Overview Of Coupling Of Data, Models And Information Through The Web Using Existing Standards

de Boer, Gerben; Baart, Fedor; Jagers, B; Becker, B.P.J.

Publication date
2014

Document Version
Final published version

Published in
Proceedings 11th International Conference on Hydroinformatics

8-1-2014

Overview Of Coupling Of Data, Models And Information Through The Web Using Existing Standards

Gerben de Boer

Fedor Baart

Bernhard P. J. Becker

Bert Jagers

11th International Conference on Hydroinformatics
HIC 2014, New York City, USA

OVERVIEW OF COUPLING OF DATA, MODELS AND INFORMATION THROUGH THE WEB USING EXISTING STANDARDS

GERBEN J. DE BOER (1,2), FEDOR BAART (1,2), BERNARD BECKER (1), BERT JAGERS (1)
(1): Deltares, Rotterdamseweg 185, 2629 HD, Delft, the Netherlands
(2): Delft University of Technology, Hydraulic Engineering Department, Delft, the Netherlands

Assessment of environmental status and integral safety requires combination of information from many sources, coming from either databases or increasingly via live model (scenario) simulations. Many of these models require input from one another, sometimes unidirectional, but more and more two-directional as well. Many protocols and frameworks are available for model coupling, often based on open standards and implementations. Previous overviews of coupling protocols have focused on data exchange volume, data complexity, invasiveness into existing models and support for specific programming languages. We extend the overview using recent developments in web-based protocols and focus on the suitability for internet-based data exchange. We also extend the focus of previous reviews by also taking the coupling with aggregated information products for end-users into account. We propose a hierarchical ordering of all standards for specific types of end-users.

INTRODUCTION

Assessment of environmental status and integral safety requires combination of information from many sources. This information is traditionally an ingest of data or an ingest from model results. Philosophically these two sources of numbers have a different nature: generation of information from data is considered an inductive process, while generation of information from model is a deductive process.

Inductively generated information is solidly based on physical reality due its reliance on measured data, but its derived information is logically not universally valid due to limited resolution and coverage of the underlying data. An example is that flood risk curves are based on only a few decades of data, whereas they cover return periods up to thousands of years. In contrast, deductively generated information is universally valid regarding its premises, but there is no guaranteed connection of these premises – and hence it's derived information - to the physical world. An example is the use of extreme value theory (e.g. Weibull distribution) for making real world predictions.

People who use multiple models are more often right (Tetlock, 2005 [1]), thus one can best use a combination of these methods: deductive methods such as numerical models that have been validated or assimilated with field data, and inductive methods such as data collections that have been constrained with physical principles such as simple mass balances.

The existing methods of coupling of data and models are perceived as either data-centric, model-centric or information-centric (Figure 1), where the user's focus is narrowed down on how to get something into one's own component and how to export it from one's own component. The fact that there are different sets of standards for coupling to data, models or graphical end-products does not help in making users view the whole as a coupled web. Thus, there is a growing need of methods that facilitate generic coupling of both data and models via the web. In fact, the terms importing and exporting could perhaps better be phased out, and be replaced by the term coupling. The word coupling implies that two sides are aware of each other, where import and export assume the other party has to adjust. Via web coupling, the already common Data as a Service (DaaS) is indeed being extended with Models as a Service (MaaS) concepts (Geller & Turner, 2007 [2]; Roman, 2009 [3]) and Tools as a Service (TaaS).

The Tools as a Service layer consists of combining data and model results and presenting or transforming them in a user suitable form. There are many examples where the Tool as a Service layer is required. The interfaces between data and models allow them to produce ever more and better information for non-specialists. The Open Geospatial Consortium (OGC) offers wide range of standards to publish information in geospatial viewers. However, for many potential users, the information in geospatial viewers is too raw, too complex or does not for their need. Recently the importance of this last step in information processing has become more apparent with the growing use of smart phones and tablets. Many sites that display model results struggle to provide both information that is useful for both information interactors (specialists, professionals) and information consumers.

Apart from the appearance of new information display devices, there is the issue that some communities do not take geospatial coordinates into account at all. On one hand, for example, insurance companies and government bodies that deal with (flood) risk assessment often prefer their information to be aggregated to mere tables, where the primary axis of representation is time, money or probability. To produce this information from a coupled DaaS/MaaS web, requires an extra processing step. This step is often seen as just-another output conversion. However, we argue that this extra processing step towards aggregated information should be considered as an extra component, as a TaaS component. This step sometimes involves a massive computation too, just like models. For instance, a flood risk assessment requires the

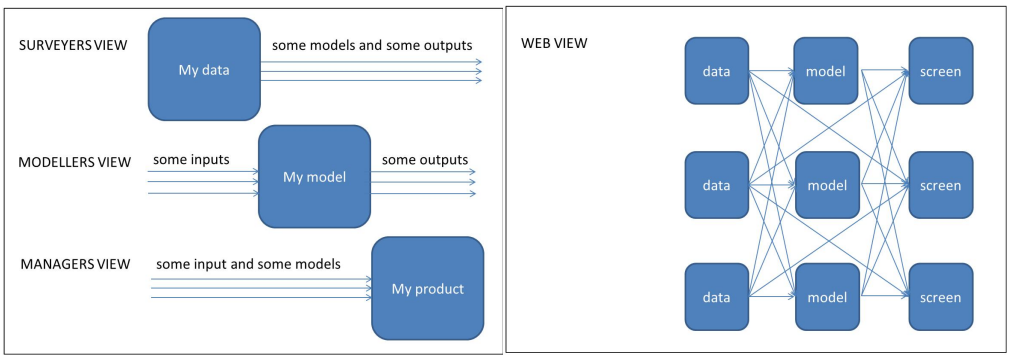

Figure 1. The classic view (left) where data producers, model programmers and data producers only think about interfaces to their own component. In contrast the web view (right) where data, models and any other operations (tools) all interlink to produce static or on-the-fly pieces of condensed information on interactive smart phones and tablet screens.

spatial convolution of maps, asset value, flood risk, vulnerability and of resilience. Depending on the data and convolution methods used, it is deductive-biased, inductive biased or hybrid.

On the other hand, often the databases of organizations contain only aggregated information that does not have a geospatial coordinate attached to it at all. Hence, if users wanted to make a geospatial plot of such overall risks, e.g. using the spatial domain associated with a certain overall risk, it is not possible to plot this without taking extra data sources into account. The simplest example is a so-called *gazetteer* which contains the polygons associated with certain geospatial names to which risk is assigned. Thus non-spatial information requires coupling to the MaaS concepts on one hand for generation from geospatial convolutions, and requires coupling to DaaS concepts on the other hand to be disseminated.

In this manuscript we state that we do not only need web standards that allow seamless coupling of DaaS service and MaaS services, as is already ongoing, but also standards that encompass the tools that process and tailor data into aggregated information (Figure 1): Tools as a Service. In this paper we examine the suitability of the model coupling techniques for web based coupling of tools. We skip DaaS standards as they are well-known from literature. We extend existing overviews with examples of coupling models using OGC Web Processing Services and websockets and assess their use for MaaS application.

REVIEW OF STANDARDS

In this section we discuss existing standards for exchange of information between data and models. Overviews of existing model coupling frameworks are given by Jagers (2010) [4] and Nativi (2013) [5]. We extend these overviews with two criteria: web-coupling and inclusion of viewing services, and propose an overarching conceptual hierarchy.

Integrating existing coupling methods with web

Some model coupling frameworks discussed in Jagers (2010) are not focused on web enablement. The CCA has not seen any development towards web enablement. The MCT (Larson, 2005) [6] is mainly focused on coupling Fortran based models through MPI. The ESMF framework (Hill, 2004) [7] also has a HPC focus with MPI and Fortran support. Web service support was added by Saint (2010) [8], who showed how the framework can be used to connect an HPC based ESMF model with an OpenMI (Gregersen, 2007) [9] based desktop model. Goodall (2011) [10] showed how an OpenMI service can run behind a WPS service.

Model visualizations

Baart (2014) [11] used the MMI protocol on top of a BMI model to enable real-time model interaction. The MMI protocol sends messages with arrays and metadata, for example CF attributes, over a message protocol. These messages were used in the 3Di project[1] to feed a WMS service that can visualize real time overland flooding and through a websocket to update hydrological structures in the browser.

Hierarchy of standards

The existing standards for DaaS services have been categorized for different kind of end-users by Baart et al. (2012) [12] (Figure 2a). They proposed a hierarchy ranging from scientists who can handle complex standards, via professionals who expect more uniformity to be cost-effective, to any end-user that expects nothing more than graphical displays of data, or

[1] http://www.3di.nu

catalogue entries. The low-level, general applicable, data standards such as netCDF or PostgreSQL have not become outdated due to the appearance of new OGC standards like WFS, WMS. On the contrary, under the hood a proper implementation of these low-level standards is essential to be able to serve an OGC WxS service in the first place. New standards appear only once previous standards have reached maturity in terms of formal and accepted descriptions, plus de facto implementations. The definition of various levels is similar to the OSI model (Zimmermann, 1983) [13] for network layers and to the 5 stars rating proposed by Tim Berners-Lee[2]. Here we propose a similar hierarchy approach for coupling of models.

The coupling of models via the web via interactive displays requires the existing models to become orders of magnitude faster. The previous generation of standards – operating in the memory of a local PC or computing cluster – is the very tools that will need to provide the very increase efficiency required to use the next generation of standards. Figure 2b places the model coupling standards discussed in the extended overviews by Jagers (2010) and Nativi (2013) into such a hierarchy. The OGC WPS standard is an example of a standard that allows to function as a TaaS. This WPS standard and the MMI-WMS coupling of Baart (2014) are in the upper right hand side of Figure 2b.

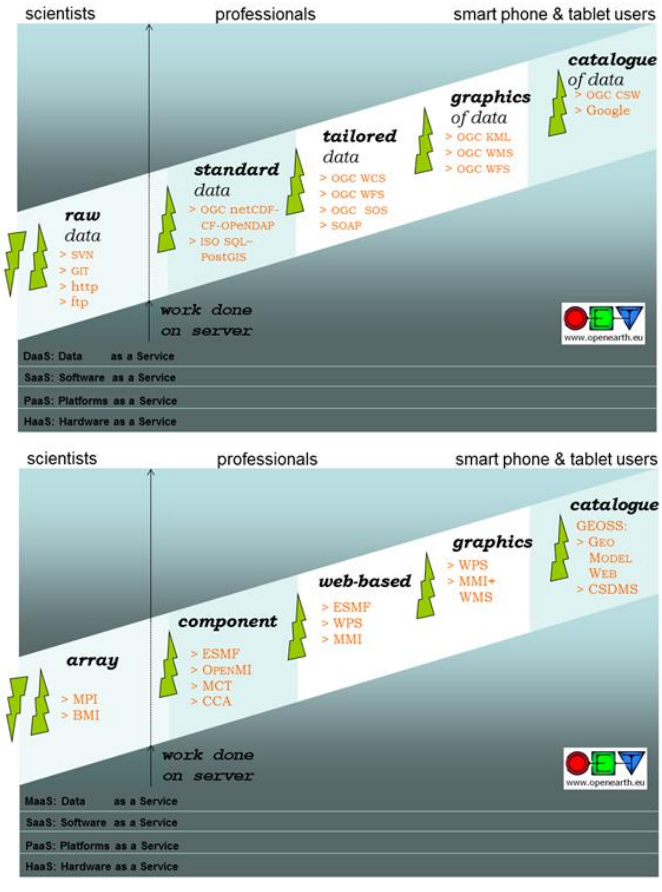

Figure 2. A hierarchical inventory of standards for web exchange of data (top) for different kinds of users, after Baart et al. (2012), and for different kinds of models (bottom).

[2] http://5stardata.info

System of systems

As more of these model services become available, the need arises for an overview of available systems. Such catalogues are in the upper right hand side of Figure 2. There are two main approaches for finding data. Some take the approach of a catalogue of manual or semi-automated generated metadata, like CSW and inspire. Others optimize their information to be crawled by automated collection of metadata, such as Google.

Several initiatives have started over the last few years to enable finding of web enabled models too. Some take the system of system approach, for example the Geo Model Web (Nativi2013), within the broader GEOSS initiative[3], where to idea is to find systems (following the publish, find, bind approach). Others aim to provide a web based infrastructure that also provides a computation environment PaaS CSDMS. An interesting research question is how such models can be found with a posteriori search engines.

CASE STUDIES

To illustrate the use of web-standards for coupling DaaS, MaaS and tailored information for end-users via TaaS, this section shows case studies for each of these services. For generation of information the workflow follows Eq. 1, where the web-interface can be at any $\Leftrightarrow$ such that server side processing is to the left of it, and client side processing to the right of it.

$$\textit{data} \quad \Leftrightarrow_1 \quad \textit{models} \quad \Leftrightarrow_2 \quad \textit{tools} \quad \Leftrightarrow_3 \quad \textit{information} \qquad (1)$$

Data

Making data available via a web interface ($\Leftrightarrow_1$ in Eq. 1) has nowadays become a required step for many data types to be used at all. For instance, many data are simply too big to be transferred to a client as a whole. OGC offers many standards for making user-defined subsets of such data available via web services (Percival, 2010) [14]. However, in some cases web services for data are insufficient. (i) The latest generation of satellites (e.g. Sentinel) produces so many terabytes that even subsets are too big. (ii) When data is stored in proprietary formats, a server with the proper license is the only way to subset the data (e.g. Oracle). (iii) Data can also be too politically/juridically complex to be allowed for web dissemination, even if it is small. For instance, the pan-European SeaDataNet does not allow users to download subset via web services, not even for open data. (iv) Data can also be too sensitive as in the EU project CIPRNet[4]. Data on hazards and vulnerabilities can fall into the hands of terrorists. Due to this risk, it is not allowed to obtain fine grained subsets of the data. For working with such big, proprietary, political or sensitive data, the web-interface has to be shifted from DaaS to the tools side ($\Leftrightarrow_3$ in Eq. 1). In this case processes run at the server that delivers only aggregated results via the web, potentially also in graphical form.

Models

More recently, making models available via a web interface ($\Leftrightarrow_2$ in Eq. 1) has become a required step to for many models to be used at all. An example of exposing a model through the internet is given by Baart (2014). In the 3Di project[1] the topography with a resolution of 0.5m x 0.5m was used to model the flooding of the vulnerable parts of the Netherlands[5]. In this

[3] http://www.earthobservations.org
[4] http://www.ciprnet.eu
[5] http://www.ahn.nl

example the challenge was the sheer amount of data. The model generates results at 10Hz for a 200Million pixel map, corresponding to an output of 8GB/s. This model is connected to an interactive web application and real time radar data using a combination of techniques. For the data exchange a combination of existing and new coupling methods were used. The rain radar data was connected using OPeNDAP and an extension to the in-memory netCDF. The model is run using the BMI protocol for the library interface and the Model Message Interface (MMI) is used to communicate model results over ZMQ and websockets. These model messages are fed to the WMS and directly to the browser. This shows that for advanced applications a single coupling method is not sufficient.

Tools

In the previous examples the tools to process data or models are placed on the server next to the data and models they operate on. This works well for big, proprietary or secret data/models. But it is also possible to make such tools available via a web interface ($\Leftrightarrow_3$ in Eq. 1) as stand-alone process. In fact, this is also becoming a required step for many complex processes to be used at all. For example, most of the tidal analysis packages used in the world have some complications preventing wider use. Classical tidal analysis packages are Fortran executables which require outdated ascii files as input, making them complex to master. Well-documented information has to be downgraded to fit these programs. And *t_tide*, a very widely used code in Matlab (Pawlowicz, 2002) [15], only works when a license of Matlab has been purchased. Offering these operations to non-Matlab users as web-service via a WPS wrapper is a solution (Boerboom, 2013) [16].

Information

Web services have been used for a comparatively long time to provide information ($\Leftrightarrow_3$ in Eq. 1) for the public. An example are flood hazard maps for inhabitants of flood prone cities as a means of raising awareness and preparedness. Flood hazard maps are in most cases generated with the help of modeling and represent a typical result of the process described with Eq. 1. This information must be provided in such a way that it is useful for non-experts. An example of such a web service is given with the flood hazard maps for the city of Cologne[6]. Here the flood hazard map is made available via web services, while previous data and modelling steps have been carried with non-web based techniques.

A simple example of the need of tailoring for such displays is the portrayal of statistical (flood risk) information on maps. The order of magnitude of probabilities ranges so much that plotting on logarithmic scales plus proper tuning of the transformed color ranges is essential. The color range is not part of the OGC WMS standard, and there is also no default layer for (log-)transformed data. Display of probability information from a coupled DaaS/MaaS web therefore requires an extra interface. We refer to this interface as the TaaS. WPS can provide such an interface, as a wrapper for the existing DaaS/Maas services as in Figure 2ab.

Such ways of presenting sensitive information on the web requires a fine-tuning to ensure proper information transfer to the end-users. This becomes even more important when the information is presented in Decision Support Systems (DSS) for professional decision makers. Here improper or untimely interpretation of information can have large consequences. An example is the rigid requirements for portrayal of Electronic Nautical Charts (IHO, 2000) [17] on ships which prevents use of common WebGIS techniques. A TaaS layer would be able to meet such strict requirements to tailoring of information.

[6] http://www.hw-karten.de/koeln/

DISCUSSION AND CONCLUSIONS

The first goal of this paper was to extend previous model coupling overviews with a discussion of the web coupling suitability. We discussed that some of the model coupling techniques have been used for web based communication (ESMF, OpenMI), while others lag behind on this aspect (MCT, CCA). New techniques (MMI), are under development that make use of developments in internet based communication (websockets, message queuing).

We found that the coupling approaches can be best considered in a hierarchy. We separate between the DaaS/MaaS and TaaS layers. One could argue if the conceptual DaaS and TaaS concepts are perhaps too generic. Is interacting with a model through one of the coupling methods not data as well? Is a model not also a tool, or is a tool not a simple model? The separation into these distinct concepts gives a focus though, both in their potential user community and in their domain. DaaS compares to philosophically inductive information, MaaS compares to philosophically deductive information, whereas TaaS offers a hybrid coupling of Daas, MaaS and TaaS. In addition, the standards used in DaaS, MaaS and Taas differ. This could lead us to reconsider whether the current trend of using WPS as a model coupling tool standard (for example Maué 2011 [19]) is the best approach. The use of WPS in a tool layer has advantages: it supports any return format, whether a picture or a pdf report or a netCDF file, it has some guidance for defining inputs in the sense that it distinguishes geospatial input from other inputs. For model coupling, where the general consensus is that the minimal interface to implement is the *Initialize*, *Run* and *Finalize*, WPS is not suitable though as it does not support an *Initialization* interface. WPS does have support for status updates, but only one-way, not on request. Furthermore, WPS does not allow stopping a started process once started, which is a major drawback for heavy processes being one of the primary motivations for migrating to server-hosted tools in the first place. So although WPS is the latest standard for model coupling and has distinct advantages such as web-coupling and tailoring of information products, it does not serve as a substitute for model coupling standards discussed in Jagers (2010) and Nativi (2013). Dedicated solution such as MMI with WMS push systems (Baart, 2013) do offer substantial improvement. WPS does offer a solution for the increased need to tailor presentation of data to specialist and non-specialist end-users. Hence is fits on top of DaaS/MaaS layers. At the final end of our proposed MaaS standards hierarchy we saw that there are initiatives to support a system of systems, allowing to get an overview of "which models can tell me something here?".

While the data as a service is becoming mature already, the model and tools as a service layers are just getting started. We will see great progress in this area in the coming years. It is likely that we will see a lot of interactive models that can be intervened with through the web. Also the operational models will start to give live updates while they're running. Enabling this requires to solve several technological challenges, a combined effort of the community and in the end, standardization.

ACKNOWLEDGEMENTS

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 312450. The European Commission's support is gratefully acknowledged. The paper has been further written with financial support and in kind contributions from the following projects: 3Di[1], Rijkswaterstaat NGHS, EU EMODnet chemistry (DG MARE), EU ECOstress (DG ECHO).

REFERENCES

[1] Tetlock, P.E., "*Expert political judgment: How good is it? How can we know?*" Princeton University Press, (2005).

[2] Geller, G. and Turner, W., "The model web: a concept for ecological forecasting", *Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International*, (2007) pp 2469–2472.

[3] Roman, D., Schade, S., Berre, A., Bodsberg, N. R., and Langlois, J., "Model as a Service" *AGILE workshop: Grid Technologies for Geospatial Applications*, Hannover, Germany, (2009).

[4] Jagers, H., "Linking data, models and tools: an overview" *Proceedings of the International Congress on Environmental Modelling and Software*, Ottawa, Canada, (2010).

[5] Nativi, S., Mazzetti, P., and Geller, G. N., "Environmental model access and interoperability: The GEO model web initiative", *Environmental Modelling Software*, 39(0), pp 214 – 228.

[6] Larson, J., Jacob, R., and Ong, E., "The model coupling toolkit: A new fortran90 toolkit for building multiphysics parallel coupled models", *International Journal of High Performance Computing Applications*, 19(3), (2005), pp 277–292.

[7] Hill, C., DeLuca, C., Balaji, Suarez, M., and Da Silva, A., "The architecture of the earth system modeling framework", *Computing in Science Engineering*, 6(1), (2004), pp 18 – 28.

[8] Saint, K. and Murphy, S., "End-to-end workflows for coupled climate and hydrological modeling", *International Congress on Environmental Modelling and Software*, (2010).

[9] Gregersen, J. B., Gijsbers, P. J. A., and Westen, S. J. P., "OpenMI: Open modelling interface", *Journal of hydroinformatics*, 9(3), (2007), pp 175–191.

[10] Goodall, J., Robinson, B., and Castronova, A., "Modeling water resource systems using a service-oriented computing paradigm", *Env. Model. Software*, 26(5), (2011), pp 573 – 582.

[11] Baart, F., "Interactive web-based flood modeling at country wide scale and planter size resolution", *Proc. Int. congres of Environmental Modelling & Software*, San Diego, (2014).

[12] Baart, F., de Boer, G., de Haas, W., Donchyts, G., Philippart, M., van Koningsveld, M., & Plieger, M., "A comparison between WCS and OPeNDAP for making model results and data products available through the internet", *Transactions in GIS*, 162(2), (2012). pp 249–265.

[13] Day, J., & Zimmermann, H., "The OSI reference model", *Proceedings of the IEEE*, 71(12), (1983).

[14] Percivall, G., "The application of open standards to enhance the interoperability of geoscience information", *Int. Journal of Digital Earth*, 3(sup1), (2010), pp 14–30.

[15] Pawlowicz, B. Beardsley, and S. Lentz, "Classical tidal harmonic analysis including error estimates in matlab using t_tide", *Computers & Geosciences*, Vol. 28, (2002), pp 929-937

[16] Boerboom, J., "Implementing the WPS Standard", GIMA MSc thesis, (2013). Retrieved from http:// https://publicwiki.deltares.nl/display/OET/wps/

[17] International hydrographic bureau, "S-52 - specifications for chart content and display aspects of ECDIS", (2010). http://www.iho.int/iho_pubs/standard/S-52/S-52_e6.0_EN.pdf

[18] Maué, P., Stasch, C., Athanasopoulos, G. and Gerharz, L., "Geospatial standards for web-enabled environmental models", International Journal of Spatial Data Infrastructures Research (IJSDIR) Vol. 6, (2011), pp 145—167.