Conceptual article

OBIS - a Data and Information Management System for the Okavango Basin

Sven Kralisch, Franziska Zander & Wolfgang-Albert Flügel

Abstract: Sustainable land and resource management and related interdisciplinary research projects like *The Future Okavango* (TFO) demand for integrated data management systems to store, describe, analyze and disseminate related information. These systems should not only take account of the possibly large range of varying data types and formats, but also have to consider different user groups and the demand for collaborative data access and data sharing. Within TFO these requirements are addressed with the development of the *Okavango Basin Information System* (OBIS). OBIS is designed as a web-based data management and analysis platform with full read/write access to all data, using open software and open standards whenever possible. This paper briefly describes the requirements that guided the OBIS development, its functions and interfaces, and gives an overview of data already stored in the system.

Keywords: environmental information; geospatial services; open-source software; resource management.

Abbreviations: TFO=The Future Okavango; OBIS=Okavango Basin Information System; OGC=Open Geospatial Consortium; RBIS=River Basin Information System; CSW=Catalogue Web Service; WMS=Web Map Service; WFS=Web Feature Service; GSOD=Global Surface Summary of the Day; NCDC=National Climatic Data Center; SOS=Sensor Observation Service; BMBF=German Ministry of Education and Research.

OBIS - Sistema de gestão de dados e informação para a bacia do Okavango

Resumo: A gestão sustentável do solo e dos recursos e os projetos de pesquisa interdisciplinar relacionados, como o The Future Okavango (TFO), demandam por sistemas de gerenciamento de dados integrados para armazenar, descrever, analisar e disseminar as informações relacionadas. Esses sistemas devem levar em conta não apenas a possível grande variedade de diferentes tipos e formatos de dados, mas também devem considerar diferentes grupos de usuários e a demanda por acesso e compartilhamento de dados colaborativos. No TFO esses requisitos são abordados com o desenvolvimento do "sistema de informação da bacia do Okavango" (OBIS). O OBIS é projetado como uma plataforma de gestão e análise de dados baseada na web, com acesso completo de leitura/escrita a todos os dados, utilizando software e padrão livres sempre que possível. Este estudo descreve rápidamente os requisitos que nortearam o desenvolvimento do OBIS, suas funções e interfaces e dá uma visão geral dos dados já armazenados no sistema..

Palavras-chave: informação ambiental; gestão de recursos; serviços geoespaciais; software de fonte aberta.

Introduction

Integrated environmental resources research in interdisciplinary projects demands for a comprehensive and holistic system analysis and process understanding (Flügel, 2011; Kralisch et al., 2012) which in turn relies on powerful and effective data management. It further depends on information sharing strategies based on best available knowledge to support researchers and decision makers. To address the demand for geospatial information sharing, a number of data format and interface standards have been developed during the last decade, mainly driven by the Open Geospatial Consortium (OGC) and other standardization organizations. Examples are the ISO 19115 standard for Geographic Information-Metadata (ISO, 2005) or the Catalogue Services Specification (OGC, 2007), both of which are widely applied in geodata infrastructures on national and international levels (INSPIRE, EU, 2007).

However, information sharing in environmental sciences also demands for the integration of geospatial data with other, non-spatial data related to specific disciplines which often feature individual data repositories. Examples are the *Data Observation Network for Earth* (Michener et al., 2011), the *CUAHSI Hydrologic Information System* (Tarboton et al., 2011) or the *PANGAEA* information system (Grobe et al., 2006). Many of these systems are constrained to specific user groups (e.g. experts), rather define a framework infrastructure instead of really

offering user and service interfaces or do not support geospatial services at all. Another problem is the often missing support for the upload and sharing of user-owned datasets. This is barely a problem of missing features in available standards, but simply a lack of support for these functions on the client's side. Regarding the data management design, many existing systems also lack a fine-grained user permission management (e.g. write-permission to owned data only). Summarizing the current situation it can be stated that – even though standards and software are readily available – integrating geodata and thematic information together with standardized services and mapping capabilities in ready-to-use client/server architectures is still challenging.

The *Okavango Basin Information*

System (OBIS, http://www.future-okavango.org/obis) addresses these deficits. Its underlying software platform is the *River Basin Information System* (RBIS, Zander et al., 2011), a client-server application built upon Open-Source software and open standards. RBIS is a modular-structured system consisting of a comprehensive and spatially enabled relational database and an application server providing web-based user interfaces to access data management and visualization functionalities. The next sections will give an overview about the requirements that guided the OBIS development, the RBIS modules used in OBIS, draft its overall system architecture and will present supported data types and available functions.

Requirements and Architecture

The development of OBIS was guided by the idea to provide a data management and exchange platform for TFO that addresses the demands of a wide range of user types, i.e. researchers, decision makers and inexperienced users. The following generic objectives were guiding its design and implementation:

- **Data types**: To address a wide range of user demands and applications, OBIS addresses the storage and management of both spatial (e.g. maps, points, polygons) and non-spatial data (e.g. time-series, documents). Flexible functions to describe, integrate and jointly access both types of data are available.
- **Open source**: To support reuse and extensibility and make use of already existing open source components,

OBIS is developed as Open Source Software.
- **Flexible and extensible**: To build a system which is scalable and easy to extend according to individual user requirements, OBIS is structured in a modular, flexible way.
- **User acceptance and trust**: To raise user acceptance and built trust, the system provides intuitive user interfaces and provides means to protect data from unauthorized access.
- **Interoperability**: To expose data on the internet and to exchange information with other applications (e.g. modelling tools), OBIS offers standard-compliant interfaces and services where possible.

In addition, TFO imposed a number of specific requirements related to its multi-disciplinary approach and the setting of regional stakeholders and potential OBIS users:

- **Accessibility**: OBIS is primarily designed as a web-based system with central data storage. To provide data access also under conditions where network availability is limited or not existing, OBIS is also available as an offline version.
- **Generalization**: To take account of the large variety of different data types and formats in TFO, OBIS provides options to store generic binary data while offering functions to easily interlink them with location and field site information.

To realize these objectives, a standard 3-tier architecture design was chosen for OBIS (Fig. 1). Here, a central application

server provides the business logic and stores/retrieves data in/from three databases containing geodata, time-series data and metadata. The interaction with clients is done via HTTP requests and XML/HTML documents over the internet, requiring only standard web-browser functionality on the client side.

The web stack used for OBIS consists of a standard Linux environment with Apache web server, PHP and PostgreSQL/PostGIS. The whole system is built based on open source software libraries, ensuring a transparent and cost-effective deployment and operation. As a platform for geodata storage and visualization, the MapServer (http://www.mapserver.org), OpenLayers (http://openlayers.org) and PostGIS (http://postgis.refractions.net) software packages were used. All data are stored within database tables whenever possible and in the file system otherwise, e.g. in the case of raw data.

Data Types

OBIS supports the management, analysis, visualization and presentation of different types of data in the context of multidisciplinary environmental assessment and planning. This includes the following data types:

- **Time-series data**: measured or simulated time-series of environmental parameters together with information about the related measurement station or simulation model
- **Map data**: spatial data layers in raster or vector format
- **Documents**: arbitrary documents in text, office or other binary formats
- **Observation data**: general information about observations together with
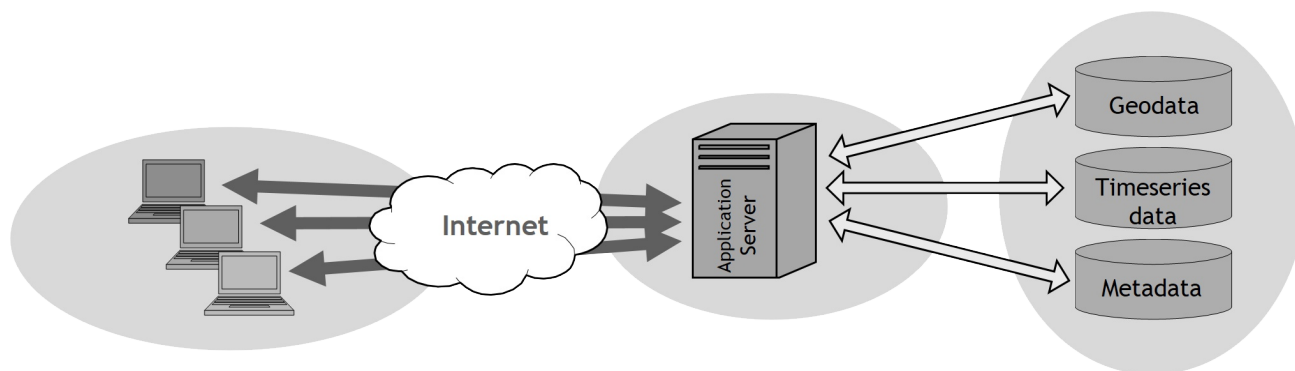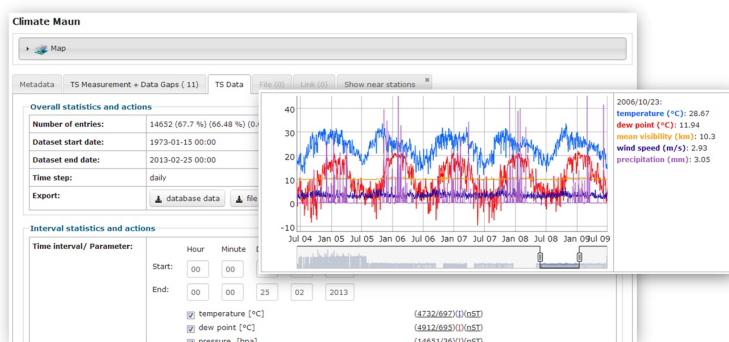

Fig. 1: OBIS common architecture.

Fig. 2: Screenshot of time-series visualization / management.

their location; detailed description of soil and vegetation samples

- **Scenario data**: description of development scenarios and linkage with related indicators and indicator values
- **Metadata**: structural and descriptive meta-information about all data stored in OBIS with description of data provenance, quality, source and others

Along with functionality to store and manage these data types, OBIS offers a variety of functions to search, analyze and access the abovementioned information. The most important are described in the following sections.

Time-series data

The OBIS time-series management module (Fig. 2) provides functions to manage measured or simulated time-series data with a special focus on environmental modelling. Time-series are stored in database tables for easy access and processing, while original data files are maintained throughout their lifetime. Main functions are:

- Manual and automated import of time-series data from various text formats, support for automated updating from ongoing measurements and mass import
- Gap detection, quality and format consistency checks, and recording of metadata during import
- Spatial analysis and search functions (e.g. distance to next station with same measured parameter), calculation of indicators (e.g. trend) and visualization of time-series data and gaps

- Gap filling toolbox with rule based selection of filling methods (e.g. inverse distance weighting with elevation correction, linear regression, nearest neighbor) and recording of detailed information on used methods and time-series data for each gap
- Visualization and export of

data in different temporal resolutions (original or aggregated), with or without gap filling
- API for direct data access from external tools, e.g. hydrological simulation models (Kralisch et al., 2009)

Time-series data in OBIS are stored together with their source, which can be a measurement station or a simulation model. In both cases the data are related to a spatial reference, e.g. a point or a representative area. The data sources are stored as separate datasets, in the case of measurement stations with name, coordinates, elevation, operation information, photos and other data. Using their location and type (e.g. climate, discharge, or precipitation station), measurement stations are automatically visualized in maps, providing map-based search functionality for stations and time-series (Fig. 3).
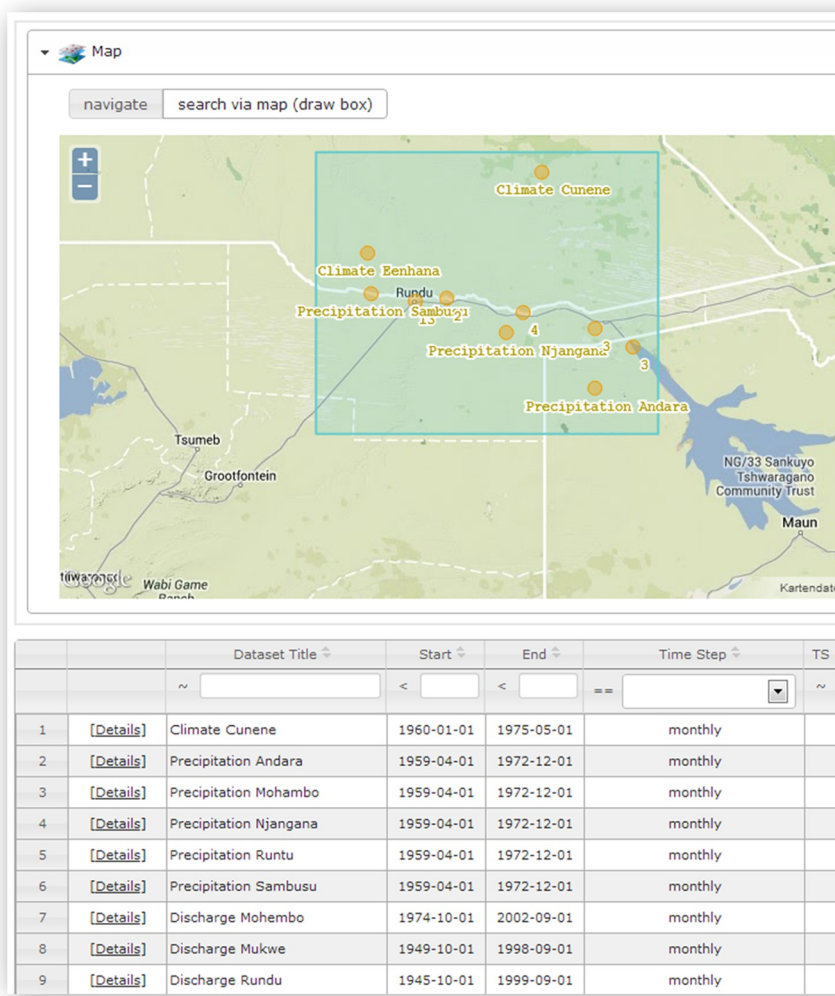

Fig. 3: Screenshot of map based station search and overview functions.

Map data

OBIS makes use of the Mapserver and OpenLayers software packages to provide web-based visualization of map data (Fig. 4). Main functions of the OBIS map data management are:

- User-friendly import and export of geodata using software assistants, data types supported are vector (ESRI Shapefile) and raster data (GeoTIFF, JPEG)
- Definition of user-defined maps combined of single data layers using an integrated map editor
- Visualization of maps based on OpenLayers in a user friendly, feature-rich frontend, including zooming/moving functionality, changing the order of the displayed layers, searching for geometry objects and displaying their attribute values

Stations and other OBIS datasets with defined location (e.g. soil and vegetation data) can be automatically linked to a map and back which opens the option to search for datasets not only by metadata, but also by their spatial location. An example is the link between a measurement station and the corresponding point on a layer/map, which will be automatically created during station creation based on type and location information.

Observation data

OBIS offers various options to describe observations, like surveys, soil, vegetation or water samplings, and other. To describe their location, all observations are assigned to a site dataset, which is represented by site name, description, and location (i.e. a point or polygon). In addition to its study site, an observation is further defined by a title, type, date and description. Further, arbitrary files (i.e. data, documents, images, etc.) can be attached to an observation dataset, allowing for broad range of possible use cases. Due to their assignment to a study site dataset, observations have a location and can easily be grouped, shown on maps and searched for by their location. Fig. 5 shows an OBIS map with study sites of

TFO.

In the case of soil or vegetation information, OBIS allows to store data on a high level of detail. For vegetation samples, information about structure (e.g. coverage, canopy height, LAI, etc.), fragmentation, diversity and others can be stored. Soil data are described by profile (e.g. FAO group/type, profile depth, etc.) and landscape information (geology, landuse, topography, etc.). Moreover, with each profile an open number of soil horizons might be stored in OBIS, containing information about its physical (e.g. silt / clay content, etc.) and chemical properties (soil sorption complex, FeO / AlO contents, etc.). Again, additional files can be stored along with the soil and vegetation datasets.

Scenario data

In order to provide information about projected impacts of future development scenarios, OBIS features functions to store and manage scenarios, indicators and their associated values. In a first step, scenarios (represented by name, description, time frame and associated binary files) and indictors (represented by
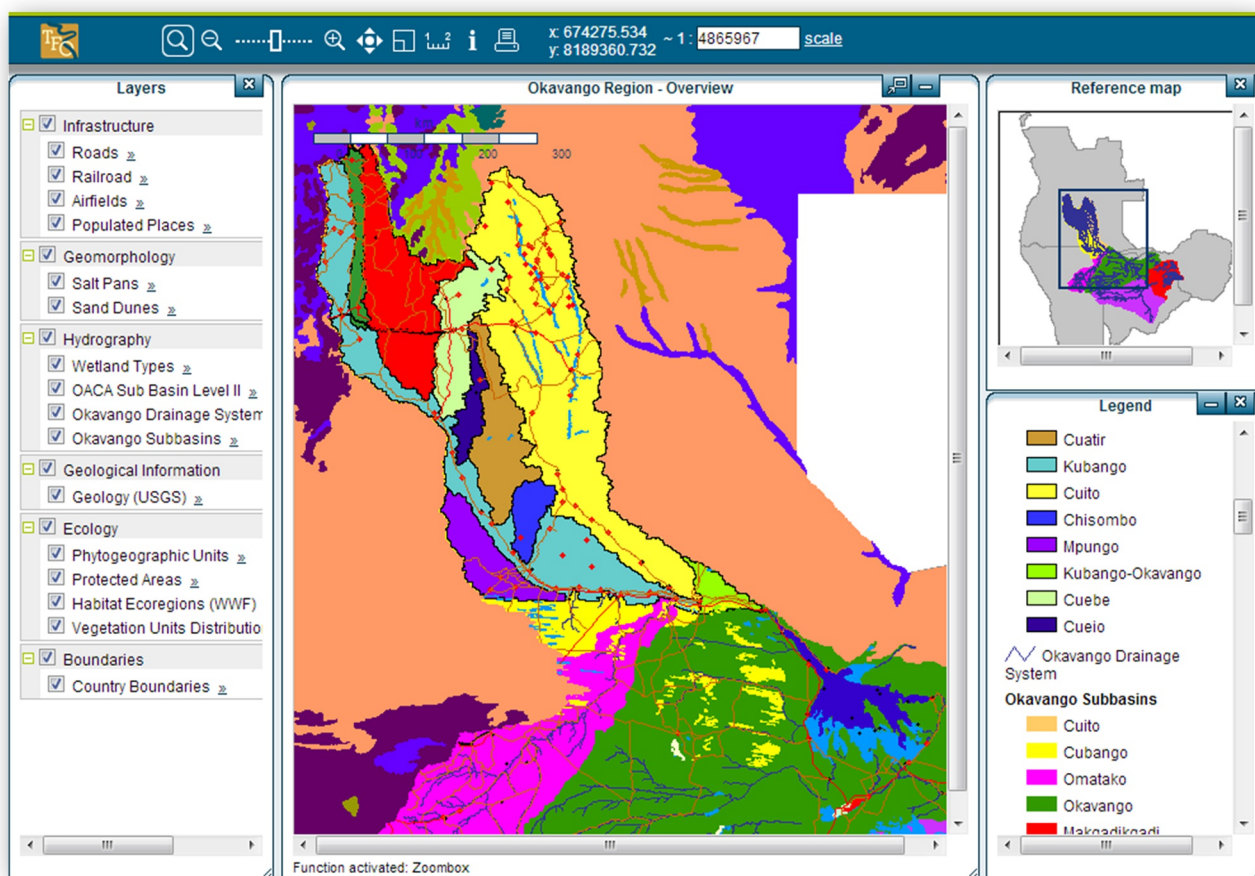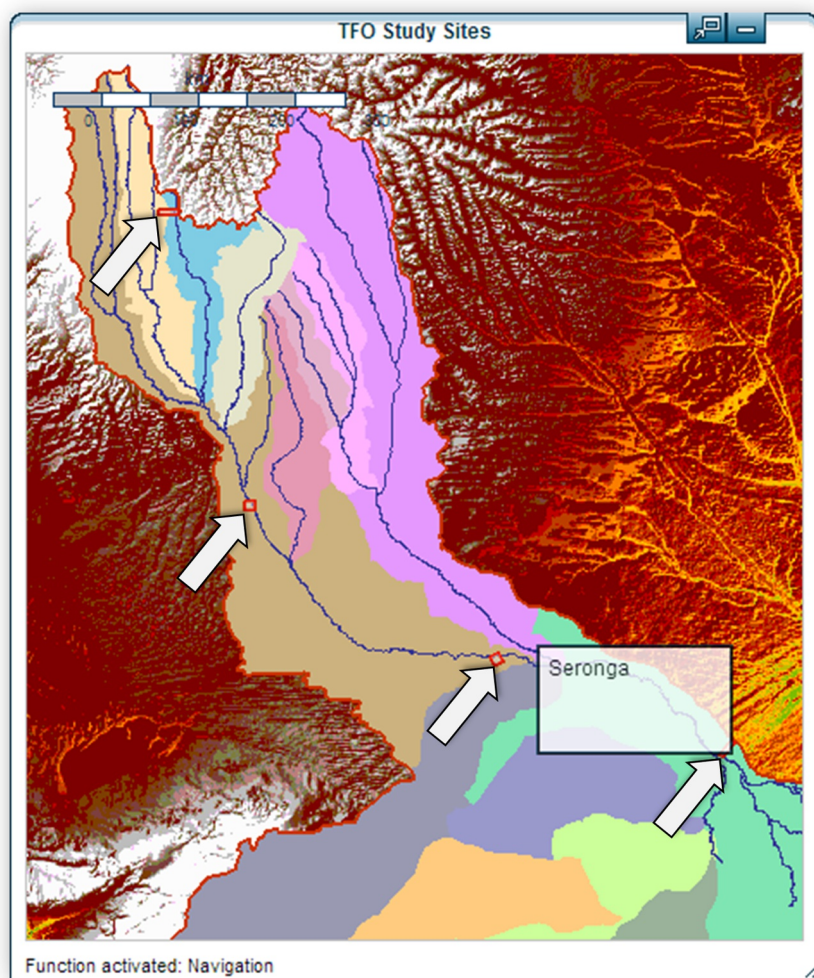

Fig. 4: Visualization of maps in OBIS.

Fig. 5: Automated visualization of study sites.

name, description, calculation method, dimension and unit) are entered separately using dedicated forms. By linking scenarios and indicators with study sites in a second step, the resulting combinations can be assigned values in OBIS, allowing a transparent and flexible association of scenario-based assessments to individual areas like e.g. hydrological catchments or administrative regions (Fig. 6). These values are represented by an individual dataset, which either can directly include the actual value or links to other OBIS datasets like time-series or maps, thus allowing for a really integrated management and visualization of project results. In order to make sure that a large variety of different types of scenarios, indicators and related values can be represented, the data model was kept rather simple, but flexible. Once scenario assessment results have been entered to OBIS, users will be able to e.g. browse for results that are related to a specific site, or to assess all information linked to a certain indicator.

Metadata

The availability of metadata is an important precondition for an efficient discovery and use of environmental information. Their storage and management therefore was a guiding design principle for OBIS. The basis is formed by the ISO 19115 standard for Geographic Information-Metadata (ISO, 2005), which is used to describe all map data. Based on this foundation,
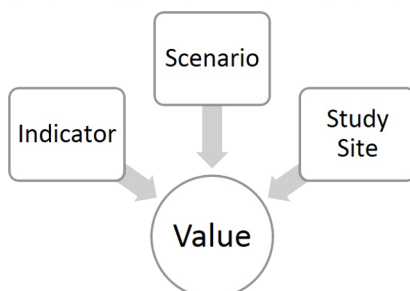
**Fig. 6: Representation of scenario-based information.**

adaptations of the ISO-defined set of mandatory and optional information were used to annotate other data types in OBIS, providing at least information about responsible persons or organizations and date of creation. For geospatial data, uploaded data layers (Shapefiles, GeoTIFFs) can be used in maps only after they have been sufficiently described with meta-information. For time-series and geodata, meta-information are automatically extracted from uploaded data where possible (e.g. spatial / temporal extent). The thorough annotation of OBIS data with meta-information is an import premise for their exposure through standardized interfaces.

New data types

Due to its flexible and modular database and application design, OBIS can easily be tailored to meet emerging demands for new data types and relationships between them. This is achieved by a description layer in the OBIS software that takes care of read/write access to database contents. This layer takes an XML document as input which describes database relations and attributes that define a certain data type, e.g. measuring stations. The XML document is then evaluated by OBIS in order to create SQL statements and user interfaces needed to perform different actions on these data, e.g. to search, browse, edit, add or delete measuring station datasets (Fig. 7).

Taking benefits from this approach, all actions have to be implemented only once and can be re-used to access all kinds of OBIS data. New data types can easily be added by (i) adopting the database design (e.g. by adding new relations and attributes), and (ii) by creating a XML view definition document defining how OBIS shall access these data.

According to the modular database and application design there are only few
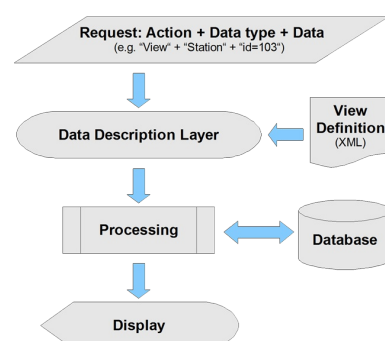
Fig. 7: OBIS information access via data description layer.
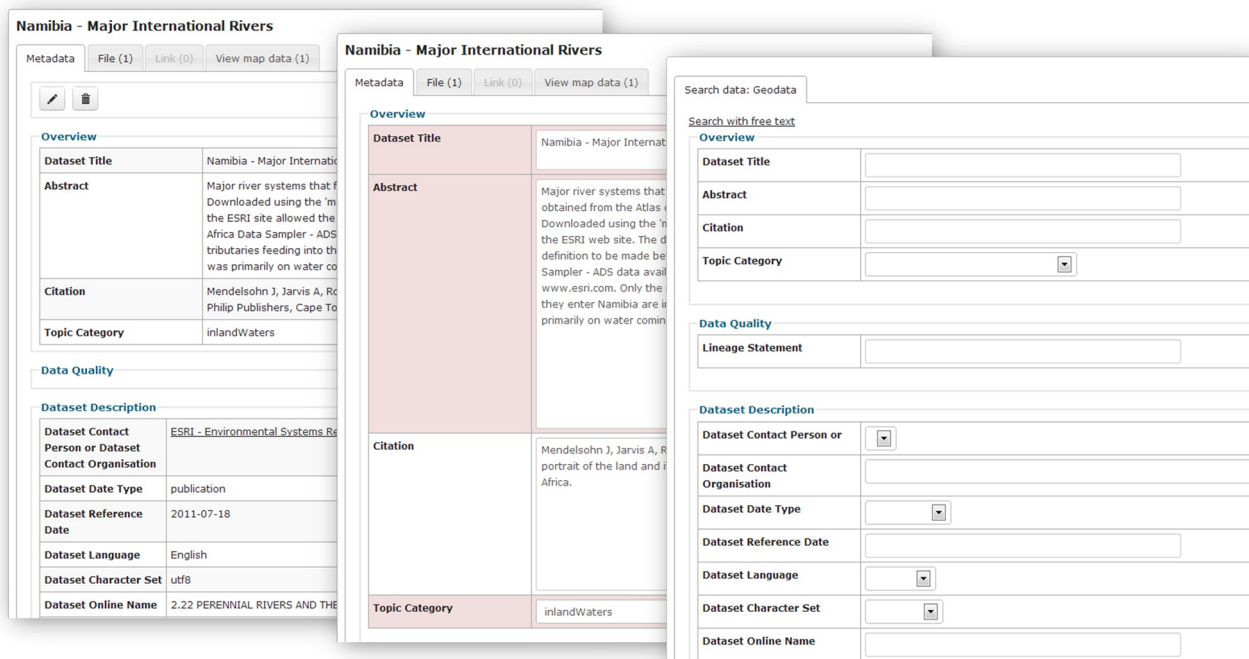
Fig. 8: Auto-generated user interfaces for data type "Geodata".

simple steps necessary to add a new data type to OBIS. At first a database model has to be designed, which describes the new data type and the connections to existing relations e.g. to the ISO 19115 *Responsible Party* relation. The second step is to add new tables and relations to the database. In the last step, the view definition XML document is created holding all information of how OBIS shall access data in the database and how they are displayed to the user. Fig. 8 shows examples of automatically generated user interfaces (view, edit, and search) for the data type "Geodata".

Generic Functions

In addition to the management of environmental data, OBIS provides common functionality for finding and protecting data, data exchange and offline deployment.

Search

OBIS provides two options to search for datasets, i.e. (i) a full-text search which will include all available data and (ii) a search based on individual data types (e.g. stations) which offers fine-grained functions to localize interesting information. While the first option will only allow to enter a single search string that is then used to perform a system-

wide search, the second option will also offer functions to search in individual attributes (e.g. all stations whose "type" attribute has the value "precipitation", to use wildcards or search exactly for a given string and perform a combined search (all grid-based geodata layers from the University of Hamburg).

Permission management

Since OBIS is designed for distributed web access by users with different levels of authorization, a flexible and transparent permission management was an important precondition for its acceptance and trustful use. This is achieved by granting access rights to combinations of single users or user groups, activities to be performed and single datasets or groups thereof. Examples of possible access permissions are:

- Granting read access for unregistered users to metadata only,
- Granting download permission for all registered users to a single time-series dataset,
- Granting permission to create / modify geodata datasets for a specific user group.

Dataset permissions are typically defined system-wide by administrators or data managers, e.g. a user might be

assigned a user group which has read and download permission for all OBIS datasets. However, dataset owners are authorized to refine these permissions for their data on a very fine-grained level, e.g. revoking a general, group-based download permission and granting it to individual users only (Fig. 9). Using this hierarchical approach, dataset owners can always take full control over their data and how they can be accessed and used by others.

In addition to detailed permission management, transparency is accomplished by means of a comprehensive access logging mechanism which is responsible to record write access to any data. This function is also used to create reports about recently added or modified datasets or to notify registered OBIS users via according emails if demanded.

Data exchange interfaces

Along with the provisioning of web based user interfaces for data and metadata management, visualization and analysis, special attention was given to offer the following standardized, service-based interfaces for accessing metadata and maps:

- **Catalogue Web Service (CSW)**: To expose a catalogue of selected metadata records (geodata and time-series data)

Fig. 9: Setting permissions for owned datasets.

on the internet, an OGC standard-compliant CSW has been implemented based on the *pycsw* software package (www.pycsw.org). Using this service, OBIS metadata can easily be exposed to and accessed from standard-compliant third party applications like the GLUES Geodata infrastructure (http://geoportal.glues.geo.tu-dresden.de), which is the joint data and service platform for the Sustainable Land Management research program funded by the German Ministry of Education and Research (BMBF) (Bernard et al., 2013).

• **Web Map / Feature Service (WMS/WFS)**: Based on the Mapserver software package used in OBIS, the WMS and (non-transactional) WFS provide functions to standardized access to maps and geometries stored in the system.

Moreover, OBIS provides several interfaces for the automated import and update of time-series data. The data sources can be local files or online data repositories and web services. An example for such a repository is the Global Surface Summary of the Day (GSOD) product of the National Climatic Data Center (NCDC) in Asheville, USA. The GSOD repository contains data from weather stations worldwide based on the Integrated Surface Database (DSI-3505) (Smith et al., 2011). GSOD data can be automatically harvested by OBIS, providing updated climate time-series for selected stations in the Okavango Basin.

Offline operation

OBIS is a web-based system to enable a collaborative use of stored data and information. However, an offline availability or easy transfer of the system might be desirable. Typical reasons are bad or completely missing internet connectivity or the duplication and transfer of the whole system to project partners. To enable offline access and simplify distribution, OBIS can be operated in a virtualized server environment using the *VirtualBox* software package (http://www.virtualbox.org) in two steps. Firstly, a virtualized version of OBIS is created by means of automated functions, resulting in an identical copy of the OBIS server environment (i.e. operating system, web and database servers, etc.) and OBIS data. The virtualized OBIS which resides in a single file can then be loaded into VirtualBox, an Open-Source Software available for all major operating systems. The operation of the virtual system is fully transparent, i.e. users accessing the system would not notice any difference to the original OBIS. If desired, only a subset of all existing data can be transferred into the virtual environment, leaving room for setups where the resulting system should allow access to a limited amount of data only.

Operation

OBIS is operationally used in TFO and at the same time continuously extended and improved. About 80 users from all around the world are currently registered to the system. In order to raise its acceptance and foster its wide-spread use, OBIS provides graphical user interfaces in German, English, Spanish, and Portuguese language that can be configured by its users. Most of currently existing users are scientists that are interested in accessing available data from the Okavango region or in providing own data to be shared with other users. However, the target user group also includes regional stakeholders and governmental or non-governmental organizations that are interested in TFO project results and baseline information offered already now or in the near future.

Currently, OBIS contains about 140 sets of geodata, 323 time-series datasets and 247 station datasets from the project region. Most of these datasets have been raised outside of TFO and were added by a small group of users aiming to provide a common set of information that supports the ongoing research activities. However, all data were added using the user interfaces described above and thus could have been entered as well by more inexperienced users. New information generated by TFO research activities is expected to enter the system within the forthcoming project lifetime, coming both from German and African scientists.

In order to ensure a sustainable and continuous operation of OBIS beyond the TFO lifetime, an implementation and further hosting in the target region by one of the African partners was found to be a stringent necessity. Here, the Okavango Research Institute (University of Botswana) was identified to provide the organizational and technical capacities required to host, maintain and improve OBIS after the end of TFO and was trained accordingly.

Outlook and Conclusion

Future work on OBIS and its underlying RBIS software platform will include the implementation of further, standardized data exchange interfaces addressing environmental simulation models, i.e. the *Sensor Observation Service* (SOS) and *WaterML2.0* data encoding standard (OGC, 2012a, OGC, 2012b). Further effort will be spent on refining and improving functions for the description and management of future development scenarios and results of related impact

assessment studies. An extension for a detailed description of environmental simulation models and results based on their application aims to improve transparency of model-based assessments. Looking forward to results from other TFO groups becoming available soon, another focus of work will be on their integration in OBIS to form a project-wide result dissemination platform.

Present results and experiences show that OBIS is a useful and adequate platform for the distributed, web-based management of environmental information. Providing storage, processing and visualization support for a large variety of data, its modular software architecture and standardized service interfaces ensure future-proof extensibility and connectivity, addressing upcoming requirements of environmental research and data management. Providing functions for an easy deployment in centralized and non-centralized setups, OBIS addresses the demands of TFO stakeholders and partners for an integrated decision support and information management platform in the Okavango region.

Acknowledgements

References

Bernard, L., Mäs, S., Müller, M., Henzen, C., Brauner, J. (2013): Scientific geodata infrastructures: challenges, approaches and directions. – International Journal of Digital Earth. 1–21. CrossRef

EU (2007): Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Official Journal of the European Union, L 108, 50.

Flügel, W.-A. (2011): Development of adaptive IWRM options for climate change mitigation and adaptation. – Advances in Science and Research **7**: 91–100. CrossRef

Grobe, H., Diepenbroek, M., Dittert, N., Reinke, M., Sieger, R. (2006): Archiving and Distributing Earth-Science Data with the PANGAEA Information System. – In: Fütterer, D.K., Damaske, D., Kleinschmidt, G., Miller, H., Tessensohn, F. [Eds.]: Antarctica. Springer-Verlag, Berlin/Heidelberg. CrossRef

International Standards Organization (2005): International Standard ISO 19115 Geographic information – Metadata. Reference Number ISO19115:2005(E).

Kralisch, S., Böhm, B., Böhm, C., Busch, C., Fink, M., Fischer, C., Schwartze, C., Selsam, P., Zander, F., Flügel, W.-A. (2012): ILMS – a Software Platform for Integrated Water Resources Management. – In: Seppelt, R., Voinov, A.A., Lange, S., D. Bankamp [Eds.]: Proceedings of the International Congress on Environmental Modelling and Software, Sixth Biennial Meeting, Leipzig, Germany.

Kralisch, S., Zander, F., Krause, P. (2009): Coupling the RBIS Environmental Information System and the JAMS Modelling Framework. - In: Anderssen, R., Braddock, R., and Newham, L. [Eds.]: Proc. 18th World IMACS/and MODSIM09 International Congress on Modelling and Simulation. Cairns, Australia.

Michener, W., Vieglais, D., Vision, T.J., Kunze, J., Cruse, P., Janée, G. (2011): DataONE: Data Observation Network for Earth - Preserving Data and Enabling Innovation in the Biological and Environmental Sciences. D-Lib Magazine 17. CrossRef

Open Geospatial Consortium Inc. (2007): OpenGIS® Catalogue Services Specification.

Open Geospatial Consortium Inc. (2012a): OGC ® Sensor Observation Service Interface Standard.

Open Geospatial Consortium Inc. (2012b): OGC ® WaterML 2.0: Part 1- Timeseries.

Smith, A., Lott, N., Vose, R. (2011): The Integrated Surface Database: Recent Developments and Partnerships. – Bulletin of the American Meteorological Society **92**: 704–708. CrossRef

Tarboton, D.G., Maidment, D., Zaslavsky, I., Ames, D., Goodall, J., Hooper, R.P., Horsburgh, J., Valentine, D., Whiteaker, T., Schreuders, K. (2011): Data Interoperability in the Hydrologic Sciences – The CUAHSI Hydrologic Information System. – In: Proceedings of the Environmental Information Management Conference 132–137.

Zander, F., Kralisch, S., Busch, C., Flügel, W.-A. (2011): RBIS - An Environmental Information System for Integrated Landscape Management. - In: Hřebíček, J., Schimak, G., Denzer, R. [Eds.]: Environmental Software Systems. Frameworks of eEnvironment 349–356. Springer Berlin Heidelberg, Berlin, Heidelberg. CrossRef

Affiliations

Sven Kralisch* (sven.kralisch@uni-jena.de), Franziska Zander (franziska.zander@uni-jena.de), Wolfgang-Albert Flügel (c5wafl@uni-jena.de)
Department of Geoinformatics, Hydrology and Modelling, Friedrich-Schiller-University
Löbdergraben 32
07743 Jena, GERMANY

*Corresponding author