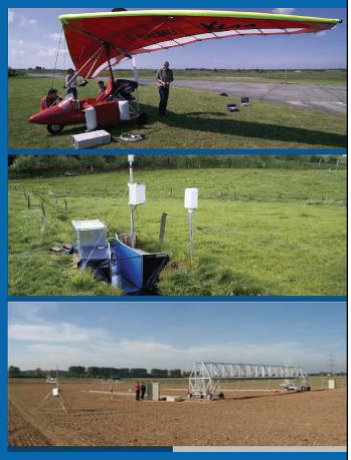
TEODOOR, a blueprint for distributed terrestrial observation data infrastructures



Ralf Kunkel[1], Jürgen Sorg[1], Martin Abbrent[2], Erik Borg[3], Rainer Gasche[4], Olaf Kolditz[2], Frank Neidl[4], Eckard Priesack[5], Vivien Stender[6]

1: Research Centre Juelich, IBG-3, Juelich, Germany (r.kunkel@fz-juelich.de)
2: Helmholtz Centre for Environmental Research, Leipzig, Germany
3: German Aerospace Center, Earth Observation Center, Neustrelitz, Germany
4: Karlsruhe Institute of Technology, Garmisch-Partenkirchen, Germany
5: German Center for Environmental Health, BIOP, Oberschleißheim, Germany
6: GFZ German Research Centre for Geosciences, Centre for Geoinformation Technology, Potsdam

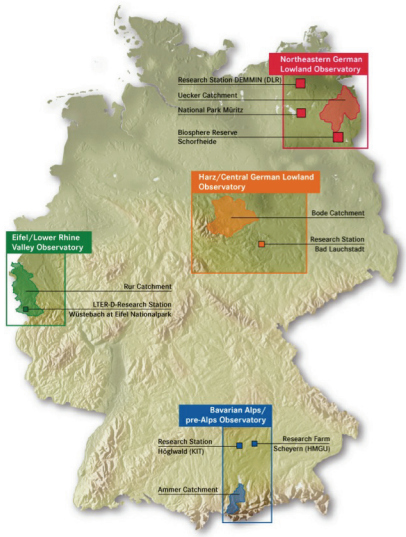
EGU General Assembly 2017, Vienna, Austria, 23 - 28 April 2017

HELMHOLTZ ASSOCIATION

The TERENO network

JÜLICH FORSCHUNGSZENTRUM

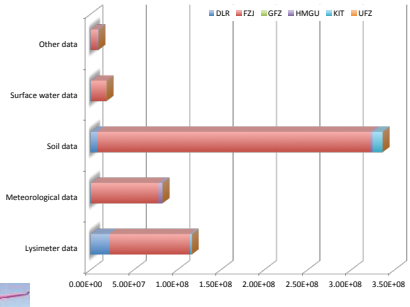
- Regional different effects of global climate change on terrestrial systems
- Global change affects all terrestrial compartments (water, soil, vegetation, atmosphere)
- Most existing observation networks focus on specific compartments and/or scientific questions
- TERENO:
  - Long-term observations (> 15 years) of hydrological and ecological parameters on different scales
  - Investigation of interaction between the different compartments
  - Bridging the gap between measurement, modelling and management
  - Currently 4 observatories, each operated by one individual Helmholtz Center
  - Project duration: 2008 until >2023



Institute for Bio- and Geosciences - Agrosphere (IBG-3)

Multi-scale and multi-compartment monitoring concept of TERENO

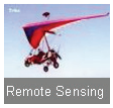
- 1065 stations
- 3 weather radar devices
- 129 lysimeters
- 400 file metadata sets




Geophysics


Groundwater monitoring


Modelling Platform


Remote Sensing


Wireless soil moisture sensor network


Lysimeters


Biodiversity monitoring

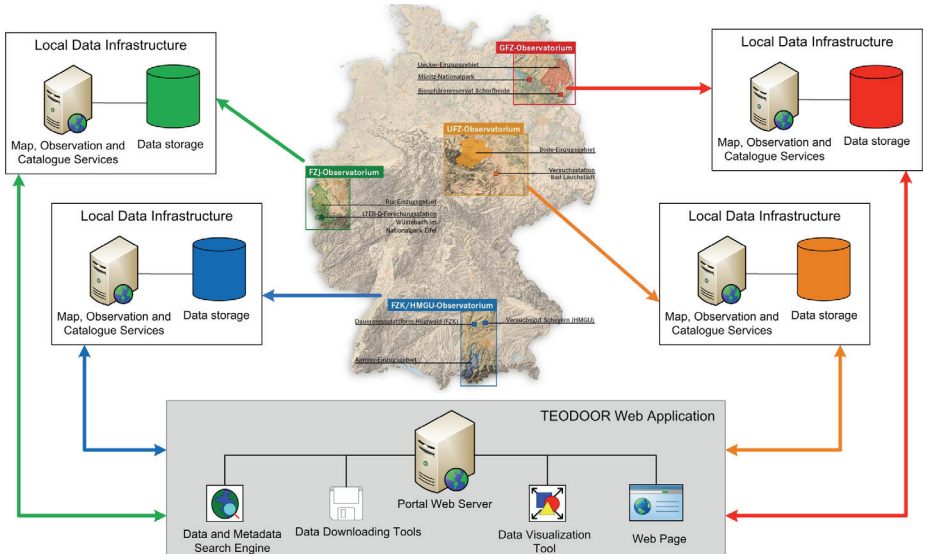

Mobile Mesocosms


Water quality monitoring


Eddy-Flux-Tower


Rainscanner

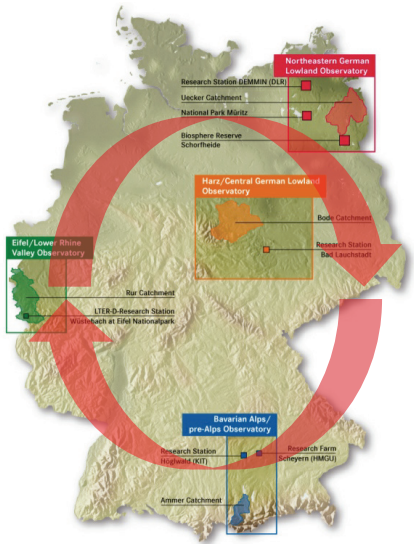
TEODOOR Distributed Spatial Data Infrastructure



Institute for Bio- and Geosciences - Agrosphere (IBG-3)

Interconnecting observatories

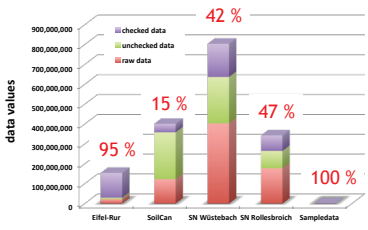
- Common data policy
  - Quality management
  - Time-limits for data delivery
  - Retention times for data publication
  - Accessibility of data
- Syntactical interoperability by consequent usage of standardized (OGC) web services and interfaces
- Semantical interoperability:
  - Common metadata profile
  - Common sensorML profiles
  - Common thesauri
  - Standardisation (parameters, units, ...)



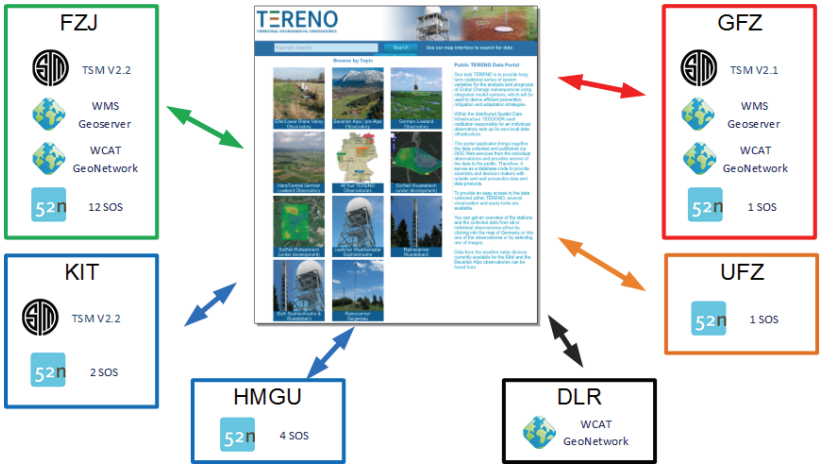
TERENO Quality management policy

Management System ISO 9001:2008
TÜVRheinland ZERTIFIZIERT
www.tuv.com ID 0010625691

- Establishing workflows for data collection, quality assessment and publication
- Nomination of responsible people
- Prohibition to circulate unevaluated data
  - **Technical inspection** (mandatory): Identification and tagging of obviously wrong data values
  - **Validity checks** (optional): Checking the continuity of time series and the definite conclusion that the observed data are representing the measured property
- Common system to assign
  - **Quality flags** (good, suspicious, bad data)
  - **Processing status** (unevaluated, quality checked,…)
- Automatic publication of quality assessed data



Data services connected to the DDP



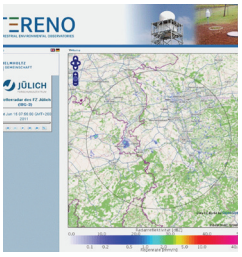
TEODOOR: The TERENO Data Portal

http.//www.tereno.net

- Central portal for information exchange, data search and data access
- Querying the TERENO Metadata Catalogues
- Connected local data infrastructures from FZJ, DLR, GFZ, HMGU, KIT, UFZ
- Custom multi-condition queries
- Predefinded queries to data
- Free access to data from more than 900 sites

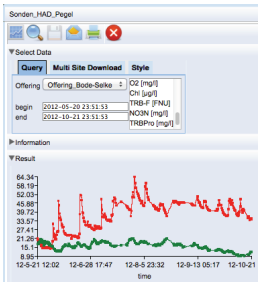
Client applications using standardized web services

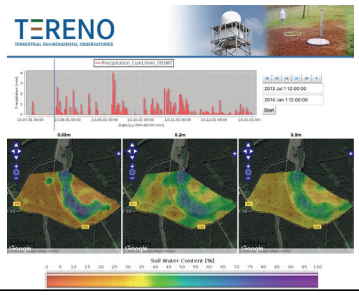
JÜLICH FORSCHUNGSZENTRUM



Animation of weather radar data using raster SOS

Visualization and download of time series data

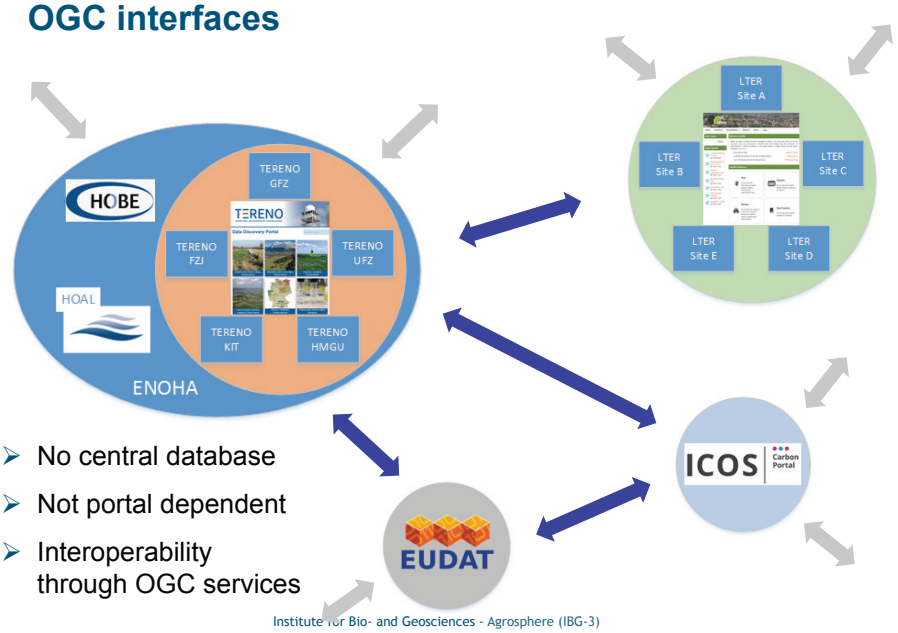




Animation of automatically interpolated soil moisture data using raster SOS

Interconnecting infrastructures using OGC interfaces

JÜLICH FORSCHUNGSZENTRUM



- No central database
- Not portal dependent
- Interoperability through OGC services

Institute for Bio- and Geosciences - Agrosphere (IBG-3)

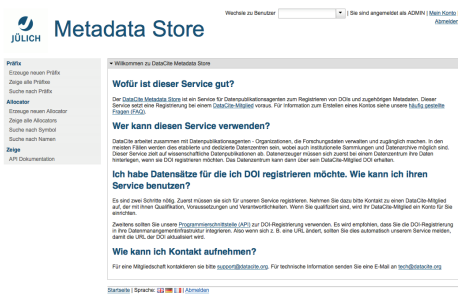
Naming issues

- Currently, naming of sites is ambiguous.
  - Different sites have identical names.
  - Sites are renamed.
  - Metadata that allow unique identification are often missing.
  - Institutions have their own naming protocols, no assurance that names are unique on a global scale.
- Access to information about observation sites
  - Need to ensure proper evaluation and facilitate interpretation of data.

DEOS - A Centralized Approach

provides a resolvable, persistent, interoperable link

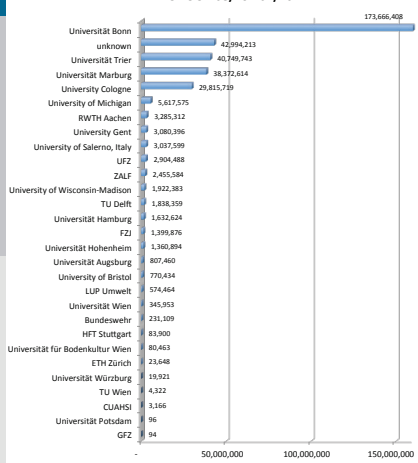
- resolvable – standard identifier syntax + network resolution mechanism (Handle System)
- persistent – through:
  - technical infrastructure (registry database, proxy support, etc.)
  - social infrastructure (obligations by Registration Agencies)
- interoperable - through a data model (semantic interoperability)

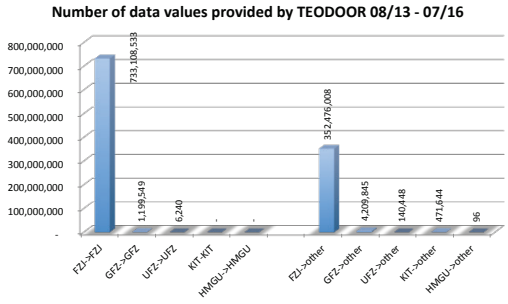


DEOS - A Centralized Approach

- Registration service currently hosted at FZJ: https://deos-id.org/deos/
- Structure: TERENO.ER012345
- Generated by DEOS or by users
- Does not replace personal or institutional names
- Building an inventory of observation facilities

TEODOOR download statistics





- Number of downloads: 3548
- Number of data series: 140,696
- Number of data values (est): 1,091,393,139
- Mean monthly dowloads: 25,000,000

Persistent data identifiers

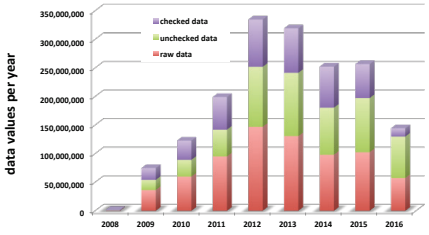
- Unique, digital identifier, allowing persistent citation of publications and data
  - Eases access to research data
  - Increases visibility of data
- Identifier refers to “landing page” containing:
  - Metadata (for station or data set)
  - Individual data sets
  - Licensing information (e.g. data policy)
- Landing page (and data, in general) hosted by issuing institution (here: GFZ)
- Internal agreement to be able to link to data from the TERENO portal
- Currently, 24 data sets/stations were identified through persistent identifiers (see https://search.datacite.org/works?query=TERENO)
- Drawbacks
  - File based approach
  - “Snapshot creation” of data from data bases required
  - Dynamical referencing to data in planning
  - Manual process
  - Hosting the same metadata on two systems in parallel (GFZ, TEODOOR)

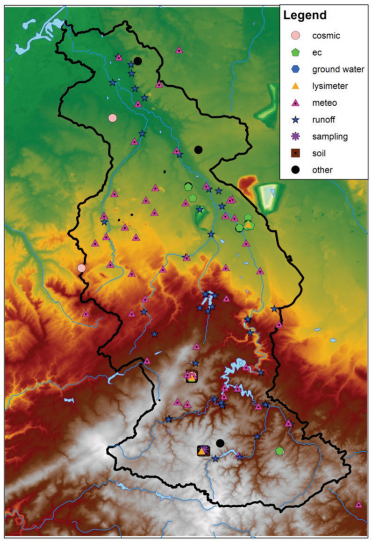
Examples

- Graswang
  - TERENO: „Graswang“ (http://tereno.imk-ifu.kit.edu/Graswang/)
  - ICOS: „DE-GWG“ (https://meta.icos-cp.eu/edit/stationentry/)
  - Fluxnet: „DW-GWG“ (https://fluxnet.ornl.gov/site/4147)
- Bad Lauchstädt
  - TERENO: „Lysimeterstation Bad Lauchstädt“ (multiple entries, e.g. http://teodoor.icg.kfa-juelich.de/ibg3searchportal2/dispatch?searchparams=freetext-lauch&metadata.detail.view.id=urn:ogc:object:feature:Sensor:UFZ:970)
  - LTER: „TERENO - Bad Lauchstaedt“ https://data.lter-europe.net/deims/site/lter_eu_de_019
- FZJ Climate Tower
  - TERENO: RU_K_001 (http://teodoor.icg.kfa-juelich.de/ibg3searchportal2/dispatch?searchparams=freetext-RU_K&metadata.detail.view.id=RU_K_001)
  - ICOS: JUE (https://meta.icos-cp.eu/ontologies/stationentry/AS/N2)

Local Infrastructure „Eifel-Rur“

- Static data (usually file based)
  - Descriptive data (reports)
  - Geodata
  - Other static data (statistics, …)
- Time series data
  - Runoff, water quality, soil, climate
    - *589 stations 10'-60', offline)*
  - Eddy-Covariance
    - *7 stations (20 Hz-10')*
  - Weather radar
    - *2 radar devices (5-10')*
  - Lysimeters (SoilCan)
    - *36 lysimeters (1'-15')*
  - Regular sampling campaign data





Time series Management System

JÜLICH FORSCHUNGSZENTRUM

Developed based on open source software and open standards.