Management and publishing of TERENO data from distributed data bases

Tereno Coordination Team Data Management

TERENO Advisory Board Meeting

26. September 2011, Blankenburg









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







TERENO distributed data infrastructure design







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')







Automated import and processing tool for site data







Data model and data publishing

- Comprehensive data model based on the CUAHSI Observation Data Model (<u>http://his.cuahsi.org/odmdatabases.html</u>):
 - 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







N

Data model used at FZJ-IBG

		역) inputid8 (INTA Nullable FK) 위] inputid9 (INTA Nullable FK) 위] outputid (INTA Nullable FK)			
G odm.datatableclasses		G odm.fileimportlog *		G odm.groupdescriptions	
1 id [INT4 PK]	Godm.datatypes 8	<pre>%] id [INT4 PK]</pre>	🖓 odm.generalcategories 🛸	1 id [INT4 PK]	G odm.labmethods *
code [VARCHAR(50)]	1] id [INT4 PK]	B archivetimestamp [TIMESTAMP WITH TIME ZONE Nullable]	M id [INT4 PK]	description [VARCHAR(255) Nullable]	+ \$] id [INT4 PK]
G definition [TEXT Nullable]	Code [VARCHAR(S0)]	[] filedatarows [INT4 Nullable]	(2) code [VARCHAR(50)]		description [TEXT Nullable]
	Gefinition [TEXT Nullable]	filename (VARCHAR(255) Nullable)	& definition [TEXT Nullable]		link [VARCHAR(255) Nullable]
		I. firstdatasettimestamp [TIMESTAMP WITH TIME ZONE Nullable]			methodname [VARCHAR(255) Nullable]
		L imported datasets [INT4 Nullable]			aname [VARCHAR(255) Nullable]
		Instruct an ettimestame [TIMESTAMP WITH TIME TONE Nullable]			organisation [VARCHAR(255) Nullable]
		B razioaraserrimesramb (miles num, with time some indiante)			
adm.logger *		adm.loggersites *		odm.metadata 2	
21 id INT4 PK	Godm.loggerfiletypes *	21 longerid [INT4 FK]	R odm.loggervariables *	1 21 id [INT4 PK]	G odm.offerings 🕆
B code (VARCHAR(SO))	11 id [INT4 PK]	21 siteid (INTA EK)	fid (INT4 PK)	& abstracttext (TEXT Nullable)	M id [INT4 PK]
Contentenailaddresses D/ADCHAD(255) Nullable]	Classcode [VARCHAR(50)]	all second further of	allowedmaxvalue [FLOAT4 Nullable]	5. liek [vapcuap/266] hullable]	Code [VARCHAR(50)]
6 contententing/of esses [VARCHUR(255) Honovej	6 description [TEXT Nullable]		B allowedminvalue [FLOAT4 Nullable]	B. configuration (VADCHAD(SEE) Multiple)	6 definition [TEXT Nullable]
L description [VEVT hullphin]			Limportfactor [FLOAT4 Nullable]	5 title [vancuan/255]]	
6 description [TEXT Hotable]			Linnerfilecolumphame (VARCHAR(50) Nullable)	Citie (VARCHOR(233))	
In romoace [TIMESTAMP WITH TIME ZONE]			B. loggerfilecolumnoumber (INT4 Nullable)	all cobiccategoryid [INTAPK]	
Lechnicalemailaddresses [VARCHAR(255) Nullable]			affretualue [FLOATA tudiable]		
b technicalwarningdays [INT4 Nullable]			Gorsecvalue (FLOATA Nullable)		
Lodate [TIMESTAMP WITH TIME ZONE Nullable]			rij agencio (INT4 Nullable FK)		
T] datatableclassid [INT4 FK]			tij categoryja (in ta nuliable FK)		
위] filetypeid [INT4 FK]			Tij loggerid [INT4 FK]		
1 sourceid [INT4 FK]			Tij offsettypeid [INT4 Nullable FK]		
们 timezoneid [INT4 FK]			一 罰 qualitycontrollevelid [INT4 FK]		
			[2] sensorid [INT4 Nullable FK]		
			1 variabletypeid [INT4 Nullable FK]		
			[2] variableid [INT4 FK]		
E odm offrattunar 8		Florte qualifierr 8		Foodm samplemedia 8	
et la (INITA DK)	Codm.projects 8	et La (INITA DV)	G odm.gualitycontrollevels 8	- 90 Ld (INTA DK)	G odm.samples 8
6. definition [TEXT Mullable]	11 id [INT4 PK]	Code (vapruap/cn))	1 id [INT4 PK]	Code (VADCHAD(SO))	fi id fint4 PK]
90 unitid [INTA EK]	Code [VARCHAR(255)]	B. definition [TEXT Mullable]	<pre>code [VARCHAR(50)]</pre>	B definition [TEXT Nullable]	description [TEXT Nullable]
All autor (usual of	& decription [TEXT Nullable]	a octimetor (rest residue)	& definition [TEXT Nullable]	a delinicon (rext itoliaole)	Labsamplecode [VARCHAR(255)]
	Firstname [VARCHAR(255) Nullable]				20 Jabmethodid [INT4 FK]
	B projectend [TIMESTAMP WITH TIME ZONE Nullable]				20 sampletypeid [INT4 FK]
	CONTRACT DIMESTAND WITH TIME ZONE Nullable				all services have for a service of
	5. surname [VARCHAR(255) Nullable]				
	a server a function of each reasonal				
G odm.sampletvpes *		G odm.sensorvariables		G odm.sitesprojects *	
11 id (INT4 PK)	Rodm.sensors R	C. description [TEXT Nullable]	B odm.sensorvariablesagents 8	11 siteid [INT4 FK]	G odm.sitestestsites
code [VARCHAR(50) Nullable]	\$[] id [INT4 PK]	sensorid [INT4 Nullable PK FK] 4	(Sposorid [INT4 FK]	21 projectid [INT4 FK]	[9] siteid [INT4 FK]
B definition [TEXT Nullable]	code [VARCHAR(50)]	variabletypeid [INT4 Nullable PK FK]	1 variabletypeid [INT4 FK]	all projection further ind	I testsiteid [INT4 FK]
deminion [rectinoned]	L currentdraw [FLOAT4 Nullable]	to note type to [in the notion of the real	양] agentid [INT4 FK]		
	description [TEXT Nullable]				
	Interfaces [VARCHAR(255) Nullable]				
	E link [VARCHAR(255) Nullable]				
	5. manufacturer (VADCHAD(255) Nullable)				
	a manufacturer (WARCHAR(255) Hullable)				
	maximpervoltage (FLOM HINDIADIE)				
	mininputvoltage [FLOA14 Nullable]				
	hodel [VARCHAR(255) Nullable]				
	powersupplytype [VARCHAR(255) Nullable]				
	type [VARCHAR(255) Nullable]				
	version [VARCHAR(255) Nullable]				
	[9] voltagetypeid [INT4 Nullable FK]				
E ete comu					
A d (with pr)	andm spatial references \$	en ogm.testsites	Codm timezones 8	M Ld (wrape)	admunits t
A address hundrarda	1 Id INTA PK	O and a human polyceth	01 Id INTACK		01 id finta pkl
address (VANCHAR(255))	B. isoeographic (BOOL Nullable)	CODE [VARCHAR(255)]	a code (VADCHAD(SO))	CODE [VARCHAR(SO)]	B. abbreviation (VADCHAD(255) Mullable)
administrativeArea [VARCHAR(255) Nullable]	 Bycographic (boot, nonable) Bycographic (boot, nonable) 	Country [VARCHAR(255) Nullable]	B. definition [TEXT todiable]	Gerinition [TEXT Nullable]	 Bobreviation [VARCHAR(253) Nullable] Code [VARCHAR(201]
b citation [VARCHAR(255) Nullable]	 name (VVRCPVR(255) Nutable) name (VVRCPVR(255) Nutable) 	b decription [TEXT Nullable]	a demición [TEXT Hutable]		CODE [VARCHOR(SO)]
city [VARCHAR(255)]	e noces [TEXT NUILable]	observatory [VARCHAR(255) Nullable]			cype [VARCHAR(255) NUIIable]
Country [VARCHAR(255) Nullable]	👦 srsid [INT4]	state [VARCHAR(255) Nullable]			
b description [TEXT Nullable]					
email [VARCHAR(255)]					
[firstname [VARCHAR(255) Nullable]					
Ink [VARCHAR(255) Nullable]					
organization (VARCHAR(255))					
L phone [VARCHAR(255)]					





Managing and publishing weather radar data







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)







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







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, Rurcatchment
 - 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 ...





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 Webpages partly realized
- Interfaces between local data bases and TEODOOR data portal via SOS in progress





TEODOOR: The TERENO Data Portal

Hierarchische Such

http://www.tereno.net

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







Portal backend: Interfaces and functions







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	Multi- ISO 19115 core		
			Resource title	1	Dataset title	М	
			Resource abstract	1	Abstract describing the dataset		
	Identification		Resource type	1			
	Identification		Resource locator	0*	Online resource	0	
			Jnique resource ocator	1*			
			Resource language	0*	Dataset language	М	
	Classification		Fopic category	1*	Dataset topic category		
			Keyword value	1*			
Keywords			Driginating controlled ocabulary	01			
	Geographic location		Geographic bounding	1*	Geographic location of the dataset		
			Femporal extent	0*	Additional extent information for the dataset	0	
	Temporal		Date of publication				
	reference		Date of last revision 1*		Dataset reference date	М	
			Date of creation				
			Lineage	1	Lineage	0	
Resolution and validity		and	Spatial resolution	0*	Spatial resolution of the dataset	0	
	Access		Conditions applying to access and use	1*			
	constraints		imitations on public	1*			
	Responsible		Responsible party	4 *	Detect responsible set	0	
	party		Responsible party role			Ŭ	
	Metadata		Metadata point of	1* Metadata point of contac		М	
	metadata		Metadata date	1	Metadata date stamp	М	
			Metadata language	1	Metadata language	C	





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 (<u>http://www.eionet.europa.eu/gem</u> <u>et</u>)
 - Additional Tereno Thesaurus







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

Schoeneseifen A80128 abstract:SOS Schoeneseifen



Data identification

TitleSchoeneseifen A80128 Abstract SOS Schoeneseifen Date

date2011-02-11T11:27:00 date type publication

Point of Contact

Name	Dr. Ralf Kunkel
OrganisationRes	earch Center Juelich GmbH
City	Juelich
Postal Code	52425
Administrative Area	North Rhine-Westphalia
Country	Germany
Phone	+49(0)2461-61-3262
EMail	r.kunkel@fz-juelich.de

Keywords

Niederschlag World

Geographic Bounding Box

6.33412981 50.50630188 50.50630188 6.33412981

Distribution Information

Online Resource

SOS IBG-3 Name

Downloadclick here to get the resource

Name Sensor Observation ServiceDescription from IBG-3 Download

Online Resource

Name Description

click here to get the Download resource

Online Resource

click here to get the resource

Metadata

Description

File Identifierec6edd99-085b-46c6-9dd7-28054c2bece1 Language eng utf8 Character set 2011-09-13T15:30:02 Date stamp Metadata Standard ISO 19115:2003/19139 Metadata Standard Version 1.0





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



delete item







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

Data visualisation in TEODOOR

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

Site Map Site Setup	INTE	ebmain ≫Internal Data Portal W	vestebach PORTAL WUE	STEBACH
Wuestebach AW 14	2			
Select Data Query Style Offering WaterQuality begin 2011-06-14.17 end 2011-06-15.17	water temperature electrical conductiv c03:00 pH value (noUnit) nitrate concentratio horide concentrationed	[degC] ▲ ity [mircoScm-1]		
► Information			1	
▼Result			is and the second	
Time 2011-06-07 22:36:00 2011-06-07 22:24:00 2011-06-07 22:24:00 2011-06-07 22:24:00 2011-06-07 22:24:00	Phenomenon WaterTemperature NitrateConcentration ChlorideConcentration OxygenSaturation PH	Data 11.88 °C 7.684 mg NO3/L 104.8 mg C/L 93 percent 6.48 -		
TEODOOR Online Data Po	ortal	Mana Marking	Ne your	State All
TERENO Presentations	(D) (ETCAS)	Contractor Rentance	Biddleton and all all a	
Meetings	的 的表示。	ALCONTRACTOR	的 的 化 化 化 化 化 化 化	
Workshops		的现在是专家	and the second	STAN STATES
Projects				
Downloads	Geogle			
Links	Grafiken ©	2011 - Nutzungsbedingunge	LANAL BREAK	11. 10/000 10/065, 6534003,56213





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)







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







- 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









- 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









- 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









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)







Roadmap for Tereno Data management infrastructure implementation

For each observatory

- List of stations to be published
- Catalogue Service set up
- Sensor Observation Services set up

end of 10/11 FZJ+UFZ done, others until 01/12 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
- Thesaurii

> TEODOOR

- INSPIRE support
- additional services and tools
- Others
 - DOI referenced datasets

INSPIRE standard GEMET + own, in development, Version 1.0 until 01/12

End of October 2011 under discussion, if required

