Scientific Data Management with Open Source Tools

An Urban Drainage Example

David Camhy, Valentin Gamerith, David Steffelbauer, Dirk Muschalla and Günter Gruber

Outline

- Initial situation
- Goals
- Technologies/Standards
- The OpenSDM approach

Initial Situation

- Proprietary DMS – Linux Server / Web GUI
- Interuniversitary Austrian project IMW2 (Novell Measurement Technologies in Water Management) – using s-can spectrolysers
- 400 million datasets since 2002, ca. 50 GB in Oracle Database
- Performance problems (Export, Visualization,...)
- Only one type of measurement station supported
- No money – no support – no further development

Goals

- Open Source, Don’t reinvent the wheel
- Use standards -> International collaboration
- Performance!
- Adequate technologies for the future (Distributed/Parallel computing, the “Cloud”)
- Store metadata and connect it to the actual values
- Make it scalable enough to allow thousands of sensors and billions of datapoints

Technologies and Standards

Open Geospatial Consortium (OGC)

- 380 members (Google, Microsoft, NASA, ESA, ...)
- Geospatial and location standards
- Examples: GML (Geography Markup Language), WMS (Web Map Service), WCS (Web Coverage Service), KML (Keyhole Markup Language), ...
- **NetCDF** (Network Common Data Form)
- Working Group: **Sensor Web Enablement (SWE)**

Sensor Web Enablement – Encodings and Services

- Observations and Measurements (O&M)
  Encoding of Measurement Data
- Sensor Model Language (SensorML)
  Description of sensor systems and processes
- Sensor Observation Service (SOS)
  Web Service to manage deployed sensors and retrieve observation/sensor data.
- Sensor Planning Service (SPS)
  Tasking Webservice for sensors and simulations
- TransducerML (TML), Sensor Alert Service (SAS),Web Notification Service (WNS), …
- EU Projects: SANY, **SUDPLAN**

The OpenSDM approach

Overview

- **Data Store** (netCDF)
- **Distributed Task Queue System**

Work in progress:

- Semantic Metadata Store / SWE compliant services
- “End user interface” (Web GUI)

netCDF – Network Common Data Form

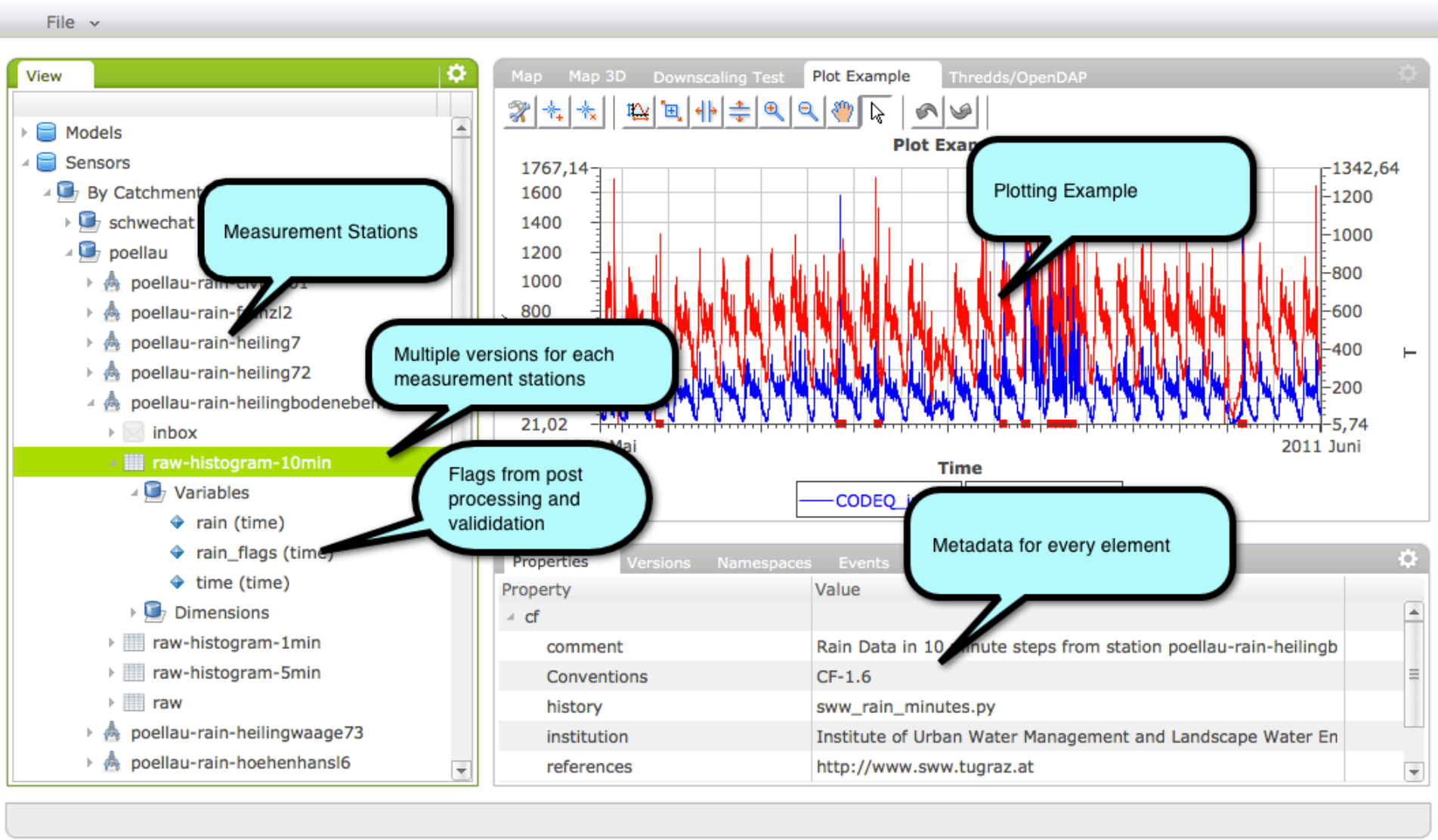
- Used a lot in „High Performance Computing“
- Self-explanatory for scientists (dimensions, variables,…)
- Many server solutions available for distributed access - OpenDAP/THREDDS, gridFTP, ERDDAP, OOSTHETYS (SOS)
- Good array-performance!
- Can be accessed in nearly every programming language
- Metadata vocabularies available (Climate and Forecast conventions)! – geo-reference, units, flagging, statistics,…
- File based: easy versioning, all metadata available directly in files
- **OGC standard** since 2011
- **Work in progress**: provide a **SOS** for data access (already working, metadata missing), indexing and queries (fastBIT indexing?)

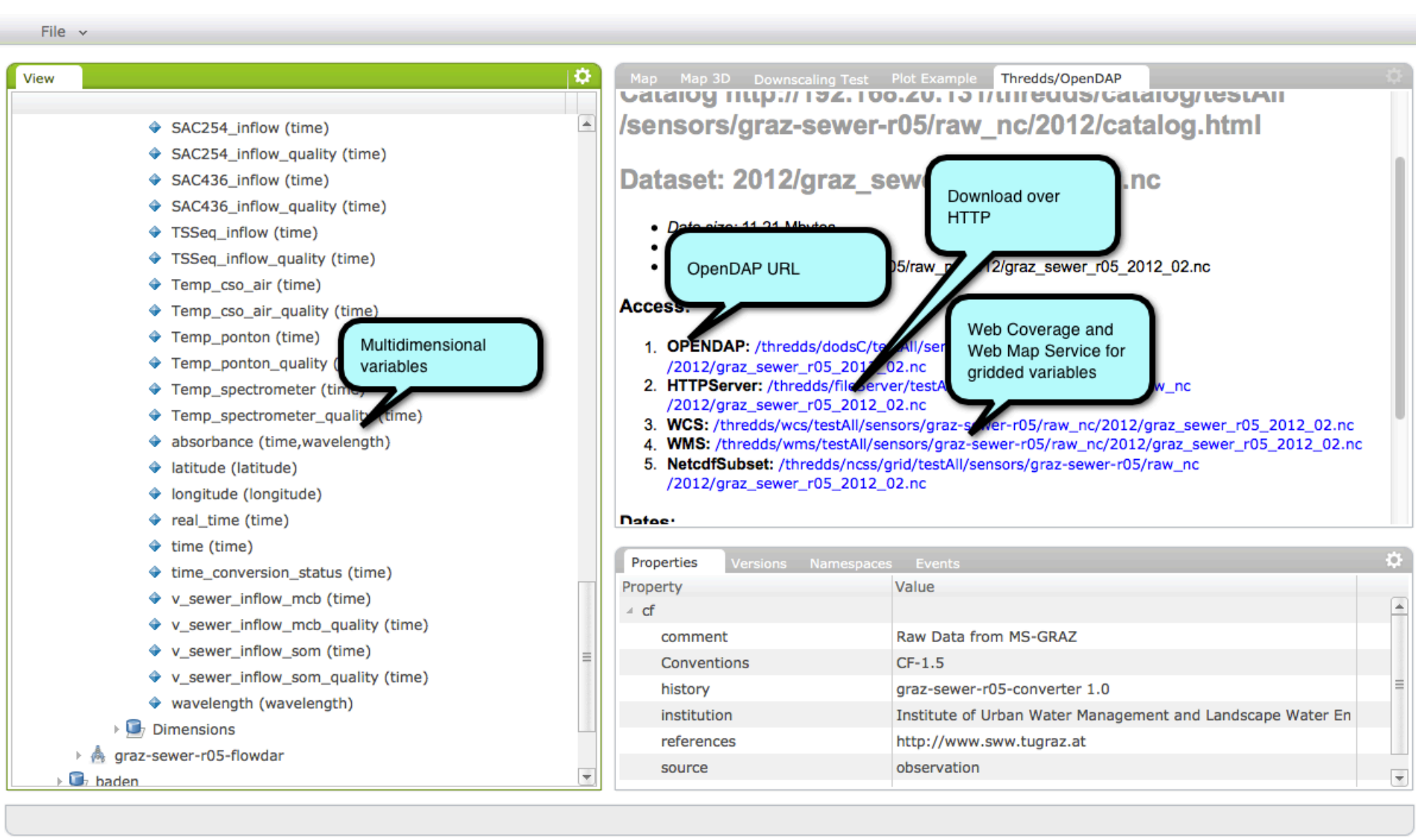
Distributed Task Queue System

- Allows scheduling/distributed execution of arbitrary task
- Allows dependent task execution
  e.g. transfer from measurement station ->
  postprocessing -> validation -> simulation
- Based on http://celeryproject.org
- Uses self-developed REST-based webservices for task monitoring and execution
- Provides a simple administration GUI
- **Work in progress**: Use a **SPS** instead of REST based webservices

Work in Progress: End User Interface

- Based on Eclipse RAP (Rich Ajax Platform)
- Server based (Java)
- Client: Web Browser (must support WebGL for 3D display)
- iPhone/Android Clients possible through same GUI protocol

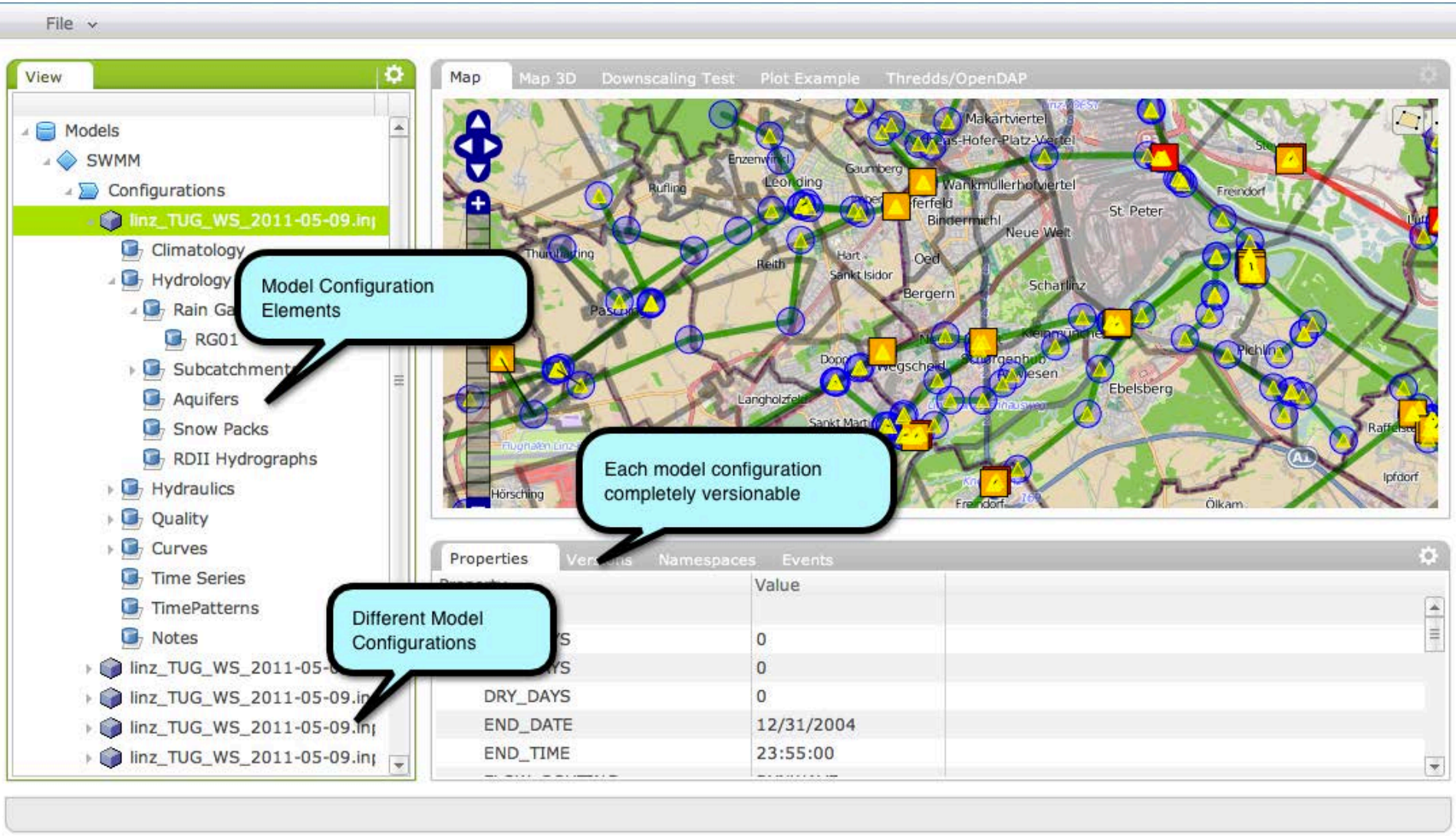

File
View
Models
Sensors
By Catchment
schwechat
poellau
poellau-rain-heiling7
poellau-rain-heiling72
poellau-rain-heilingbodeneben
inbox
raw-histogram-10min
Variables
rain (time)
rain_flags (time)
time (time)
Dimensions
raw-histogram-1min
raw-histogram-5min
raw
poellau-rain-heilingwaage73
poellau-rain-hoehenhansl6
Measurement Stations
Multiple versions for each measurement stations
Flags from post processing and valididation
Map
Map 3D
Downscaling Test
Plot Example
Thredds/OpenDAP
Plotting Example
1767,14
1600
1400
1200
1000
800
21,02
1342,64
1200
1000
800
600
400
200
5,74
T
Mai
2011 Juni
Time
CODEQ_
Properties
Versions
Namespaces
Events
Property
Value
cf
comment
Rain Data in 10 minute steps from station poellau-rain-heilingb
Conventions
CF-1.6
history
sww_rain_minutes.py
institution
Institute of Urban Water Management and Landscape Water En
references
http://www.sww.tugraz.at
Metadata for every element

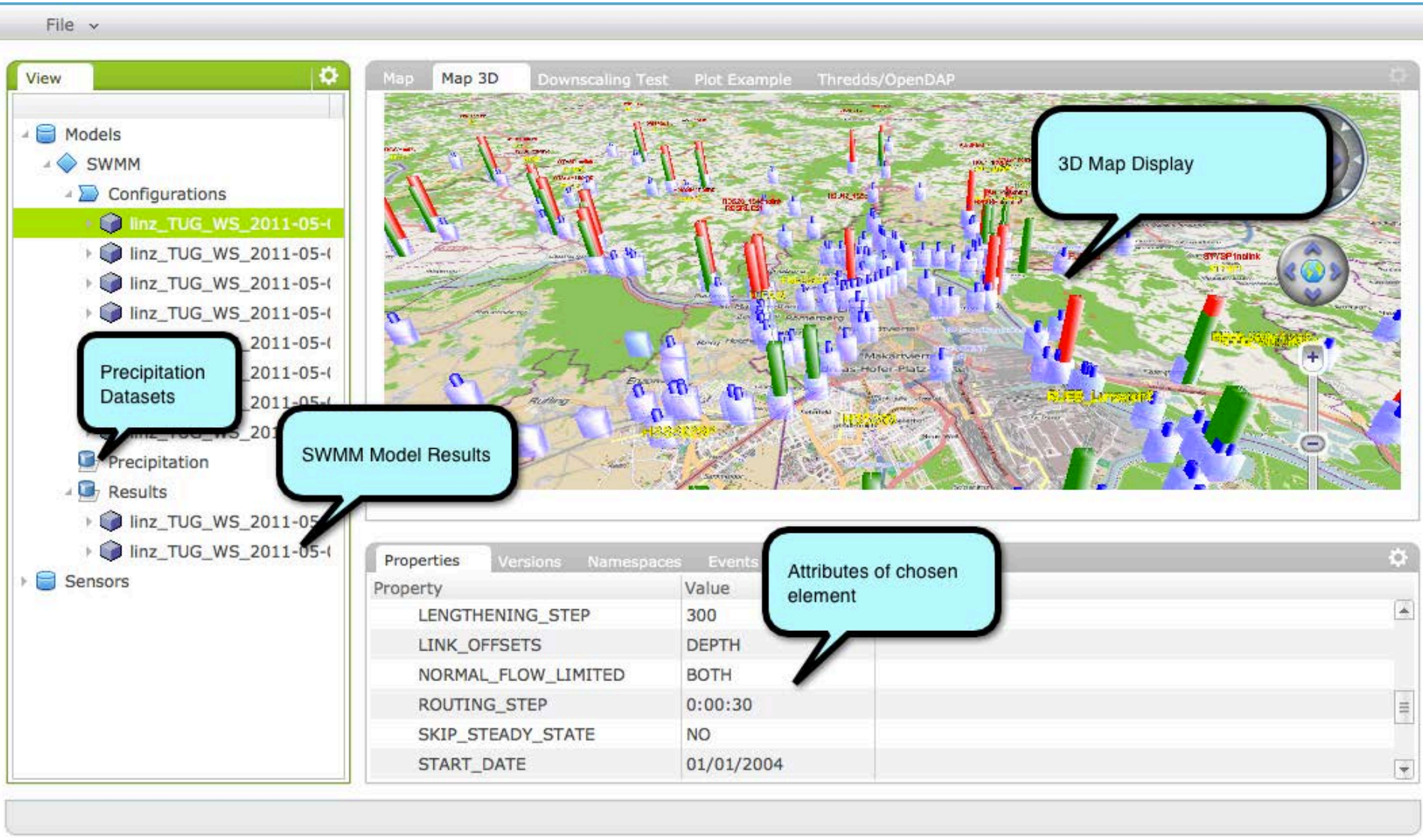

File
View
SAC254_inflow (time)
SAC254_inflow_quality (time)
SAC436_inflow (time)
SAC436_inflow_quality (time)
TSSeq_inflow (time)
TSSeq_inflow_quality (time)
Temp_cso_air (time)
Temp_cso_air_quality (time)
Temp_ponton (time)
Temp_ponton_quality (
Temp_spectrometer (
Temp_spectrometer_quality (time)
absorbance (time,wavelength)
latitude (latitude)
longitude (longitude)
real_time (time)
time (time)
time_conversion_status (time)
v_sewer_inflow_mcb (time)
v_sewer_inflow_mcb_quality (time)
v_sewer_inflow_som (time)
v_sewer_inflow_som_quality (time)
wavelength (wavelength)
Dimensions
graz-sewer-r05-flowdar
baden
Multidimensional variables
Map
Map 3D
Downscaling Test
Plot Example
Thredds/OpenDAP
/sensors/graz-sewer-r05/raw_nc/2012/catalog.html
Dataset: 2012/graz_sew
.nc
Download over HTTP
OpenDAP URL
12/graz_sewer_r05_2012_02.nc
Access:
1. OPENDAP: /thredds/dodsC/te
/2012/graz_sewer_r05_201
02.nc
2. HTTPServer: /thredds/file
/2012/graz_sewer_r05_2012_02.nc
3. WCS: /thredds/wcs/testAll/sensors/graz-s
r05/raw_nc/2012/graz_sewer_r05_2012_02.nc
4. WMS: /thredds/wms/testAll/sensors/graz-sewer-r05/raw_nc/2012/graz_sewer_r05_2012_02.nc
5. NetcdfSubset: /thredds/ncss/grid/testAll/sensors/graz-sewer-r05/raw_nc
/2012/graz_sewer_r05_2012_02.nc
Web Coverage and Web Map Service for gridded variables
Dates:
Properties
Versions
Namespaces
Events
Property
Value
cf
comment
Raw Data from MS-GRAZ
Conventions
CF-1.5
history
graz-sewer-r05-converter 1.0
institution
Institute of Urban Water Management and Landscape Water En
references
http://www.sww.tugraz.at
source
observation

Work in Progress: Semantic Metadata Store

- Directed acyclic graph structure with access control lists
- Each node in the graph can be further described by attributes
- Each node can also be a file which is stored in blocks in a distributed manner
- Ontologies can be created by the user in a GUI.
- **Model/simulation integration**: SWMM model prototype
- **RDF** representation available / “semantic web”
- Connection to “semantically prepared” markup languages like **SensorML**.
- Contains Metadata of future **SOS** and **SPS** services

Prototype: SWMM model integration




File
View
Models
SWMM
Configurations
linz_TUG_WS_2011-05-(
linz_TUG_WS_2011-05-(
linz_TUG_WS_2011-05-(
linz_TUG_WS_2011-05-(
Precipitation
Results
linz_TUG_WS_2011-05-(
linz_TUG_WS_2011-05-(
Sensors
Precipitation Datasets
SWMM Model Results
Map
Map 3D
Downscaling Test
Plot Example
Thredds/OpenDAP
3D Map Display
Properties
Versions
Namespaces
Events
Attributes of chosen element
Property
Value
LENGTHENING_STEP
300
LINK_OFFSETS
DEPTH
NORMAL_FLOW_LIMITED
BOTH
ROUTING_STEP
0:00:30
SKIP_STEADY_STATE
NO
START_DATE
01/01/2004

Thank you!