INFORMATION HANDLING IN INTERDISCIPLINARY, HYDRO-ENVIRONMENT ENGINEERING PROJECTS

Molkenthin Frank[1] & Notay Vikram[1] & Li ChiYu[1]

[1]Chair Environmental Informatics, BTU Cottbus, Germany, Konrad-Wachsmann-Allee 1, 03046 Cottbus
E-mail: frank.molkenthin@tu-cottbus.de

Abstract

Information handling in water related interdisciplinary projects are a demanding challenge due to the mass of data from field work, laboratory experiments and use of simulation models. This paper describes a general concept for information handling and management in an interdisciplinary research unit using a generalized information modeling approach for multi-scale physical state variables in combination with metadata and Web services based information systems following the INSPIRE initiative. The concept has been applied using an innovative hydroinformatics system so-called “Turtle” in a research unit handling heterogeneous mass data from extensive field measurements, laboratory experiments as well as simulation models from different disciplines such as hydrology, hydrodynamics, geo-hydraulics, geo-physics and soil mechanics.

Introduction

Projects in water related research and engineering are using mass of information originated from field measurements, laboratory experiments and numerical simulations. The efficient handling and the useful utilization of such information require suitable information systems to support efficient interdisciplinary and distributed project collaboration as well as the integration of models and data on different scales. Normally information management is not seen as critical core part of projects, effort on this task pays off more on the medium and long term perspective than on the short term level. This leads to a lower level of focus on information management during the project preparation and the project performance. Simple, conventional solutions such as file systems for ASCII data files, excel sheet files and separated GIS files are applied to survive in the flood of incoming mass data. Medium term oriented reporting and long term oriented archiving is considered by personal “solutions”. This works in smaller projects with non-changing staff but not anymore on interdisciplinary projects, when mass of heterogeneous data is collected to be used by experts from different disciplines at any time over long time periods.

As the amount of available data by modern sensor, remote sensing and Web technology is increasing rapidly, there is a demand on suitable information management systems to deal with this problem on short, medium and long term tasks. As a vicious circle, no information management system is available when data is available, or if a sophisticated information management system is set-up, nobody wants to use it (see Nelson, 2009). Data base systems and GIS software can be used as basic supporting tools, but did not consider the different interdisciplinary aspects of water related projects. To overcome this problem hydroinformatics systems are requested as flexible solution considering different level of project complexity, different spatial and time scales as well as the different, discipline oriented view on the water related project topics. This paper describes an approach for a such system by combining metadata and Web service oriented information management (such as introduced by INSPIRE, 2007) and trans-disciplinary information modeling of physical state variables based on generalization and specialization. The system has been applied in an interdisciplinary hydro-environmental research project.

Research Unit “Grosshang”

The interdisciplinary research unit “Grosshang – Natural Slopes” (see http://www.grosshang.de) deals with the “Coupling of Flow and Deformation Processes for Modeling the Movement of Natural Slopes” (Hinkelmann, Zehe 2006, 2011). The research unit is structured in five highly interdisciplinary sub-projects. The central sub-project supports the other sub-projects with a Web-based information system to manage multi-scale physical state variables and to integrate information from field and lab experiments as well as numerical simulations.

Target of the research unit is to simulate the long-term deformation of large mountain sides. Extensive field measurements are performed to identify the relevant processes and belonging scenarios for the slope Heumöser in Ebnit/Austria. The measurements cover hydro-meteorological, geo-hydrological and geo-physical state variables on different time scales and spatial distribution. The quantity and heterogeneity of measurement data is growing during such project due to ongoing permanent measurements and additional short period measurements. These mass data are used to identify the relevant natural processes and to specify suitable scenarios for further investigations. Simulation models from hydrology, hydrodynamic, multi-phase groundwater flow and soil mechanics are used and partial coupled towards interdisciplinary, multi-scale simulation components to get a better understanding of these processes and scenarios. Laboratory experiments are performed for model parameter identification and verification. Both, simulation models and laboratory experiments, are also increasing the amount of heterogeneous mass data to be handled in the research unit.

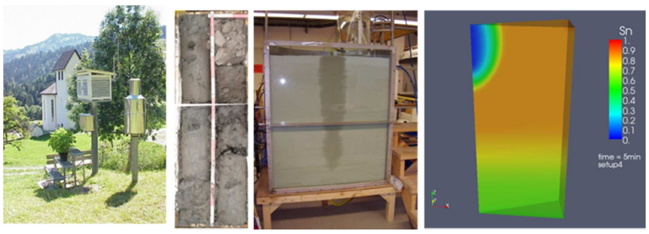

Figure 1: Data from field, laboratory and simulation

The different simulation models, field and laboratory measurements deal with physical state variables on different scales in time and space. The integration and coupling of the different models requires the exchange or sharing of multi-scale physical state variables during runtime or in a sequence of partial processes. The heterogeneity, complexity and interdisciplinarity of this research unit is a challenge for information modeling, management and engineering in hydroinformatics. The demands on information management can be structured by time perspective in three types of support services:

- Information Acquisition and Handling
  This is a short term task during and after field/lab measurement for data pre-processing to transform raw data to usable information.
- Information Analysis and Management
  This is a medium term task following the information acquisition running within the time window of the research unit. This task supports the core activities of the project partners and their collaboration based on information sharing.
- Information Archiving and Storage
  This is a long term task for the time after the end of the research unit. It ensures the persistent availability of relevant information and allows an efficient information retrieval for future projects.

Information systems have to consider these three tasks with their different requirements and conditions.

Information Modeling Concept

Key idea of information management within the research unit is the generalized information modeling of all relevant physical state variables using the basics from mathematics (set theory, tensor algebra) and computer science (object-oriented modeling) in combination with metadata and Web services technology. All physical state variables are modeled by tensor classes and managed by sets. This modeling concept is called “TensorML” following other XML based modeling concepts. The tensor classes are structured by tensor rank, tensor dimension and tensor structure (Molkenthin 2006, 2009). Examples for the tensor rank are scalar, vector and matrix. Physical state variables are depending on different numbers of coordinates (such as space and time), defining the tensor dimension. For any tensor dimension related discretization structures are defined, such as regular and irregular topological structures based on points as known from grid modeling in numerical schemes. The physical state variable values and coordinates (space and time) are linked to these points within the topological structures. The tensor structure considers combination of tensors on the tensor object level or on the tensor value level. These set combinations includes functional relationships between different tensor objects such as material laws, Q/h relations as well more complex dependencies described by AI or numerical models. Set operations on tensor sets enable the management of several physical state variables independent form their origin, scale and relationship.
The classifications by tensor rank, tensor dimension and tensor structure lead to several tensor classes covering all typical kinds of data for measurements and simulation models. Based on the object-oriented approach the described tensor classes contain - besides the data for the physical state variables values - methods for management, manipulation and analysis such as tensor operations, tensor analysis and set operations. Tensor objects also contain metadata describing all relevant information about the origin (sensor, model), spatial and temporal coverage, contact person and so on. The modeling concept is described by UML and implemented by Java classes. Tensor objects and sets are described by a specific XML-application/notation called TensorML.

Information Handling Concept

The described approach has been used to develop a flexible information system with the code name Turtle. The system consists of tools for user interfaces (editor, analysis and visualization), report generator (interactive HTML report), information archiving, import/export as well as metadata management. It is linked with GIS systems such as WebGIS services for displaying maps within a browser.

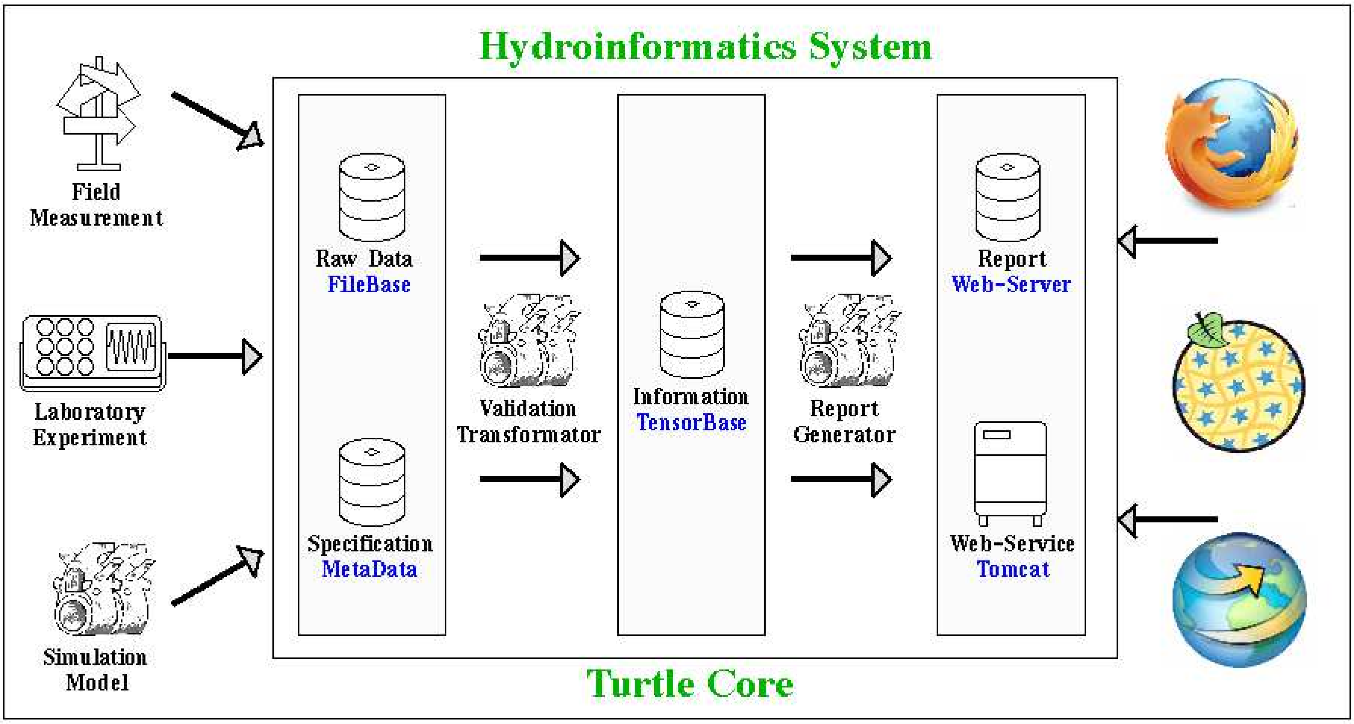

Figure 2: Basic concept of the data management in the research unit

Turtle integrates the traditional separated pre- and post-processing, databases and simulation processors in a holistic but flexible hydroinformatics system. The basic concept of information handling in the system is shown in Figure 2. The raw data from field, lab and simulation are the input data as they are generated by the sensor device interfaces in the field and lab as well as simulation models. They are collected, validated and pre-analysed in the raw data file base. Related metadata specifications allow a complete description towards long term persistent and accessible information. In a validation process the raw data information are transferred in a generalized tensor base including several filter, interpolation, standardization and transformation methods to bring all information on a standardized level independent form their origin in the raw data file base. As tensor objects are semantic independent units including data and methods the tensor base can be used for any kind of information analysis and management in the research unit. Interactive Web reports are generated to access the tensor base in a distributed environment via the Internet and to ensure a long term oriented archiving and storage of the information. Besides standard Web pages Web/GIS services and metadata services are supporting this with modern functionality and Web interfaces to other systems such as spatial data infrastructures (SDI) and GIS systems. Some details of the introduced components of the information system Turtle are described in the next chapters for the example of field measurements. Same principles are used for laboratory experimental data and results from simulation models but cannot be described in details in this short paper.

Field Measurements: Raw Data Structure

Turtle has been applied for the field measurements at the slope Heumöser nearby Ebnit/Austria. The related raw data are stored in data files, using mainly proprietary ASCII formats depending on the sensor/station configuration and technical equipment. Several permanent and temporal field measurements are available, examples are:

- meteorological stations operated by city of Dornbirn at 5 locations
- hydrological/meteorological station with 21 sensors operated by Karlsruhe Institute of Technology (KIT)
- creek measurements (weirs) by KIT
- tracer measurements by KIT
- 3 borehole measurements by KIT
- nanoseismic network by Univ. Stuttgart
- geoelectric/seismic investigations by UFZ Leipzig and Univ. Potsdam

The available data are handled in three steps:

- classification by metadata including filter and validation rules
- integration in the file base as original raw data
- transformation in the tensor base of Turtle with flexible time scale using validation and transformation rules

Figure 3: Field measurement

File Base: Information Acquisition and Handling

The file base consists of a huge amount of raw data files with different data formats from the different sources. The file base is analyzed for format validation and measurement metadata such as time window, sampling rate, type and unit of the measured physical state variable and measurement gaps. The raw data itself is validated by rules. Examples are value ranges and filter methods. Value ranges define valid ranges of data, such as physical rules (e.g. air humidity: 0.0% to 100.0%) or sensor based limitations and specifications of the used sensors. Filter methods can be used to define specific operations for specific time windows such as NaN values for out of operation, or data offset for wind direction instrument shift. Time shift operations are used to transform all data from local measured time zones (CET, CEST, UTC) in one standard time coordinate system (UTC).

Metadata: Model, Station, Sensor, Tensor

The performed field measurements are described in a suitable information model and related metadata scheme based on ISO 19115. All measurements from one source (data service) are combined in one file model structured by stations. Each station consists of several sensors, the sensor itself is linked to one or more measured physical state variables modeled as tensor objects. File, model, station, sensor and tensor do have metadata properties to describe all relevant properties, such as location, operation time windows, value range and tolerance, sampling rate, units, serial number. Time and location (spatial coordinates) are using as coordinates for the other attributes. This allows the description of time depending calibration parameters, time windows of non-operations and maintenance as well as (e.g. seasonal) changing locations of sensors.

Tensor Base: Information Analysis and Management

The raw data are transformed into the tensor base using standard properties e.g. for time, unit and location to get a standardized set of physical state variables described my tensor objects (“TensorML”). Data from different sensors with the same physical state variable are combined in one time series. Gaps are filled by interpolation, where suitable and possible. The tensor base consists of tensors for a predefined project time window (1998-2011) and time scales for all physical state variables. Basic regular time step is 600 sec with the time origin of 01.01.1990 00:00:00 UTC. Further time scales can be specified, typically time step resolution of 3600 sec (1h), 21600 sec (6h), 86400 sec (1d) and 604800 sec (7d) are used. The transformation between the time scales is part of the tensor base using typically mean value or sum-up methods. Separated measured state variables can be composed by tensor operations such as wind direction and speed towards wind vector or soil moisture at different depth towards a 2D (time, depth) scalar field.

The transformation of raw data in the tensor base considers the measured state variables. Further interesting physical state variables can be included as depending tensors. The relationship is again a tensor, which might be time depending. Examples are the API (Antecedent Precipitation Index) function for precipitation or the discharge tensor within creeks. For the discharge the measured physical state variable is the water level. Based on the weir geometry and instrument configuration related Q/H relationships can be specified for specific time windows.

Web Report: Information Archiving and Storage

The information system generates an interactive Web report with the whole information content of the file base and the tensor base. The use of JavaScript and JavaServlet technology to generate any Web page dynamical on demand for any tensor object, tensor set or tensor operation allows flexible navigation through the existing information scheme. The Web report consists of three main parts:

- metadata based structure report
- raw data report
- tensor base report and data (export) access

The metadata report represents the metadata specification of the raw data file base. Metadata search functionality base on GeoNetwork (http://geonetwork-opensource.org) and GeoServer (http://www.geoserver.org) as GIS frontend provides a powerful interface for information retrieval. The raw data files are completely linked and reported in tables and diagrams, so all data information can be found in the report and in the raw data files. The tensor base report is interactive to allow the reader to surf through the information base for different time scales and time windows. The Table Viewer reports all data in an alphanumeric manner. The Data Viewer is also based on tables but summaries the physical state variables by hourly, daily, weekly, monthly, annual mean, min, max values as well as gaps, variance and standard deviation. The Image Viewer is the graphical report of the tensor base with all related diagrams of the physical state variable on every time scale for every suitable time window.

The access to the data base via the Internet is realized by a JavaServlet component of the information system. User can ask for any physical state variable in any time resolution and duration as well as in any combination with other variables (e.g. rainfall data, creek discharge, soil moisture and borehole declination). Besides standard table and diagram presentation, the data can be downloaded as Excel, ASCII or XML file (data export) or be analyzed in a tensor editor JavaApplet component or via Web/GIS interfaces in GIS with full application functionality.

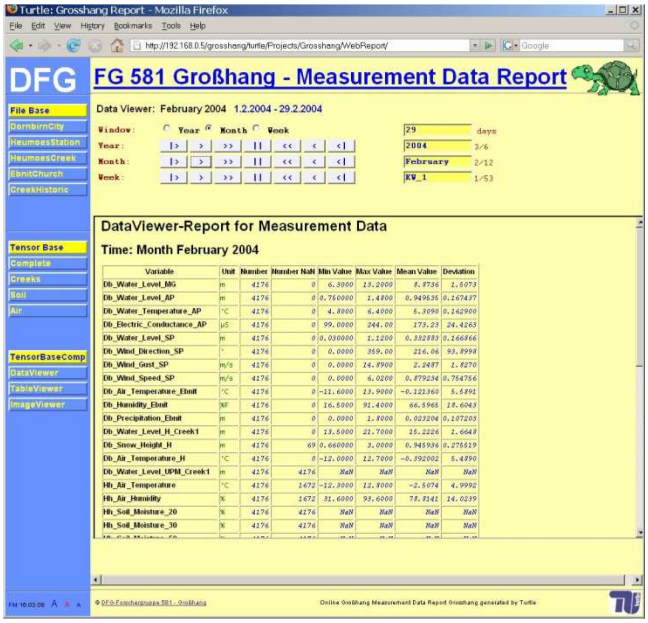

Figure 4: Data Viewer report component

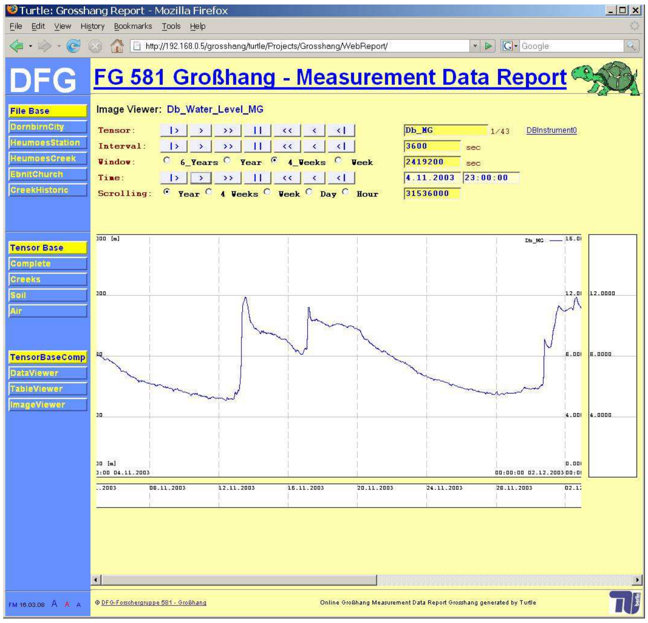

Figure 5: Image Viewer report component

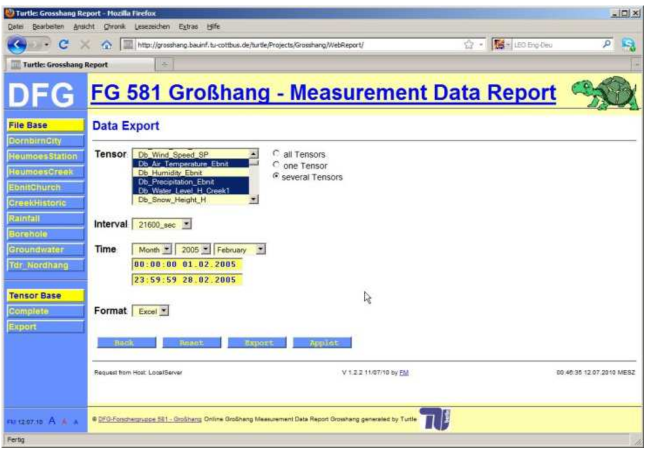

Figure 6: Data export component of the Web report

Data Archiving and Storage

Long term archiving of data is important for interdisciplinary projects to ensure data discovery and retrieval. In the research unit Großhang – "Natural Slope", such data archiving services are based international standards such as ISO 19115 (geographic information - metadata) and ISO 19119 (geographic information – services). Information retrieval is implemented according to the INSPIRE conform GDI-DE technology (German spatial data infrastructure). This holds for the raw data in the file base as well as for the content of the tensor base. The interactive Web reports with metadata search functionality enables the full access to the tensor base and to the raw data file base. In this way Turtle as information base combines all three important tasks: short term information acquisition and handling, medium term information analysis and management and long term information archiving and storage.

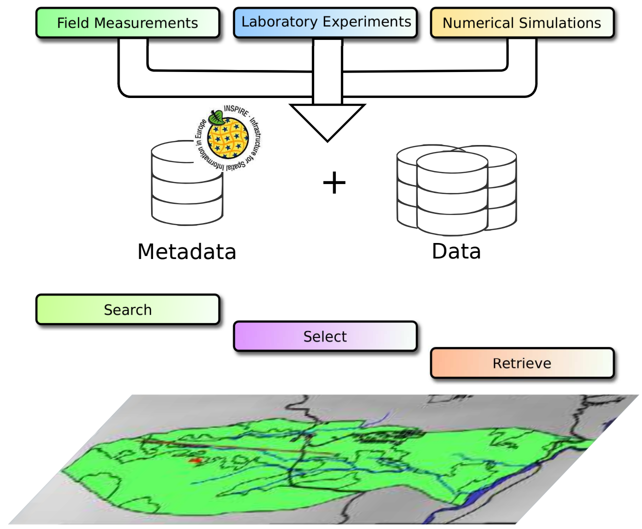

Figure 7: Metadata concept for information retrieval

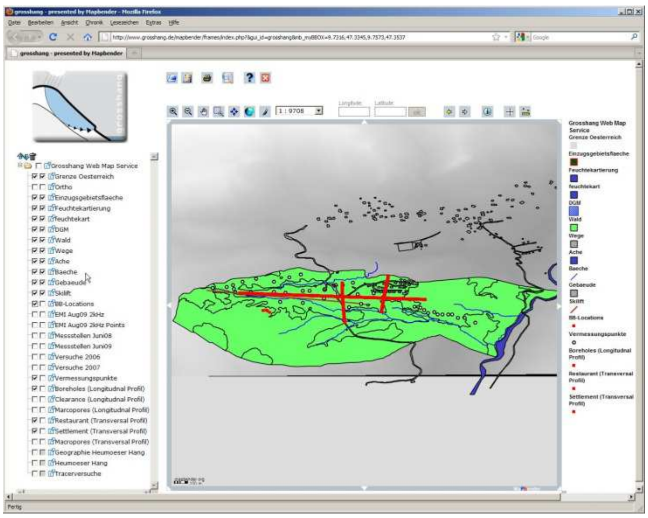

Figure 8: WebGIS interface to project information

CONCLUSIONS

Interdisciplinary water related engineering projects require a flexible handling of information from different disciplines on different time and space scales. This paper described a holistic approach to handle and manage mass information from field, laboratory and numerical simulation in one hydroinformatics system using a generalized modeling concept for physical state variables (“TensorML”) in combination with metadata and Web/GIS techniques such as used by the INSPIRE initiative of the EU. The application in the research unit “Grosshang” demonstrates the suitability of this approach and the potential for further activities in this field. Two examples for ongoing activities are introduced: The tensor based modeling concept is also used to couple different simulation model by an independent tensor exchange broker. This broker can similar to OpenMI exchange information on demand including mapping and scaling but also using tensor object relationships to represents physical behavior from material laws, Q/h relationships to complex process described by differential equation systems. Second example of future innovation is an event and process identification and scenario composition/generation component to feed simulation models based on natural observations.

Water is the base of human life. The challenge of a transparent, fair and sustainable water management on Earth in the age of the information society needs such kind of information system. Initiatives such as INSPIRE and the European Directives such as Water Framework Directive, Urban Waste Water Directive, Flood Directive, Drinking Water Directive are forcing these tasks on all levels. The importance on suitable, flexible information handling and management in water related projects for short, medium and long term perspective is increasing and demand innovative concepts and solutions.

Acknowledgments

The work is partially funded by the Deutsche Forschungsgemeinschaft (DFG) within the Research Unit 581 “Grosshang”. The presented work is based on a close collaboration within the sub-projects of the research unit, special thanks goes to all colleagues in particular to Jan Wienhöfer, Falk Lindenmaier, Erwin Zehe and Reinhard Hinkelmann.

References

Hinkelmann, Zehe (2006). *Kopplung von Strömungs- und Transportprozessen für die Modellierung von Großhangbewegungen*. Hydrologie und Wasserbewirtschaftung, Heft 1, pp. 51-54

Hinkelmann R., Zehe E., Ehlers, W. & Joswig M. (2011). *Special Section on Landslides: Setting the Scene and Outline of Contributing Studies* Vadose Zone Journal, pp. 473-476.

DFG research unit FG 581 Großhang (2006-11): *http://www.grosshang.de*

Lindenmaier (2008): *Hydrology of a large unstable hillslope at Ebnit, Vorarlberg: identifying dominating processes and structures*, Ph.D.-thesis,, Univ. Potsdam, http://opus.kobv.de/ubp/volltexte/2008/1742/

Molkenthin, Hinklemann, Lindenmaier & Zehe (2006). *Web-based information system for multi-scale physical state variables*, in Proceedings 7thHydroinformatics Conference, Nice/France, ISBN 81-903170-4-0, pp 2122-29

Molkenthin, Meienberg, Hinkelmann (2009). *Information Management of Multi-Scale Physical State Variables in an Interdisciplinary Research Unit*, DFG-Data Management Workshop, Cologne, http://www.tr32db.uni-koeln.de/workshop/workshop09.php

Bryn Nelson (2009): Empty Archive, in NATURE | Vol 461|10 September 2009

INSPIRE, Directive 2007/2/EC, http://inspire.jrc.ec.europa.eu/ (visited 31 Jan 2012)