Jun 18th, 9:00 AM - 10:20 AM

An information platform fostering re-use of water data

Ralf Denzer
Environmental Informatics Group, ralf.denzer@enviromatics.org

Sascha Schlobinski
cismet GmbH, sascha.schlobinski@cismet.de

Gerben Boot
DELTARES, gerben.boot@deltares.nl

Frank Keppel
DELTARES

Erik de Rooij
DELTARES

International Environmental Modelling and Software Society (iEMSs)
7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA,
Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2014-proceedings

An information platform fostering re-use of water data

Ralf Denzer [1, 2], Sascha Schlobinski [2], Gerben Boot [3], Frank Keppel [3], Erik de Rooij [3]

[1] Environmental Informatics Group, Germany (ralf.denzer@enviromatics.org),
[2] cismet GmbH, Germany (sascha.schlobinski@cismet.de),
[3] DELTARES, The Netherlands (gerben.boot@deltares.nl)

Abstract: The re-use of publicly funded governmental data has received a lot of attention recently. In the environmental domain, it is clear that improved public services need exchange of data across governments at all levels. Governments keep producing information products at all levels, and some of them are more or less readily available. Reporting obligations in the EC, for instance demanded by the Water Framework Directive (WFD), are direct inputs to European datasets. The technology base for advanced systems is supported by the OGC suite of standards and by encodings for water information.

Still re-use of data is not happening at large scale, and platforms, particularly those developed in R&D projects often do not survive long after the project. Recent discussions in the community have identified a fundamental gap between concepts, research and implementation. This gap is given by a) a distance between the modelling community / modelling environments and infrastructure developers / providers, b) a lack of tools supporting the uptake of infrastructures being built and a lack of understanding, how "re-purposing" (re-use under different context) of data can be supported.

This paper describes the concept of a water information platform which aims at making re-purposing of water information for water science seamlessly possible. This platform will be developed in the context of a European FP7 research project.

***Keywords**:* Water information platform; spatial data infrastructure; OGC services; data repurposing

1 INTRODUCTION

The re-use of publicly funded governmental data has received a lot of attention recently. In the environmental domain, it is clear that improved public services need exchange of data across governments at all levels.

In Europe, policy driven programs like INSPIRE[1] and Copernicus[2] have made large efforts in recent years to provide data and services for various purposes. At the same time, technology driven programs such as "ICT for Risk Management" and "ICT for Sustainable Development" have attempted to provide a technology base to build large scale service networks, in particular by funding a number of ICT projects in EU-FP6 and EU-FP7, for instance published in Usländer(2010) and Douglas (2008), and most recently, in the EnviroFI project[3].

[1] http://inspire.jrc.ec.europa.eu
[2] http://www.gmes.info
[3] http://envirofi.eu

At the same time governments keep producing information products at all levels, and some of them are more or less readily available. Reporting obligations in the EC, for instance demanded by the Water Framework Directive (WFD[4]), are direct inputs to European datasets like those maintained in WISE[5]. The technology base for advanced systems is supported by the OGC suite of standards[6], and by encodings for water information[7].

Still re-use of data is not happening at large scale, and platforms, particularly those developed in R&D projects often do not survive long after the project. Recent discussions in the community have identified a fundamental gap between concepts, research and implementation. This gap is given by:

1. a distance between the modelling community / modelling environments and infrastructure developers / providers, as identified in Argent (2007)
2. a lack of tools supporting the uptake of infrastructures being built and
3. a lack of understanding, how "re-purposing" (re-use under different context) of data must be supported by collaboration of humans as well as software tools, as recently identified at the Australian Water Alliance Symposium WIRADA (WIRADA 2011, Denzer 2011a).

This paper describes the architecture of a water information platform which aims at making re-purposing of water information for water science seamlessly possible. This platform is currently developed in the context of a European research project, in an evolutionary process, and a first version has just gone on-line.

2 CONCEPTUAL VIEW OF AN IT PLATFORM SUPPORTING WATER DATA RE-USE

As part of the SWITCH-ON project[8], a Spatial Information Platform (SIP) is under development, which will support the reuse of information products for the water science domain (figure 1). The concept of the SIP has been developed as part of the project preparation by the co-authors of this paper, and the implementation is under way since the end of 2013.

On the input side, the SIP will ensure access to external catalogues. For seamless integration, tools for conversion and schema mapping facility may be necessary. Metadata about external sources and transformation on them will be created as part of the SIP Catalogue. The catalogue and data storage will be based on a common information model, which will be developed in close collaboration between IT partners and water scientists. This information model will use (and extend where necessary) standards of the OGC suite of services, specifications and information models. Internally the platform will serve the project, through interactive retrieval interfaces both for humans and software clients (lowest platform layer).

Where new information products are created within the project, the platform will support the publication of result datasets into the platform catalogue, which will be exposed to the outside world through interactive interfaces and service API's. The platform will also provide a system management facility which supports the management of data models, catalogues, access rights, technical interfaces etc.

A set of tools will also be part of the platform. These tools will be made available in modular fashion and can be used a) by internal processes of the platform (for instance where conversion is concerned), b) by internal and external users through interactive user interfaces, c) internal and external software components through service API's.

While this toolbox will contain some general purpose tools, for instance for visualisation, the main focus is to support water-related analysis and processes, the repurposing of data in the water context, and the documentation of scientific analysis through metadata.

[4] http://ec.europa.eu/environment/water/water-framework
[5] http://water.europa.eu
[6] http://www.opengeospatial.org
[7] http://www.opengeospatial.org/projects/groups/waterml2.0swg
[8] http://www.water-switch-on.eu

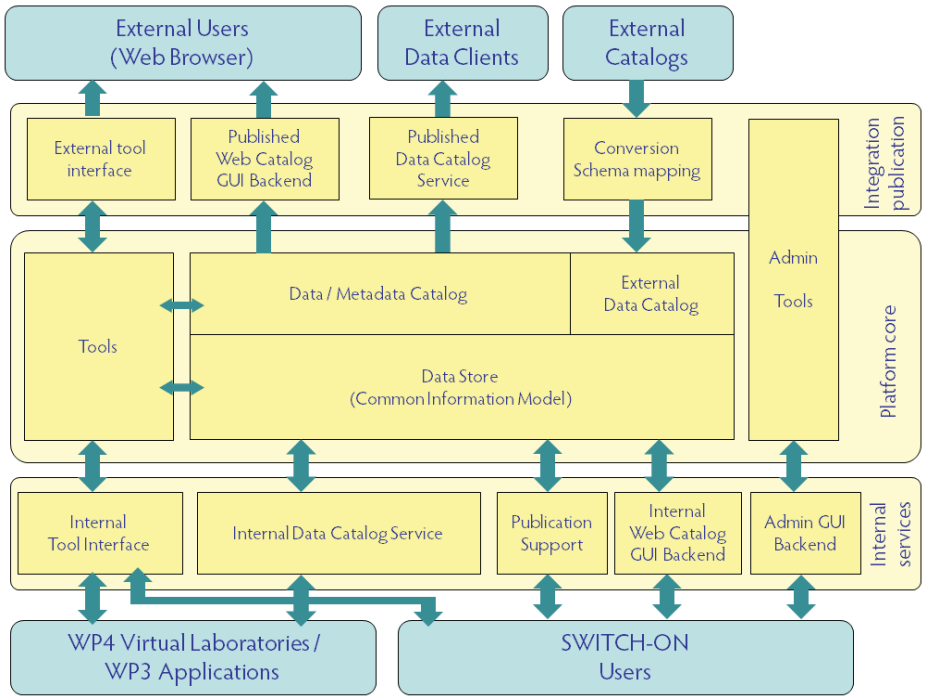

Figure 1. Conceptual view of the Spatial Information Platform (SIP).

3 END USERS OF THE PLATFORM

The SIP is intended to serve mainly two types of end users. First, water scientists will be supported to find open data and tools relevant for their individual experiments and in return publish their own datasets as well as the results of repurposing or research. Second, the platform will be available to businesses which use the available information (open data published through the SIP) as input for value-added software products. During the project, fourteen such value-added applications will be piloted. The platform will support both researchers (through GUI's) and developers involved in the creation of value-added applications (through API's) to access open data and tools relevant for their specific context.

4 RELATED WORK

Water-related information is progressively made available on-line by a large variety of actors. However, "on-line" means different concepts to different people. The following concepts are established approaches:

1. presenting data through an interactive discovery process in a web environment, with *no access to the original data*, and limited output capabilities (like maps)
2. making datasets available for *download*, including or not including interactive discovery
3. providing data API's for inclusion into value-added software, through *proprietary interfaces*
4. providing data API's for inclusion into value-added software, through *technically standardised* interfaces
5. providing data API's for inclusion into value-added software, through *technically standardised* and *semantically harmonised* interfaces, based on common information models

Each approach may be based on entirely open or more closed user communities, and with or without registration and authentication.

As the project specifically addresses the re-use of open government data, information sources from public authorities, but also those emerging from science, are particularly important. Water data from operational monitoring and reporting of authorities is available at many web sites for viewing and / or downloading. In Europe, the EEA plays and important role by providing water information through WISE and the water data centre[9]. Data does not only include monitored or reported data but also fundamental information like basin networks (EEA 2012). Over the past years, the Australian government has made large efforts in building water information platforms, information systems and tools (BOM 2013). Similarly, in the United States, CUASHI[10] acts as an alliance to improve water information and associated ICT. At international level datasets like those from the GRDC (global data runoff centre)[11] are important sources. The GEOSS data core also lists many water-related information resources (GEOSS 2013).

5 TECHNICAL BASELINE OF THE CORE PLATFORM

The core platform will be implemented with a stack of freely available open source software compiled in the CIDS product suite of CISMET GmbH. This software suite consists of a set of software components, application programming interfaces (APIs), management and development tools, services and applications, with a special focus on interactive solutions which need to integrate geo-spatial systems with databases, sensor networks, document-oriented systems, unstructured information sources and numerical models. CIDS is particularly suited for solutions which have to be built across existing heterogeneous information systems, which may be under control of different organizations.

CIDS uses freely available open source software exclusively, is itself freely available as open source software, distributed as Github[12] projects, and is compliant with all major OGC services. CIDS has been used in numerous projects since 1999, see Güttler(2000). For recent projects see Denzer(2011b), Schlobinski(2011) and Denzer(2013).

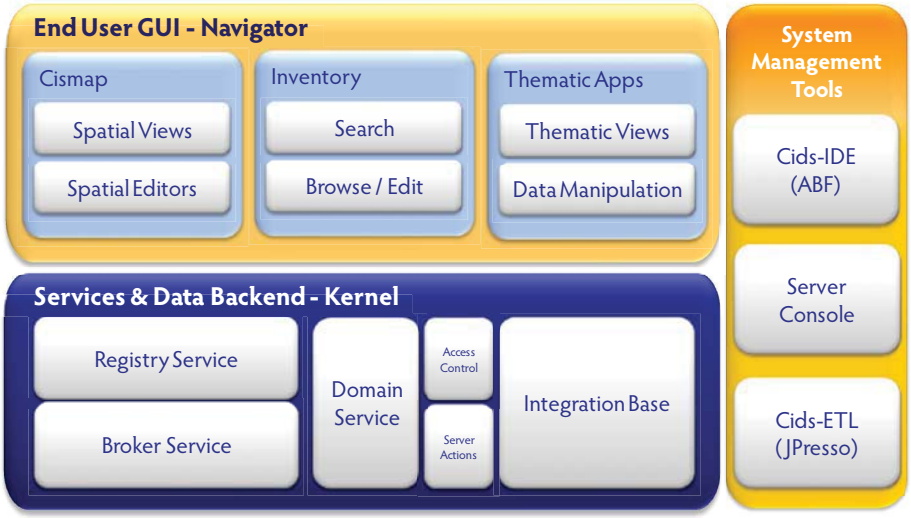

Figure 2. CIDS architecture.

A CIDS Service Network is configured and developed with an interactive development and configuration environment (CIDS-IDE). As open source software the CIDS platform uses many open

[9] http://water.europa.eu/themes/water/dc
[10] http://cuashi.org
[11] http://www.bafg.de/grdc
[12] https://github.com/cismet

source tools to provide its functionality. Among the most prominent are Netbeans RCP[13] and PostgreSQL/PostGIS[14]. For a detailed description of the platform, and its use in the context of a governmental water information system, see Hell(2013).

6 ANTICIPATED TOOLS FOR RE-PURPOSING OF WATER DATA

The CIDS platform is a generic infrastructure which is independent of the application domain. Each CIDS installation is configured with the actual information models for the particular solution, and is extended with custom viewers, data manipulation GUI's, visualization and processing components. When necessary, external processes (e.g. numerical models) are linked via service calls.

Regarding the water-science specific tools, a custom shop with tooling is foreseen within the SIP. This tooling will supply the more domain (hydrology) specific functionalities needed to more efficiently enable re-purposing of water data. Often, once data is found, it will require one or more processing steps before it can actually be (re)used. For example: the source data may be a BIL file while the model requires a NetCDF file or the source data can be raw sensor data still containing outliers and missing values. Besides that, it might be necessary to interpolate missing values, aggregate time steps or convert between measurement units.

The above examples describe three types of processing steps that are often required to prepare data before use; Conversion, Cleaning and Transformation. Data conversion is the process where the data is converted from one format to the other. Data cleaning is the process of detecting and correcting invalid records in a dataset and data transformation is the process in which data is transformed using some kind of algorithm.

In order to successfully re-purpose data it must be possible to incorporate one or more processing steps. Because the number of processing possibilities are limitless and it is not desirable to have a separate interface for each individual processing type a generic interface is required. A good candidate for this interface is the OGC-WPS[15] (Web Processing Service). Firstly it fits the requirement to act as an interface for all conventional processing types. Secondly it is an open OGC standard and therefore blends well with the other OGC style interfaces in the SIP. Finally it fits the concept of a distributed system. This is important as the data sources being re-purposed are potentially distributed (Baart, 2014).

As proof of concept the Data Conversion Module (DCM) produced by Deltares will be made available as an OGC-WPS. The DCM is an application that can be used to perform numerous data-conversion, cleaning and transformation processes. As project deliverable the DCM can host a number of data conversions and transformations such as: Grid conversions (between ArcInfoAscii, BIL, GRIB1, GRIB2 and NetCDF-CF), scalar conversions (between SWE, WaterML2, CSV and NetCDF-CF), interpolate missing values and time step (dis)aggregations (Grijze, 2014).

Custom tooling to supply specific and/or additional functionality on data discovery, data transformation, data visualization or data dissemination to support data re-purposing, is to be made available as services through one open standard: OGC-WPS. The set of services to be provided by the SIP can change depending on the actual domain needs.

[13] http://netbeans.org/features/platform
[14] http://www.postgresql.org
[15] http://www.opengeospatial.org/standards/wps

7 FIRST VERSION

The first version of the SIP has become available early in the project. It is based on the CIDS Navigator. The SIP data model is based on extensions of existing standards from OGC and the Comprehensive Knowledge Archive Network (CKAN) of the Open Knowledge Foundation[16]. This data model will be published during the course of the project.

The three basic use cases implemented cover the following core functionalities:

1. browse, find and view data,
2. import and export data and
3. re-purpose data.

Use case 1) includes geographic and thematic browsing supported by a catalogue of resources and search functionalities. Use case 2) supports end users in publishing their resources (select a resource, format and associated meta-data for import into the SIP) with the help of a wizard. Use case 3) supports end users to apply algorithms to a chosen set of data, and to store the newly created data as a new resource (re-purposing).

A video demonstrating these first use cases of the SIP is available at youtube[17].

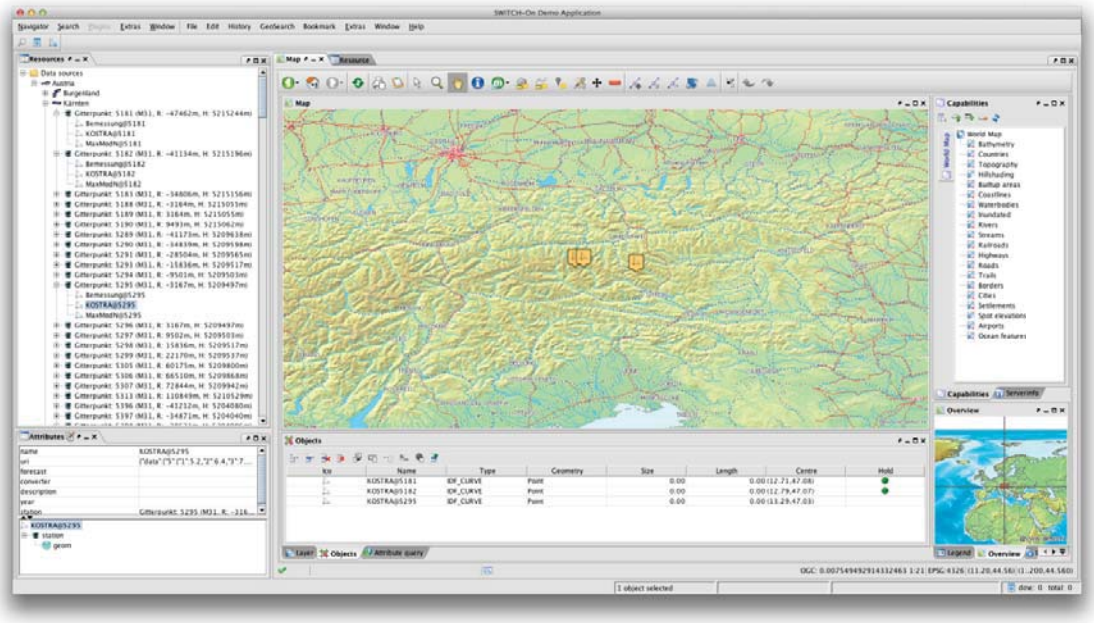

Figure 3. GUI of the first SIP version.

8 FUTURE WORK

The SIP will be developed over the coming 2 years in an incremental fashion. Data sources will be successively integrated, new use cases will be supported, and water-related tools will be incorporated into the service network. The continuous use of external data sources will very much depend on their availability, quality and documentation of data provenance.

[16] https://okfn.org/
[17] http://www.youtube.com/channel/UCOAPooTB-gFIsnCAN4qu5Xw

ACKNOWLEDGEMENTS

The SWITCH-ON project is funded under the European Framework Program FP7 (contract number 603587).

REFERENCES

Argent R.M., R. Denzer, R. Güttler (2007). Modelling Environments and Use of Service Infrastructures, MODSIM 2007, International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand, December 2007, pp. 833-838

Baart, F., G.J. de Boer, W. de Haas, G.V. Donchyts, M.E. Philippart, M. van Koningsveld, M. Plieger. A comparison between WCS and OPeNDAP for making model results and data products available through the internet in: Transactions in GIS Vol.16 nr. 2 (2012) pp. 249-265, Blackwell Publishing Ltd

BOM (2013). Austrian Government, Bureou of Meteorology, Improving Water Information Programme, Progress Report, Advances in water information made by the Bureau of Meteorology in 2013, http://www.bom.gov.au/water/about/publications/document/progress_report2013.pdf

Denzer R., 2011a. Hydroinformatics: Interoperability, standards and governance of water information infrastructures, Proceedings of the WIRADA Science Symposium, http://www.csiro.au/WIRADA-Science-Sympsosium-Proceedings

Denzer R., Torres-Bejarano F., Hell T., Frysinger S., Schlobinski S., Güttler R., Ramírez H., 2011b. An Environmental Decision Support System for Water Issues in the Oil Industry, Environmental Software Systems Vol. 9 - Frameworks of eEnvironment, Springer Advances in Information and Communication Technology 359, pp. 208-216, 2011

Denzer R., Schlobinski S., Gidhagen L., Hell T., (2013) How to build Climate-Change enabled EDSS, in: Hrebicek J. et al (eds.), Environmental Software Systems, Fostering Information Sharing, IFIP AICT 413 (2013), pp. 464-471, Springer

Douglas J., Th. Usländer, G. Schimak, J. Esteban, R. Denzer, 2008. An Open Distributed Architecture for Sensor Networks for Risk Management, Journal Sensors 2008, 8, pp. 1755-1773

EEA, 2012. EEA Catchments and Rivers Network System, ECRINS v1.1, ISSN 1725-2237, http://www.eea.europa.eu/publications/eea-catchments-and-rivers-network

GEOSS (2013). Group on Earth Observation, GEO-IX, 22-23 November 2012, Report of Data Sharing Working Group, Document 13 ,https://www.earthobservations.org

Grijze, A.H., P.J.A. Gijsbers, E. de Rooij, O van den Akker. Technology behind the Deltares Open Archive In: D.M.Ames and N.Quinn (Eds)Proceedings of the 2014 International Congress on Environmental Modelling and Software (iEMSs), San Diego, California, USASession A2: Sharing Scientific Environmental Data and Models.

Hell T., Kohlhas E., Schlobinski S., Denzer R., Güttler R., An information system supporting WFD reporting, in: Hrebicek J. et al (eds.), Environmental Software Systems, Fostering Information Sharing, IFIP AICT 413 (2013), pp. 403-413

Güttler R., Denzer R., Houy P., (2000). An EIS Called WuNDa, Environmental Software Systems Vol. 3 (2000) – Environmental Information and Decision Support, pp. 114-121, Kluwer Academic Publishers

Schlobinski S., Denzer R., Frysinger S., Güttler R., Hell T., 2011. Vision and Requirements of Scenario-Driven Environmental Decision Support Systems Supporting Automation for End Users, Environmental Software Systems Vol. 9 - Frameworks of eEnvironment, pp. 51 - 63, Springer Advances in Information and Communication Technology 359, 2011

Usländer Th., Denzer R., 2010. Requirements and Open Architecture of Environmental Risk Management Information Systems, in: B. van der Walle (ed.), Information Systems for Emergency Management, M. E. Sharpe Publishers, ISBN 978-0-7656-2134-4, pp. 344-368, 2010

WIRADA 2011. Proceedings of the WIRADA Science Symposium 2011, http://www.csiro.au/WIRADA-Science-Sympsosium-Proceedings.