

# Design of a Software Framework Based on Geospatial Standards to Facilitate Environmental Modelling Workflows

Vanessa Watson<sup>a</sup>, Kym Watson<sup>b</sup>

<sup>a</sup>*Fraunhofer Application Center System Technology AST, Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Am Vogelherd 50, 98693 Ilmenau, Germany (vanessa.watson@iosb-ast.fraunhofer.de)*

<sup>b</sup>*Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Fraunhoferstr. 1, 76131 Karlsruhe, Germany (kym.watson@iosb.fraunhofer.de)*

**Abstract:** Environmental modelling workflows are often tedious due to heterogeneity of data sources, types and formats. This paper discusses how a software framework based on Open Geospatial Consortium (OGC) standards may facilitate the workflow steps and improve their flexibility by supporting data access, processing, and decision-making. The software framework is designed based on OGC standards, which can be realized in any modern programming language, to encapsulate processing algorithms (so called fusors) that fuse the environmental input data and run a model based computation. Such a software framework can improve the integration of new data sources into environmental applications as well as the data exchange between scientists who can then focus on the real application problem at hand. The open source model MODFLOW, which is worldwide one of the standard codes for groundwater flow simulation, is taken here as an example data fusor. A further objective of this software framework is to offer the complete modelling workflow as a web service.

**Keywords:** environmental modelling; open geospatial standards; MODFLOW; fusion

## 1 INTRODUCTION

Environmental modelling in particular and geosciences in general rely on data from many different sources: e.g. geological profiles, digital elevation models, meteorological data, hydrological data on catchments such as groundwater levels and river discharge, land use data, just to mention a few. A scientist is confronted with the enormous task of accessing the data with different protocols, and pre-processing the data for the task at hand. Pre-processing includes conversion of the data format, units and coordinates, the possible removal or replacement of outliers and exceptions (like missing values), adapting temporal or spatial scales and assembling all input data into data structures amenable for further processing. Only after this step can the domain specific modelling and analysis begin. The scientist normally wants to make the modelling and analysis results (e.g. tables, graphs) available to other scientists for further downstream studies. Currently the steps in this workflow are normally conducted with many customized software tools at great effort with a definite lack of interoperability and adaptability for re-use.

These steps are often tedious due to the heterogeneity of the data types and formats. They are usually setup for a particular scenario with the aid of scripts written in languages

such as MATLAB, R or Python. The model itself is a software module with specific input and output data sets as well as parameters. This approach is difficult to sustain when new data sources or models are to be incorporated.

This general modelling scenario can be very well illustrated for a case study using the USGS open source groundwater flow model MODFLOW. Groundwater flow modelling is an ideal example to test the implementation of a software framework based on geospatial standards to facilitate the work flow since it has become a standard tool in environmental sciences, comprises a sequence of different work flow steps and requires a variety of data sources and formats [Anderson and Woessner, 2002; DGG, 2002]. Defining the purpose of the groundwater flow model and selecting data sources, for example from remote or in-situ sensors, or databases are the first steps in the groundwater modeling process. Furthermore, the collected data in general needs to be processed (e.g. data cleaning, format conversion and spatial interpolation of point information). Based on the geological and hydrological conditions and the dominant processes of the groundwater flow system a model based algorithm is selected. After calibrating and validating the selected model with different sets of field data, the model is executed using additional parameters, e.g. to control convergence. In the next steps, model results need to be visualized, stored for possible further processing and compared with the actual field situation to assess if the model needs to be adapted. For the principles of groundwater flow modelling see the hydrological guide of the German Geological Society DGG [2002] and Anderson and Woessner [2002].

The paper discusses how open standards may be used to facilitate the workflow steps and improve their flexibility. A further topic is to offer the complete workflow as a web service to allow decision makers to apply it with limited options and only a browser installed. The main standards considered below are those of the Open Geospatial Consortium: Observation & Measurement Model, SensorML and the Sensor Web Enablement (SWE) family, together with UncertML to describe uncertainty. The objective is to define a flexible software framework to access and process environmental data with integrated environmental models.

The standards apply not only to the original sensor data, but also to data derived from any process of observation and measurement, including a model based algorithm. The sensors may be in-situ, remote (such as a satellite or other image based system), or ex-situ (based on sampling and laboratory analysis). In current practice, the web services typically access a database of processed sensor data rather than the actual sensors. However, there is a strong trend to make more and more sensor data as well as modelling and processing services accessible over the web as the platform for diverse interoperable applications using standardized interfaces. Major international initiatives promoting these developments are INSPIRE<sup>1</sup>, GEOSS<sup>2</sup> and the Future Internet, as explained in Havlik et al. [2011]. INSPIRE is Directive 2007/2/EC of the European Parliament and of the Council establishing an *Infrastructure for Spatial Information in the European Community*. INSPIRE mandates normative Implementing Rules and informative guidelines to ensure the accessibility and interoperability of spatial data sets and services. GEOSS (Global Earth Observation System of Systems) is an intergovernmental programme, coordinated by the Group on Earth Observations. GEOSS is a 10 year global programme that aims to provide to the broad environmental science and user community support tools for the monitoring, analysis and modelling of various environmental phenomena in a range of societal benefit areas.

---

<sup>1</sup><http://inspire.jrc.ec.europa.eu>

<sup>2</sup><http://www.earthobservations.org>

## 2 OVERVIEW OF OGC STANDARDS

The Open Geospatial Consortium (OGC) is a non-profit, international voluntary consensus standards organization that is leading the development of standards for geospatial and location based services (see Reichardt and Percivall [2011]). The standards are applied to a wide range of domains, including hydrology, geology, meteorology, oceanography, land cover, and many other scientific fields in geosciences. The primary objective of the standards is to support information discovery, access, processing, fusion and decision-making.

For an introductory overview of OGC standards and their application in the geosciences, the reader is encouraged to browse the on-line tutorial Percivall [2010a]. It gives an overview of the OGC web service standards Web Map Service (WMS), Web Feature Service (WFS), Web Coverage Service (WCS), and Catalog Service for the Web (CSW) as well as the Sensor Web Enablement (SWE) standards presented in more detail below. The OGC SWE standards allow sensor data to be accessed in a uniform way over the Internet with web services and XML encodings.

The SWE services of relevance to this paper are: (a) *Sensor Observation Service (SOS)* to access sensor observations with queries filtering on the phenomenon (property) and the spatial and temporal domains of the observations, (b) *Sensor Planning Service (SPS)* to parameterize and task (schedule and execute) assets such as sensors, sensor platforms (e.g. satellites), models or even persons (e.g. to conduct ex-situ measurements). The main operations of the SPS are DescribeTasking (to get the tasking parameters), GetFeasibility (to ascertain if the asset can be tasked with the given parameters) and Submit (to actually execute the task). During the execution the operations GetStatus and Cancel are available. In this paper we apply the SPS to models and the model result(s) become new observations for a SOS, i.e. *the model is considered to be sensor*.

The services are accompanied by two fundamental information models:

- Sensor Model Language (SensorML). It describes the static metadata of a sensor such as manufacturer, operator, measurement quantities and procedure and its accuracy, as well as the sensor position, if fixed. Note that a sensor data fusion or processing procedure with input / output variables and parameters can be described as a sensor.
- Observation & Measurement Model (O&M Model) is the core information model. It describes the relations between an observation, the associated feature of interest (what was observed) and the associated sensor procedure (how was the observation performed). The O&M Model can be extended to describe the uncertainty of observations in a syntax called UncertML.

It is important to realize that model results and more generally, the result of any data processing or fusion algorithm, can be viewed as a new observation whose associated sensor procedure is the applied processing or fusion algorithm. Thus the SOS and SPS work together in unison to enable processing chains in a homogeneous service framework. All SWE services and information models are specified with XML schemas that allow their realization in any modern programming environment with largely generic software modules, the so-called "SWE enablers". The specific application software can then build on these enablers to access sensor data and produce derived data sets with all necessary descriptive metadata. The objective is to facilitate the integration of new data sources into applications and improve data exchange between applications. Application

programmers can then focus on the real application problem at hand rather than on data collection and basic manipulation.

### 3 CASE STUDY MODFLOW

#### 3.1 Overview of MODFLOW

The open source software MODFLOW published by the U.S. Geological Survey is one of the standard three-dimensional groundwater flow models used worldwide. It has a modular program structure whereby simulation and hydrological processes are divided into packages, for example the River Package deals with the simulation of the effects of rivers, the Evapotranspiration Package simulates the transpiration of plants and direct evaporation from groundwater and the Strongly Implicit Procedure Packages solves the system of simultaneous finite-difference equations. The many different packages provided for MODFLOW allow a wide range of modelling applications in the area of water supply management, contaminant and heat transport and groundwater/surface water interactions.

MODFLOW solves the following three-dimension groundwater flow equation with the assumption of constant groundwater density using a block-centered finite-difference approach [Harbaugh, 2005; Domenico and Schwartz, 1997]:

$$\frac{\partial}{\partial x}(K_{xx} \frac{\partial h}{\partial x}) + \frac{\partial}{\partial y}(K_{yy} \frac{\partial h}{\partial y}) + \frac{\partial}{\partial z}(K_{zz} \frac{\partial h}{\partial z}) + W = S_s \frac{\partial h}{\partial t} \quad (1)$$

where  $K_{xx}$ ,  $K_{yy}$ ,  $K_{zz}$  [ $m/s$ ] are the values of hydraulic conductivity along the  $x$ ,  $y$  and  $z$  coordinate axes, assumed to be parallel to the major axes of the hydraulic conductivity tensor,  $h$  [ $m$ ] is the potentiometric head,  $W$  [ $s^{-1}$ ] is the volumetric flux per unit volume representing sources / sinks of water,  $S_s$  [ $m^{-1}$ ] is the specific storage of the porous material and  $t$  [ $s$ ] is time.

The groundwater flow equation, together with the specification of boundary conditions of the aquifer system (e.g. of type Neumann, Dirichlet, or Cauchy) and specification of the initial heads, constitute a mathematical representation of a groundwater flow system.

In MODFLOW an aquifer system is replaced by a 3D discretized domain consisting of cells in which the medium properties are assumed to be uniform. The size of the model grid is specified by the user in terms of the number of rows, columns and layers. Hydrostratigraphic units which are geological layers with similar hydraulic properties are represented by the model layers. Irregular model boundaries can be created by defining inactive (no flow) cells. For each cell a flow equation is written and hydraulic head calculated.

#### 3.2 Input Data, Parameters and Output Data

In order to run a MODFLOW simulation, model geometry, aquifer hydraulic properties, initial hydraulic head distribution and boundary conditions must be specified. The input data is read from separate external files, whereby the input system was designed for optimal flexibility. Tables 1 and 2 list the input data required for a transient MODFLOW model as a set of data workflows. Each workflow is described in a block of 3 table rows: information on the data source, processing and the MODFLOW input respectively. In a transient simulation, where head changes over time result in a change in storage of the

groundwater system, simulation time is divided into stress periods over which forcing (e.g. well abstraction, recharge, boundary flow) of the system is constant. Each stress period, in turn, is discretized into time steps. Depending on the complexity of the model, one has to add more input data to this basic table. For example if one wants to set up a model with pumping wells and a river, one has to specify location, depth and pumping rate of the wells in addition to river stage, river bottom elevation and river bed conductance. For further information about the latest MODFLOW version see Harbaugh [2005].

**Table 1.** Modelling workflows (1)

Data Source	Geological data: profiles of boreholes, vertical electrical or magnetic soundings
Processing	Grouping the different soil types of the borehole profiles into units with similar hydraulic properties
<b>MODFLOW input</b>	<b>Model geometry</b> with number of layers and number and width of the rows and columns. Aquifer/layer type (confined, semi-confined, unconfined), <b>Bottom elevation of a layer</b> (to calculate layer thickness), <b>Effective porosity</b>
Data Source	Geological data: profiles of boreholes, location of faults and impermeable rocks. Hydrological data: location of catchment boundary, groundwater level measurements
Processing	derive boundary conditions for the model domain based on the field situation (e.g. catchment boundary and impermeable rocks are no flow boundaries ⇒ Neumann boundary condition)
<b>MODFLOW input</b>	<b>Boundary conditions</b>
Data Source	Remote sensing data: Digital Elevation Model
Processing	Interpolation of elevation values to model cells
<b>MODFLOW input</b>	<b>Elevation of top layer cells</b>

In addition to the input data, parameters need to be specified which control the execution of MODFLOW. Depending on the groundwater flow problem a solver package which solves the equations of the simulation and a convergence criterion need to be selected.

MODFLOW allows the user to specify the amount, type and frequency of output data which is written to several different files in binary format. For example MODFLOW saves a detailed run record and simulation report with a summary of the input data of all packages. In addition, hydraulic heads after each time step and the drawdowns (difference between the initial and calculated hydraulic heads) can be written to output files. To check the quality of the numerical simulation (the continuity statement of groundwater flow has to be met), MODFLOW calculates a volumetric water balance budget for the entire model area and a percentage error term at the times requested by the model user. The overall ground water modelling workflow comprises the steps of Table 1 followed by the execution of MODFLOW itself, each of which can be tasked by a SPS with results available via an SOS. Tables 1 and 2 serve as an illustration of the typical steps required to process information from heterogeneous sources in order to generate the required input data for a downstream model.

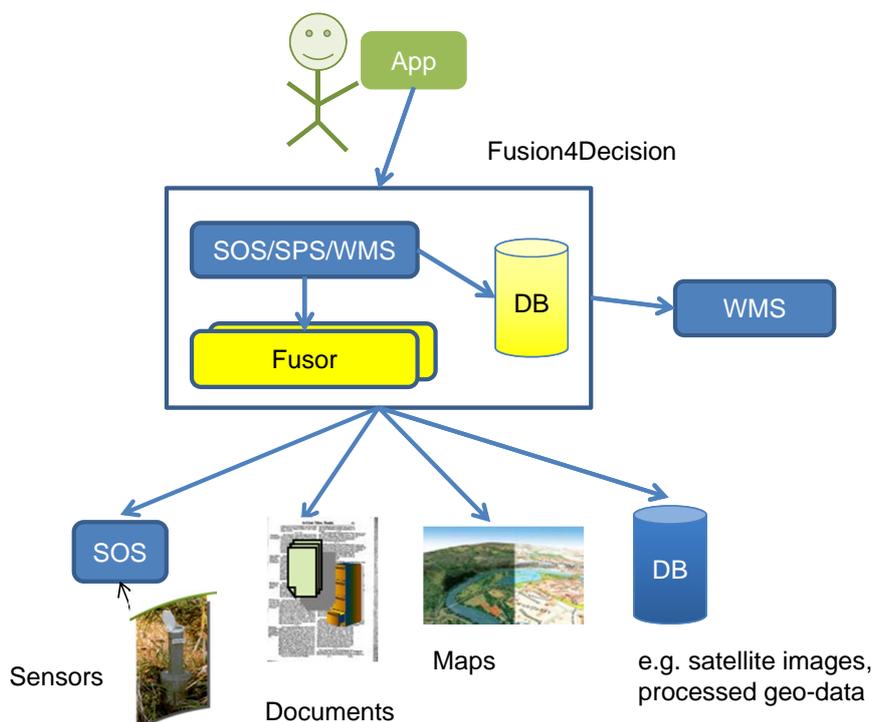
**Table 2.** Modelling workflows (2)

Data Source	Hydrogeological data: pumping tests
Processing	Apply analytical models to determine storage coefficient and hydraulic conductivity from draw down curves
<b>MODFLOW input</b>	<b>Storage coefficient / hydraulic conductivity for layers</b>
Data Source	Hydrological data: measured groundwater levels
Processing	Interpolation of measurements to model cells
<b>MODFLOW input</b>	<b>Initial hydraulic head</b> distribution in model area
Data Source	Remote sensing data: aerial photographs of model area and field observations of landuse types, meteorological data: air temperature, incoming and outgoing solar radiation, relative humidity, wind speed, sunshine hours, precipitation
Processing	derive landuse map from images and field observations, apply analytical/empirical equations to estimate evapotranspiration for each landuse type, interpolate measured precipitation, estimate groundwater recharge with water balance
<b>MODFLOW input</b>	<b>Groundwater recharge rate</b>
Data Source	depends on type of stress (e.g. extraction wells: pumping schedule with time periods and rates of pumping)
Processing	derive length, pumping rate and time steps of each stress period
<b>MODFLOW input</b>	<b>stress periods and time steps</b>

#### 4 SOFTWARE FRAMEWORK

The software framework *Fusion4Decision* illustrated in Fig. 1 with details in Table 3 uses the OGC web services SOS, SPS and WMS to encapsulate processing algorithms, the so-called *Fusors*. MODFLOW is the example Fusor considered here. The term Fusor has been coined to describe a general fusion or processing of input data, including through a model based computation. Another common fusor is the spatiotemporal interpolation of measurement data. The Fusor is in any software code that can be integrated into a Java environment, such as MATLAB, R, Python, and C variants. For further background information on applying OGC standards for fusion of geospatial information and decision making see Percivall [2010b] and the references therein.

The web service operations are based on XML files according to OGC schemas for the requests and responses. Helper software in Java has been written to parse and generate these XML files for the service profiles used in the framework. The Fusors can then be developed independently without detailed knowledge of XML and the OGC schemas. An important feature of *Fusion4Decision* is that a Fusor result is itself described as an Observation as in the O&M Model with all relevant metadata on how the result was obtained. As for MODFLOW, the Fusor results comprise a set of files (numeric, graphic) that are listed as links in the Observation result. These files can then be accessed as needed for downstream processing and visualization. Scientists and decision makers can then be given results customized for their respective purpose.



**Figure 1.** Software Framework Fusion4Decision based on OGC standards

**Table 3.** Usage of the SPS and SOS operations

SPS Interface	Function
GetCapabilities	lists the available Fusors
DescribeTasking	lists the configurable parameters of the Fusor described in SensorML
GetFeasibility	checks the validity of the parameters of the Fusor
Submit	starts execution of the Fusor
GetStatus	returns the execution progress
Cancel	aborts execution of the Fusor
DescribeResultAccess	yields the SOS Request for the Fusor result (SOS Observation)
SOS Interface	Function
GetCapabilities	lists the available Fusor results (derived Observations)
DescribeSensor	description of a parameterized Fusor as executed
RegisterSensor	registration of a new Fusor described in SensorML
GetObservation	query for observations fulfilling criteria
InsertObservation	insertion of new Fusor result
GetObservationById	IDs are managed by an underlying information management system
GetFeatureOfInterest	yields associated feature of the Fusor result (e.g. coordinate grid as a Sampling Feature)
GetFeatureOfInterestTime	time interval of data for the feature

*Fusion4Decision* includes an information management system to manage all files associated with the Fusors and generated Observations. This supports usage scenarios requiring provenance and reproducibility as well as general server entry housekeeping. The information management system used is WebGenesis<sup>®</sup> of Fraunhofer IOSB, but other solutions are possible depending on the overall functionality. The software framework *Fusion4Decision* has been developed in Java and provides the web service interfaces SOS, SPS and WMS within a service oriented architecture SOA. Table 3 shows how the SPS and SOS operations are used to run a Fusor.

## 5 CONCLUSIONS AND RECOMMENDATIONS

The use of the OGC standards such as SOS, SPS, WMS, SensorML and O&M, can facilitate the design of software frameworks to support the processing of environmental data and environmental modelling workflows. The main advantages are two fold: (a) the integration of data sources and the pre-processing of data can be done more systematically with less effort, and (b) a complete set of metadata is maintained throughout the workflow thus enabling downstream processing and re-use of results and models in other applications. The complete modelling workflow can be offered as a web service for scientists or decision makers. The vision for the future is that more and more environmental data sets and models will be equipped with standard interfaces. The international initiatives promoting this development will have to ensure compatibility and provide techniques to integrate existing data sets and models.

## REFERENCES

- Anderson, M. P. and W. W. Woessner. *Applied Groundwater flow modeling - Simulation of flow and advective transport*. Academic Press, 2002.
- DGG. *Hydrogeologische Modelle. Ein Leitfaden mit Fallbeispielen*. Deutsche Gesellschaft für Geowissenschaften, 2002.
- Domenico, P. A. and F. W. Schwartz. *Physical and Chemical Hydrogeology*. John Wiley, Chi, 2 edition, 1997.
- Harbaugh, A. W. *MODFLOW-2005, The U.S. Geological Survey modular ground-water model - the Ground-Water Flow Process*. U.S. Geological Survey Techniques and Methods 6-A16, 2005.
- Havlik, D., S. Schade, Z. A. Sabeur, P. Mazzetti, K. Watson, A. J. Berre, and J. L. Mon. From Sensor to Observation Web with Environmental Enablers in the Future Internet, <http://www.mdpi.com/1424-8220/11/4/3874/>, doi: 10.3390/s110403874. *Sensors*, 11(4):3874–3907, 2011.
- Percivall, G. OGC and the Geosciences - IGARSS 2010 Tutorial, <http://www.ogcnetwork.net/GeosciencesTutorial>, July 2010a.
- Percivall, G. OGC Fusion Standards Study Engineering Report, OGC 09-138, [http://portal.opengeospatial.org/files/?artifact\\_id=36177&version=2](http://portal.opengeospatial.org/files/?artifact_id=36177&version=2), March 2010b.
- Reichardt, M. and G. Percivall. OGC Activities to Bridge Research to Practice, [https://portal.opengeospatial.org/files/?artifact\\_id=47671](https://portal.opengeospatial.org/files/?artifact_id=47671), April 2011.