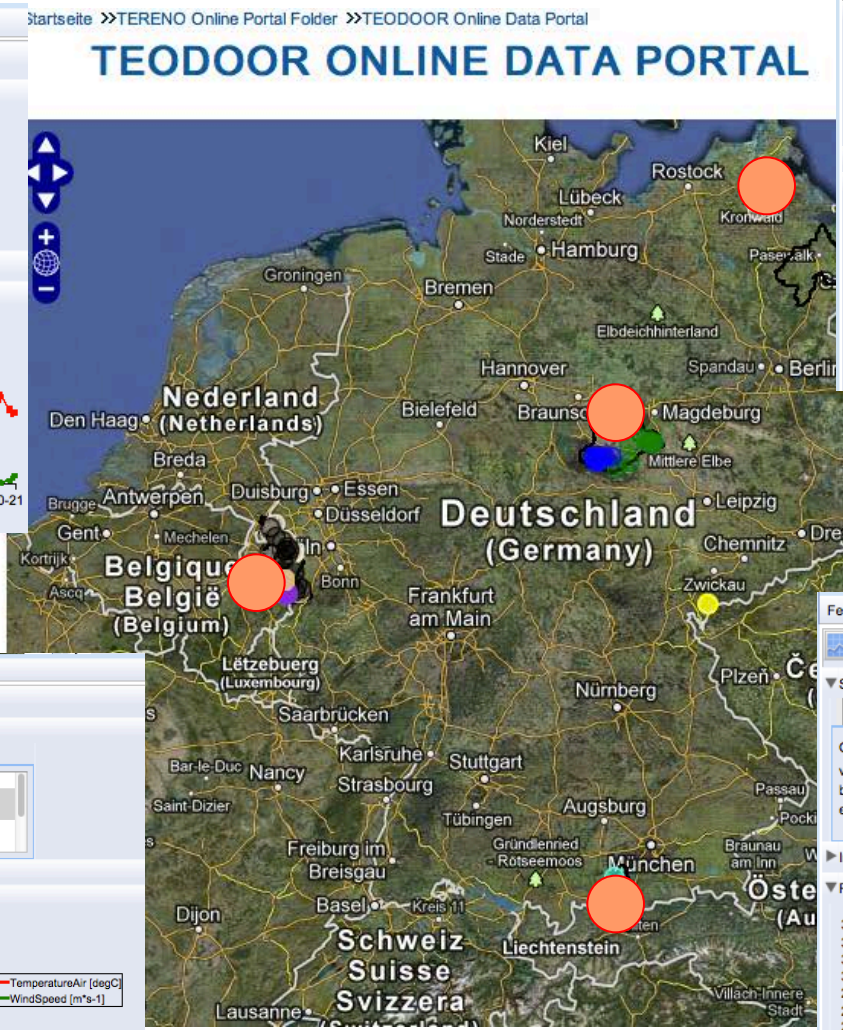
- Connecting to OGC-SOS services
- Graphical selection of stations
- Display of:
  - Latest observations
  - Offerings
  - Available parameters
  - Station information (sensorML metadata)
- Visualisation of station data time series
- Data download (E-Mail, notification of responsible persons)
- Multisite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

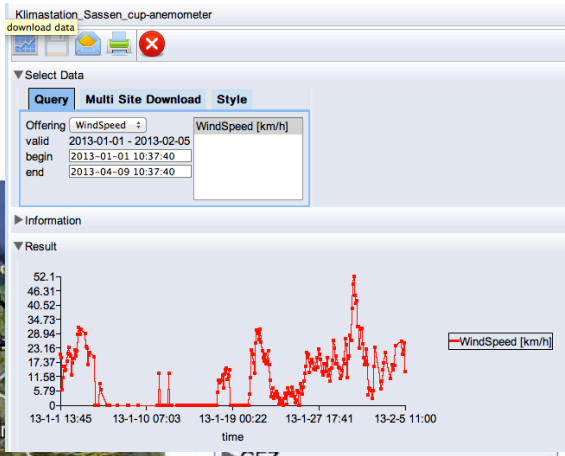


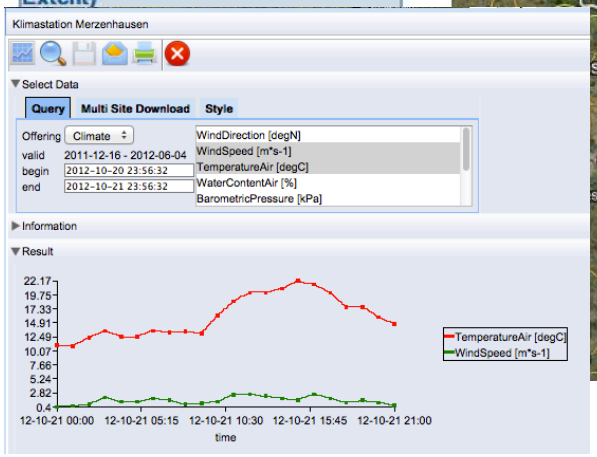
TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

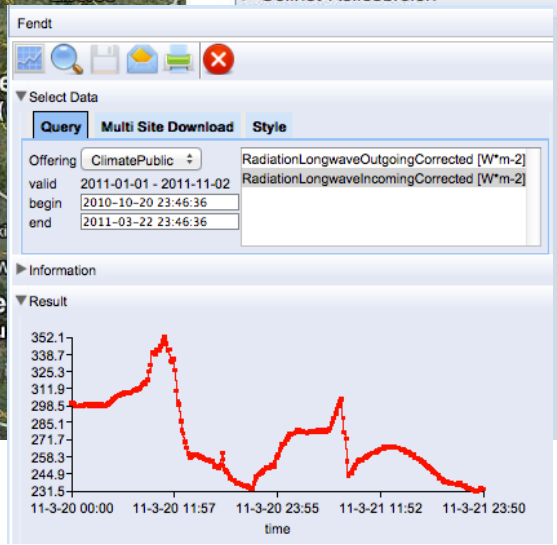
Data visualization in TEODOOR

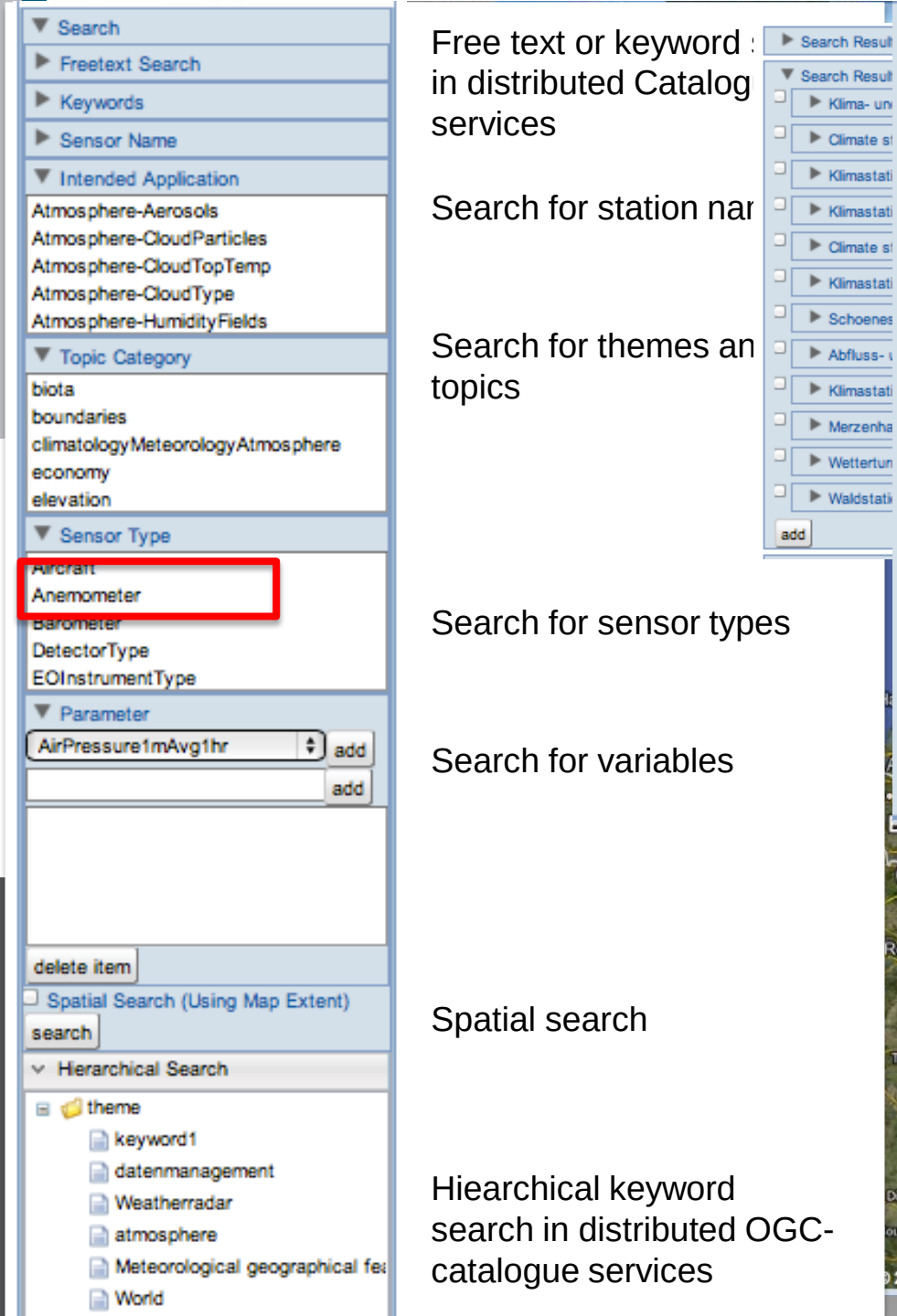












Free text or keyword s
in distributed Catalog
services

Search for station nar

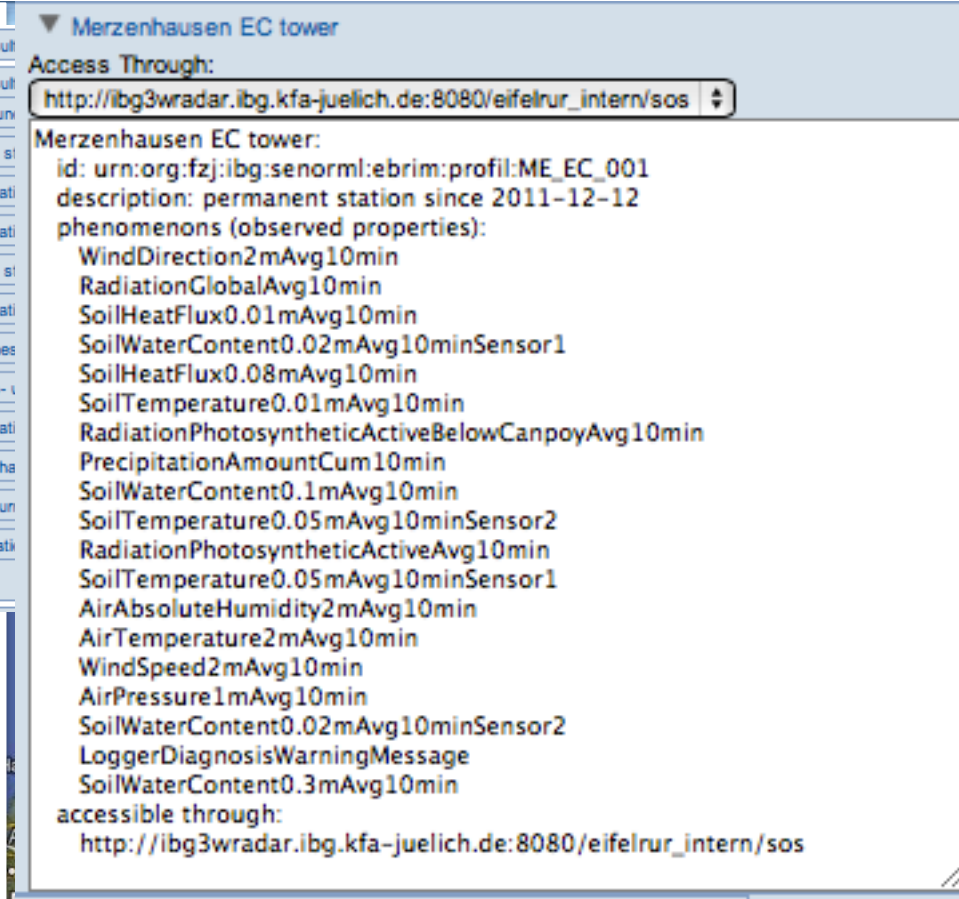
Search for themes an
topics

Search for sensor types

Search for variables

Spatial search

Hiearchical keyword
search in distributed OGC-
catalogue services



TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Data Types and QC Methods

Type	Descriptions	QC method
D1	Automatically-uploaded data; not quality controlled when imported into TERENO database.	Automatic quality assurance procedures & visual inspection
D2	Manually-uploaded, and externally quality-controlled data	Only approval-by-upload needed
D3	User-inserted data; not quality controlled when inserted into TERENO database	Visual inspection

Type D1

Type D2

Type D3

Data Processing Status

Processing Status	Descriptions
Level 1	Raw data (unevaluated)
Level 2a	[1]Externally quality controlled data (without internal approval)
Level 2b	[2]Internally controlled data (with automatic quality assurance procedures (e.g., range check) and visual inspection)
Level 2c	Externally quality controlled data with internal approval
Level 3	Data derived from one or more Level 2(b,c) data series

[1]External : Outside TERENO data infrastructure
[2]Internal : Within TERENO data infrastructure

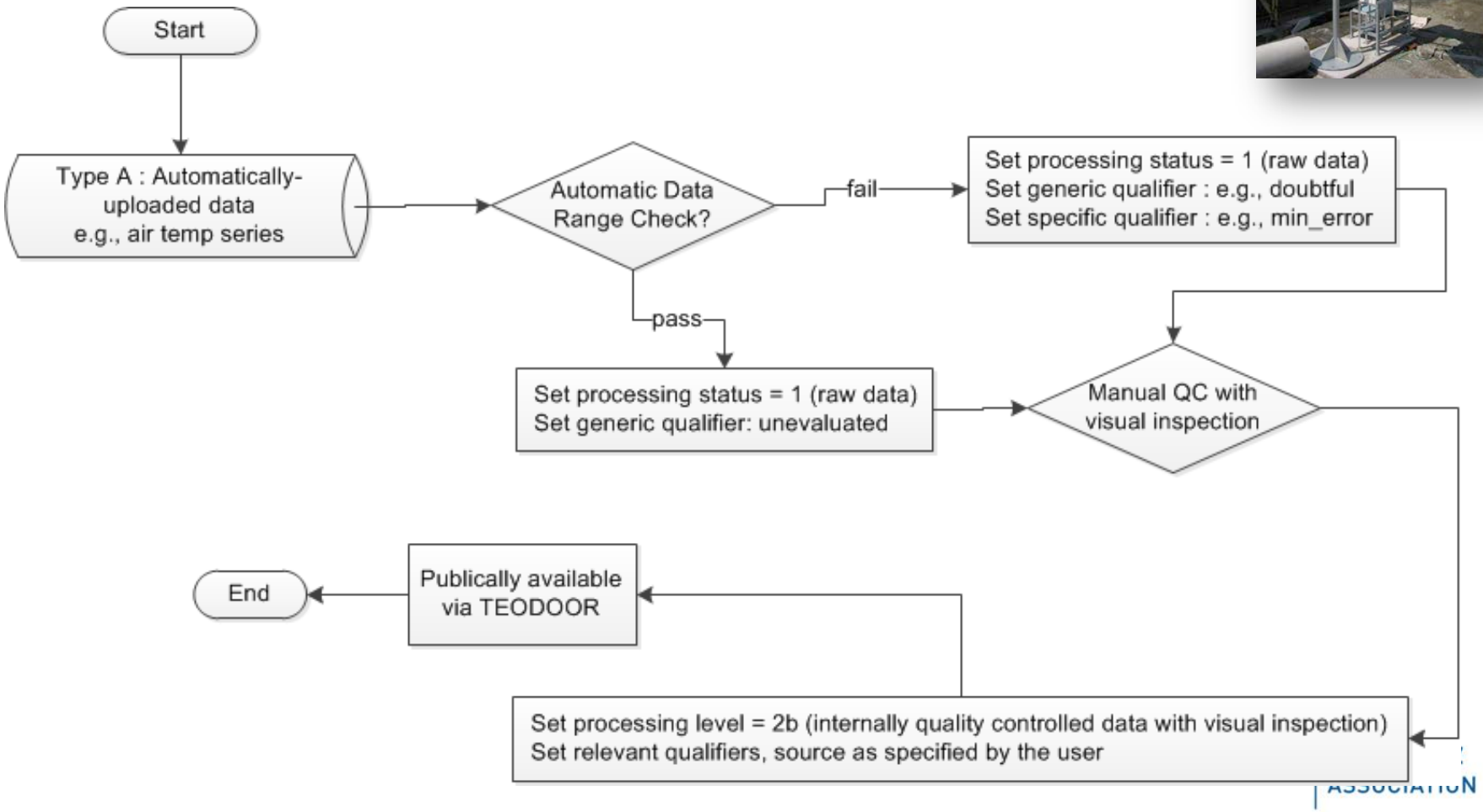
Qualify Flag (Qualifiers) Convention

- Generic qualifiers represents the common quality status of the data and they are fixed.
  - Examples: `unevaluated, ok, bad, doubtful, gapfilled, missing`

- Specific qualifiers extend generic qualifiers
  - they provide additional information about the causes and possible direction of bias, and/or the magnitude of data variability.
  - Specific qualifiers are varied according to observed properties and specified by the experts.
  - Examples: `isospike, minerr, maxerr, interpolated, extrapolated`

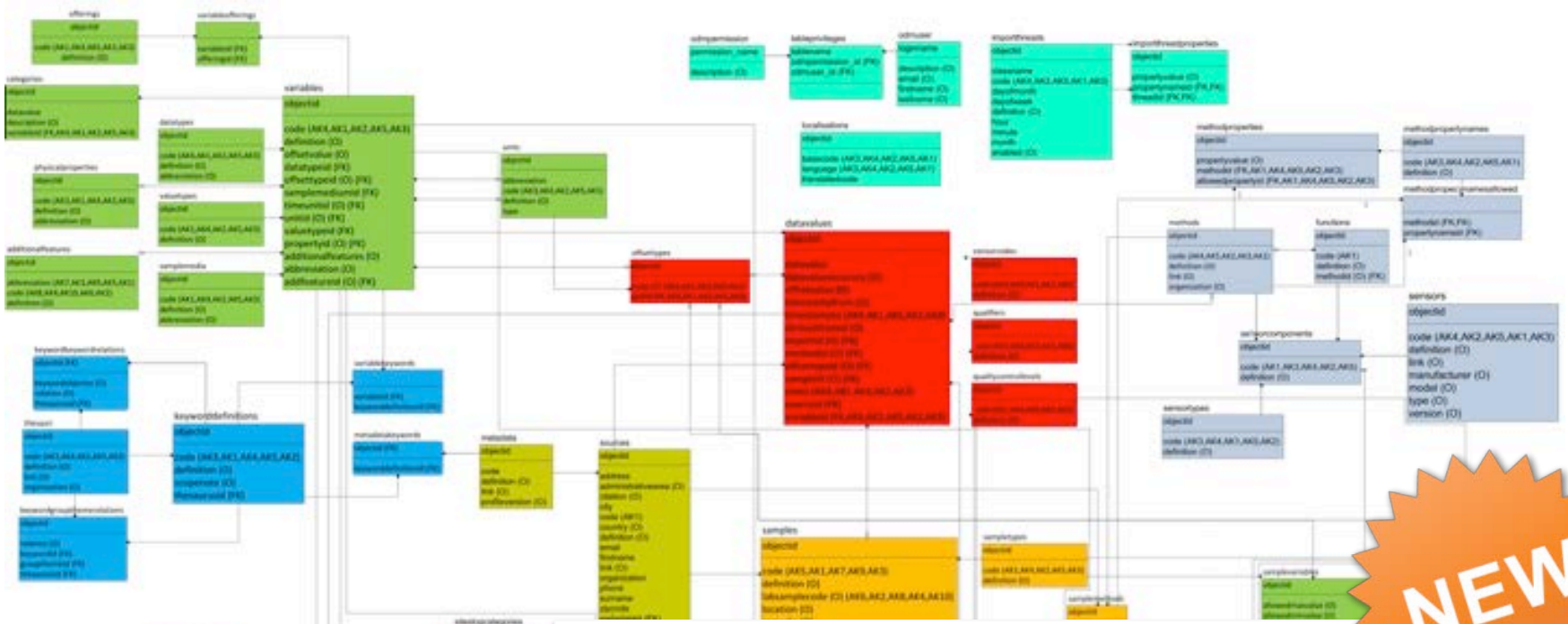
TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Workflow of D1: Automatically-uploaded data



TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Observational Data Model

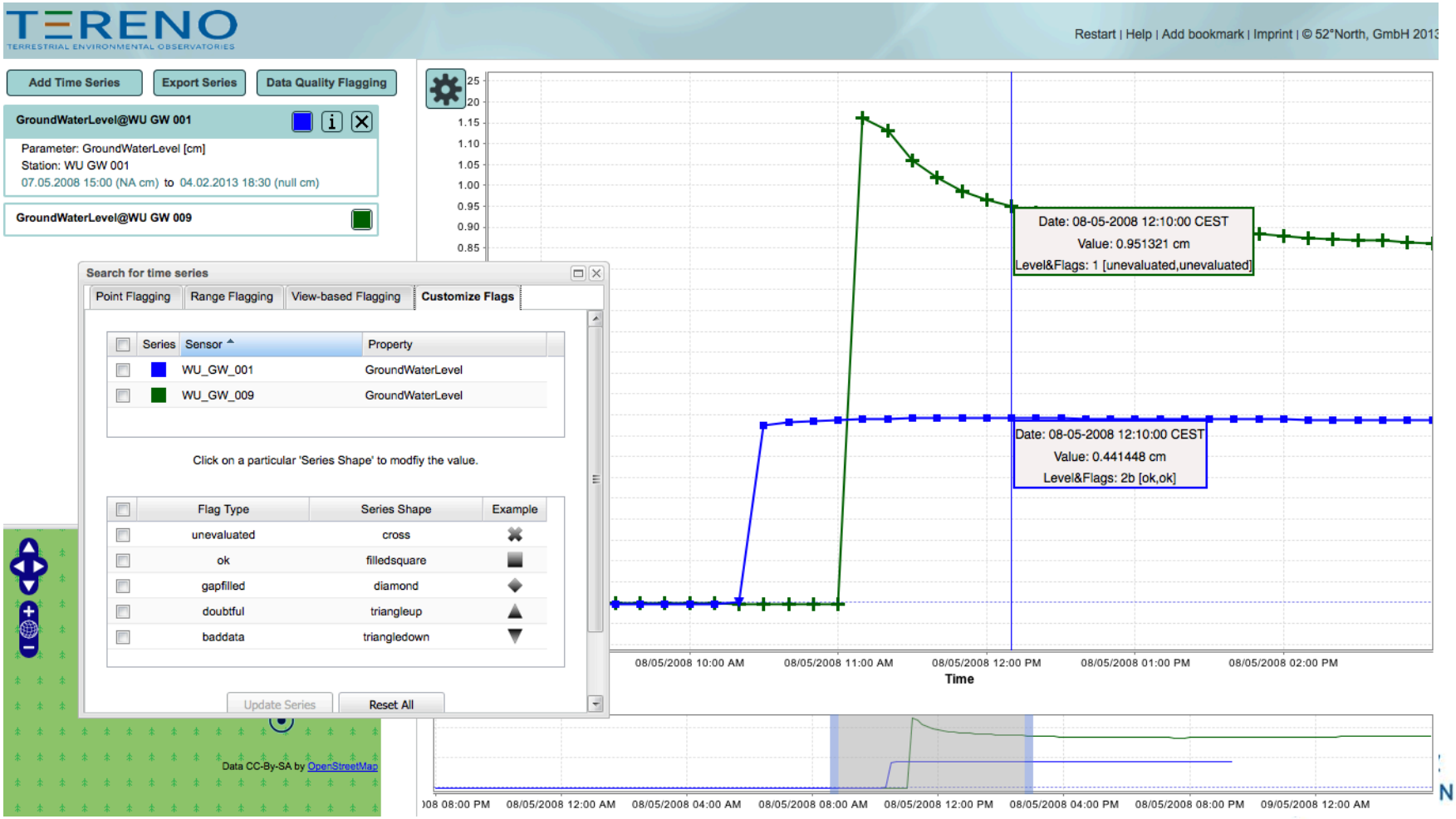


NEW

- Qualifier
- Processing status
- Source-Variables Mapping
- Variables-Qualifiers Mapping

TERENO
TERRESTRIAL ENVIRONMENTAL OBSERVATORIES

Quality flagging tool



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