Wosniok, Christoph; Lehfeldt, Rainer

A Metadata Profile for Numerical Modeling Systems

A METADATA PROFILE FOR NUMERICAL MODELING SYSTEMS

Christoph Wosniok[1] and Rainer Lehfeldt[2]

ABSTRACT

Sharing the details of model configurations for selected simulation scenarios is complex and often vague. Due to the long list of relevant parameters involved and the specification of input and output files, a sustainable documentation of a numerical study includes the description of the used numerical engine, of the applied boundary conditions and of the simulation results with standardized metadata. These data files can be included in Spatial Data Infrastructures and therefore be disseminated in an interoperable way. In this paper, we present an ISO metadata profile for computational modeling, which is able to depict both the numerical core and the configuration settings of a modeling system.

1. MOTIVATION

Metadata are the essential vehicle for transporting information through Spatial Data Infrastructures (SDIs). Describing and searching for data are generally executed on metadata, which finally link on the actual data sources, usually in form of standardized and therefore interoperable web services. The ISO 19115: Geographic Information – Metadata (2003) provides the generic structure for metadata in major national and international SDIs, where the about one dozen ISO core elements build the ultimate backbone for all kinds of metadata. Domain-specific metadata profiles extend this ISO core with individually required elements.

During the uprising of SDIs in the past years, numerical models have mostly been a side topic in SDIs due the complexity of models and model runs. A comprehensive description of numerical models comprises several data types, most of which are categorized as common geo-data files: data containing some kind of geographic localization. These are input and output files, boundary conditions or the grid on which a model could run. The numerical modeling core processes these files as part of the input parameters during the execution of a model run. Additionally, there are several parameters concerning the simulating behavior and the execution of the software of the actual calculation unit of a model. These parameters are not part of metadata profiles generally used to describe geo-data. (Wosniok and Lehfeldt 2012)

However, to reproduce a model run and consequently its simulation result requires a standardized description of these detailed parameters. Including them next to the additional files in a SDI benefits both information sharing and reproducibility of simulation results for legal and scientific purposes. Decisions often more and more rely on modeled data as for example proposed by the European Water Framework Directive (European Parliament 2000).

---

[1] Research Assistant, German Coastal Engineering Research Council, Federal Waterways Engineering and Research Institute, Hamburg, Germany (christoph.wosniok@baw.de)
[2] Senior Researcher, German Coastal Engineering Research Council, Federal Waterways Engineering and Research Institute, Hamburg, Germany (rainer.lehfeldt@baw.de)

It also follows the idea of the Model Web (Geller and Turner 2007), which is a concept for a network of connected and interacting models. Several elements of such a network are subject of ongoing projects in the Seventh Framework Programme of the European Union; however, the set up of the metadata profile is not among them.

2. RELEVANT PREVIOUS WORK

The theory of model capabilities and its descriptions have been investigated since 1967 by Chorley and Haggett (1967). Besides the actual software handling, documentations of modeling software generally provide detailed descriptions of physical modeling capabilities like used methods, equations and generalizations. However, there was no commonly used pattern for these descriptions. In 1994, the International Association for Hydraulic Research published "Guidelines for Documenting the Validity of Computational Modelling Software" (IAHR 1994) providing a structure to document methods and capabilities for the semantic validation of numeric modeling systems. Accordingly, documents showing the technical integrity have been published for modeling systems such as UnTRIM, Telemac 2D or Delft-3D-Flow. The guidelines provide an outline for mandatory chapters, but the actually delivered validation documents are composed of much unstructured text within the chapters containing too much aggregated and unstructured information. This loose structure proves the use of validation documents as foundation for interoperable information exchange as insufficient.

The nowadays dominating XML format was officially release by the W3C in 1998[3] and serves as the basic exchange format for all kind of data not only in SDIs. It is also used for the reference implementation of the ISO 19115. In its latest version of 2010, the standard specifies nearly 450 possible elements to describe geographic data. This broad choice enables very specific descriptions for the specific domains by developing metadata profiles, as it has been done in for example, the GeoSeaPortal of the German Federal Maritime and Hydrographic Agency (BSH) (Soetje 2008), the European portal SeaDataNet[4] or the Marine Data Infrastructure Germany (MDI-DE) (Lehfeldt et al. 2011).

Models have been considered at some points in the ISO 19115. For example the categorizing element hierarchyLevel (element #6) allows to specify "model" or *presentationForm* (element #368) offers "modelDigital", defined as a "multi-dimensional digital representation of a feature, process etc". However, due to the generic nature of the ISO further specifications as for example input and output files can not be found.

Hill et al. (2000) developed a Content Standard for Computational Models (CSCM) defining specific metadata elements for modeling, for instance, input- and output data or processing and hardware requirements. The CSCM provides a structure to describe complete modeling scenarios, including details on single datasets and values of used parameters. However, it has, similar to the IAHR validation documents, the shortcoming of most descriptions being given in unstructured text.

The CSCM followed closely the development of the standard ISO 19115 "Geographic Information – Metadata". Therefore, most of the general metadata elements in the CSCM have a close match in the ISO. It was possible to map CSCM elements on ISO 19115:2010 Core elements for the *hierarchyLevel* "model" (Wosniok and Lehfeldt 2012); therefore the interoperability for using metadata elements from the CSCM in an SDI is given. However, the CSCM still has a large number of free text fields, leaving large portions of relevant details unstructured.

[3] http://www.w3.org/XML/
[4] http://www.seadatanet.org/Standards-Software/Metadata-formats

3. USE CASE

To clarify our intentions of providing an extensive metadata profile for numerical modelling systems and its model runs, we applied a use case from an ongoing research project. Along an example of a model run with the aim to detect morphodynamic changes in the German Bight, we identified elements relevant to be described in structured metadata. The AufMod Project, a project to exploring the sediment flux in the German Bight using a multi-model approach (Kösters et al. 2010). We fitted the identified relevant metadata elements of the project model run into the structures of an adapted CSCM. An excerpt from the metadata for our use case is depicted in table 1.

Large portions of metadata for numerical modelling systems can be taken from the ISO 19115. The generic standard offers sufficient options and compound elements as for example for spatial and temporal descriptions. Relevant metadata may comprise:

- An underlying digital elevation model or grid defining the resolution of the model.
- Input parameters steering the model. For a morphodynamic model this could be water level, current, salinity, suspended particle matter, waves, sediment or climate parameters like wind or air pressure. It should be possible to define start values for each parameter on each node of the grid. Datasets can be defined in standard metadata sets, but have to fit to the grid.
- Boundary values need to be defined at the edges of the grid. They can vary over the regarded modeling time span.
- Physical parameters crucially influence the simulation results of a model run, for example, in case of the morphodynamic model, the bottom friction. These parameters can be defined by constants, algebraic equations or individual models which can be switched on or off. Each of those options allow different levels of approximation, thus need to be defined unambiguously.
- Finally, the operator of a model can refine modeling properties further defining the detail and performance of a model, as for example the number of vertical layers in the water column or the number of fractions of sediment and bedload.

The grid usually comes in an external file, start and boundary values as well as single parameters for further processing can be provided in separate files. This depends on how parameters should be represented in the model, regional differences in parameter values are usually added in an own file, global constants do not require this. Within the Federal Waterways Engineering and Research Institute, a custom format to define values on regions is used. This format enables to set parameter values based on a defined parameter list. Parameters for start and boundary values are often defined in constants or in pre-processed time series.

Driving model parameters are set as constants, functions or, again, models. There are for example models for friction, turbulence, settling velocity, some of those nested models are model system specific modules as for example Dredgesim for UnTRIM, which enables to add dredging measures in the modeling process.

Table 1 Model run specific metadata for the use case.

	Contents	Example Values
Grid	Grid file	
Start Values	Water Level	0 m
	Current	0 m/s
	Salinity	0.35%
	Suspended Particle Matter	0 mg/l
	Waves	Hs = Tp = 0
	Climatology	Submodel
	Sediments:	
	- Grain Size Distribution	Function
	- Porosity	40%
	- Sediment inventory / stratigraphy	Area data set
Boundary Values	Water Level	20m
	Current	0 m/s
	Salinity	0.35%
	Waves	Fetch length/ none
	Wind	Velocity/ Direction
	Air pressure	Submodel/ none
	Fresh water inflow (in combination with:) Sediments concentration/ mass flow	Location-time-variant 1 kg/m³
Driving Model parameters/ Model capabilities	Grain Size Distribution/ Number of fractions	Function/ 1
	Transport Mode	Suspended/ bed load
	Sediment Layers (stratigraphy and stratification)	Layer thickness, concentration
	Bedload calculation	Van Rijn 1994
	Layers of Waterbody	50
	Number of SPM and bed load fractions	5/5
	Friction model	Global Constant
	Settling Velocity	Stokes (constant)
	Morphologic acceleration	Off
	Turbulence model	k-e (Rodi)
	Coriolis	Beta-Plane
	Exchange layer	5dmaxglobal (15cm)
	Modules (e.g. Dredgesim)	Off

In terms of metadata, separate files are considered as metadata sets valid on its own. They are handled as normal geo-metadata sets. This includes information on names, origins, different versions, producers, spatial- and temporal validity, geometry and so on. However, there is always the danger of inconsistencies to the parameters described in the metadata of the model run.

4. MAPPING

In order to arrange the identified metadata from our use can into the CSCM we have to examine the capabilities of the standard from the year 2000. In this section, we describe the adaptions made to the CSCM. We focus on the model specific metadata elements and omit general elements covered by the ISO. We list examples values for the start value of water level (table 2) and bed load (table 3). Although we need to make adaptions for our purposes and due to modifications of the ISO 19115 since the year 2000, we aim to preserve previously conducted work on the CSCM.

The first four and the 10th CSCM section can easily be mapped on the ISO as those contain general information on description, use, geographic and temporal extent which are covered under the ISOs' *MD_Identification* compound element (#23). We suggest applying the ISO elements therein which offer similar, and often more extensive, description options than the CSCM does. Elements concerning the numerical core and scenario descriptions are covered in sections 5 to 9. Information described here can to the greatest extend not be found in the ISO and is therefore be used as additional elements completing our ISO profile. We therefore only examine elements from these sections in the following.

4.1 System Requirements

In section 5, general system requirements for hardware and software are requested. These information are certainly important, however, in terms of describing and searching model runs, linking to software documentation covering these topics should be sufficient. In order to keep the metadata profile open for various models, the section 5.1 *Expertise Required* remains, although in our case there is always special expertise needed to operate the model.

4.2 Input Data Requirements

Summarized, section 6 provides input parameters with its relation to the rest of the model run. To keep the description clear and atomic, each start or boundary value requires a complete description of section 6.

The single input data files are referenced in element number 98 *Data Input File*. The URL should point to another metadata set, possibly within the same SDI the model run is described in. In the following elements of this section, the single parameters are described in detail similarly to the description of a complete geo-metadata set. A large issue is the possible redundant description of a dataset both in the linked metadata set and the description given for the model run metadata set, which easily leads to inconsistencies between both descriptions. However, the structure given by the CSCM tries to avoid this by setting flags as a conditional for further descriptions. If an external metadata set is given in number 98, parts of this section can be skipped during metadata filling. Unfortunately, there is no guarantee that the linked metadata set actually contains all the possibly mandatory elements listed in the CSCM. In fact, the ISO 19115 core compliance is the only guarantee for the linked metadata in common SDIs. Therefore automatic tools are needed for a comparison and completing the fields with the values of the linked metadata set, for example via the OGC Catalogue Web Service (CS-W) (Voges and Senkler 2007).

We propose to completely replace subsection 6.1. *Input Data Extent and Resolution* with the ISOs' compound element *EX_Extent* (#334). It contains geographical descriptions like a bounding box, vertical and temporal descriptions more broadly than the structure given by the CSCM. To cover possible overlaps to linked metadata sets, this compound element has only to be completed if there is no external metadata set.

Section 6.2 describes the role of the parameter; the CSCM uses the more generic term construct. Here, the parameters are specified, as the usage in the model could differ from the description in the linked metadata set. We extended codelist 7 by adding “Function – 05” and “Other

Model – 06" to the list. We also added an element called *Input Data Classification* to differentiate start values, the underlying grid and boundary values.

Section 6.3 again depends on the availability of a linked metadata set, if this is not available, the link to the actual data (#121) should be contained in therein.

Table 2 Input Data Requirements for a water level input data set. Table is based on CSCM specification, version 1.2 (Hill et al. 2000). Adaptions are marked bold. For the purpose of a clear depiction, we completed more values than necessary.

CSCM No	Element name	Definition	Obligation	Domain	**Example**
97	Input Data Extent and Resolution	Temporal and spatial extent and resolution for which the model was designed.	O	sec 6.1, lines 101-107	***See linked metadata set***
98	Input Data File	URL address to an external file containing description of the data input requirements in detail.	O	free text	**http://mdi-dienste.baw.de/.../recordId=5f26a5be-219f-4d55-b965-c1a868a1b792**
99	Input Modeling Construct Description	Parameter and variable constructs of the model.	C- Is there no reference to an external file containing this information provided in line 98?	sec 6.2, lines 108-119	
100	Input Dataset Description	A description of a dataset required in the processing of the model.	C- Is there no reference to an external file containing this information provided in line 98?	sec 6.3, lines 120-124	
	Sec. 6.1: Input Data Extent and Resolution (Optional, Non-Repeatable)				
-	EX_Extend from ISO	information about horizontal, vertical, and temporal extent	O		
	Sec. 6.2: Input Modeling Construct Description (Conditional, Repeatable)				
108	Name of Construct	Name assigned by the model or modeler to the specified model construct.	M	free text	**Water level**
109	Construct Classification	Functional properties of the specified construct.	M	**code list 7**	**04 – fixed parameter**

110	Construct Description	Description of the specified dataset.	M	free text	**Water level values for all elements of the model grid**
111	Construct Input Source	Method in which the construct is introduced to the model.	M	(fixed model setting, dataset member, user input)	**Dataset member**
	Input Data Classification	**Purpose for the model run**	**M**	**New Code List: grid, Boundary data, start values**	**Start values**
112	Dataset	Dataset in which the construct is found.	C- was "dataset member" selected as "Construct Input Source" (line 111)?	dataset name selected in line 120	
113	Construct Type	Data type of the construct.	M	free text	**ASCII**
114	Construct Units	Standard of measurement of given construct. (feet, meters, coded values, etc.)	C - Is the construct represented by units of measure?	free text	**meter**
115	Minimum Value	Minimum value accepted for processing in the model.	O	free real	**-500**
116	Maximum Value	Maximum value accepted for processing in the model.	O	free real	**10**
117	Default Values	The default value(s) assigned by the modeling software and/or modeler.	C - Does the model come with default value(s) for the parameter?	free text	
118	Construct Repeatability	Indication of how many times this construct occurs in the input. (Zero signifies an "optional" construct)	M	0 to N	**1**
119	Construct Comments	Any additional comments required to describe the particular input construct	O	free text	
	Sec. 6.3: Input Dataset Description (Conditional, Repeatable)				

120	Name	Name assigned by the model or modeler to the specified input dataset.	M	free text	**Water level**
121	Input Dataset File	URL address to an external file containing descriptions for the particular dataset.	O	free text	**ftp://baw-data/waterlevel.xyz**
122	Conceptual Data Structure	A textual description expounding on the concept of the required dataset.	C- Is there no reference to an external file containing this information provided in line 121?	free text	
123	Computational Representation	The physical data structure of the dataset required for the model.	C- Is there no reference to an external file containing this information provided in line 121?	free text	
124	Dataset Repeatability	Indication if more than one of these datasets may be provided. (Zero signifies an "optional" dataset.)	M	0 to N	**1**

4.3 Data Processing

The data processing section of the CSCM consists only of three elements, where only the programming language is mandatory. However, this is the place to go into details of the numerical modeling core and what is depicted in table 1 under “Driving Model parameters”. We propose a structure similar to the input data in the previous section. It has to be designed very broad, as the parameters have varying values from simple numerical values, as for example the number of sediment layers, up to coupled models itself, as the friction or the velocity. Therefore, a large number of elements is marked as optional.

Table 3 Data Processing metadata structure.

No	Element name	Definition	Obligation	Domain	Example
200	Processing Parameter Extent and Resolution	Temporal and spatial extent and resolution of the parameter.	C- Is the parameter represented in a file containing a spatial or temporal extend?	sec 7.1.	*See linked metadata set*
201	Processing Parameter metadata File	URL address to an external file containing description of the data input requirements in detail.	O	free text	http://mdi-dienste.baw.de/.../recordId=1234abcd-5678-efgh-b965-c1a868a1b792
202	Processing Parameter Description	Parameter construction of the model.	M	sec 7.2	-> 7.2
	Sec. 7.1: Processing Parameter Extent and Resolution (Optional, Non-Repeatable)				
203	EX_Extend from ISO	information about horizontal, vertical, and temporal extend	O		
	Sec. 7.2: Processing Parameter Description (Conditional, Repeatable)				
204	Name of Parameter	Name assigned by the model or modeler to the specified model Parameter.	M	free text	Bed load transport
205	Parameter Classification	Functional properties of the specified Parameter.	M	code list 7	07 – other model
206	Parameter Description	Description of the specified dataset.	C- if there is no Processing Parameter metadata file	free text	Van Rijn 1994
207	Parameter Input Source	Method in which the parameter is introduced to the model.	M	(fixed model setting, dataset member, user input)	Fixed model setting
208	Parameter Type	Data type of the parameter.	O	free text	-
209	Parameter Units	Standard of measurement of given parameter. (feet, meters, coded values, etc.)	O	free text	-
210	Minimum Value	Minimum value accepted for processing in the model.	O	free real	-
211	Maximum Value	Maximum value accepted for processing in the model.	O	free real	-

212	Default Values	The default value(s) assigned by the modeling software and/or modeler.	O	free text	-
213	Parameter Comments	Any additional comments required to describe the particular input parameter	O	free text	-
214	Processing Parameter Dataset File	URL address to an external file containing the particular dataset or a documentation of the parameter	C- if there is no Processing Parameter metadata file (line 201)	free text	
215	Computational Representation	The physical data structure of the dataset required for the model.	C- Is there no reference to an external file containing this information provided in line 121/ 201?	free text	
216	Dataset Repeatability	Indication if more than one of these datasets may be provided. (Zero signifies an "optional" dataset.)	M	0 to N	1

4.4 Output data

The CSCM offers an own section for output data. However, we propose to use the *dataset* description in ISO19115 format. There are rarely differences to other geodata sets which would require an other metadata structure (Wosniok et al., 2011). However, by referencing on the respective other metadata set via an URL, information can be connected.

4.5 Calibration Efforts and Validation

In section 9, the CSCM proposes elements calibration and validation of the used model. These elements are similarly vague and based on free text as the IAHR validation documents. As the measures for model calibration and validation are highly specific for each model, giving a structure would hardly be feasible. Considering the similar contents, we propose to link to the IAHR validation document instead of using the metadata elements proposed by the CSCM.

5. CONCLUSION

Going into details of a model engine involves balancing between too little and too many details. When performing a model run, most of the used data and parameters are used in other model runs as well. This fosters a structure of distributed metadata, similar to the known service oriented architecture paradigm: A metadata set can be accessed via a single URL, which is the base for reaching single elements of the metadata set. This concept can be realized with the CS-W standard.

A standardized exchange of - not only - modeling methods such as described here is demanded and helpful when dealing with several participating parties. This approach of structuring the elements of modeling therefore supports a transparent evaluation of modeling results and consequently leads to more precise assessments.

ACKNOWLEDGMENTS

The work presented here is funded by the Federal Ministry for Education and Research (BMBF) through Project Management Jülich (PtJ) under grant number 03KIS089 (Marine Data Portal). The authors gratefully acknowledge this support as well as the contributions from co-workers and other partners to this research project.

REFERENCES

Chorley, R.J. and Haggett, P. (1967) Integrated Models in Geography, Worshire & London.

European Parliament (2000) “Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000: establishing a framework for Community action in the field of water policy”.

Geller, G.N. and Turner, W. (2007) “The Model Web: A Concept for Ecological Forecasting”, Geo-science and Remote Sensing Symposium, IGARSS 2007, IEEE International, pp. 2469 – 2472.

Hill, L., Crosier, S. J., Smith, R.T. and Goodchild, M. (2001) ”A Content Standard for Computational Models”, D-Lib Magazine, Vol. 7, No. 6 (digital).

IAHR (1994) “Guidelines for Documenting the Validity of Computational Modelling Software”, International Association for Hydraulic Research, P.O. Box 177, 2600 MH Delft, The Netherlands.

ISO/ TC211 (2003) “ISO19115:2003 Geographic Information – Metadata”.

Kösters, F., Plüß, A., Heyer, H., Kastens, M. amd Sehili, A. (2010) "Validating a hydrodynamic framework for long-term modelling of the German Bight". EGU General Assembly 2010, Vienna, Austria.

Lehfeldt, R., Melles, J., Bauer, M. (2011) "Concept And Implementation Of The German Marine Data Infrastructure" In: Roger Longhorn und Stefania de Zorzi (Eds.): Proceedings of CoastGIS 2011: 10. International Symposium on GIS and Computer Mapping for Coastal Zone Management. Marine and Coastal Spatial Planing, Vol.2. Venezia: Corila, pp 50-54.

Soetje, K. C. (2008) "To be on the right path from BOOS to an integrated pan-European marine data management system", Proceedings of: US/EU-Baltic International Symposium, 2008 IEEE/OES, Tallin, pp 1-7.

Voges, R. and Senkler, K (2007). OpenGIS Catalogue Services Specification 2.0.2 - ISO Metadata Application Profile OGC.

Wosniok, C., Bauer, M., and Lehfeldt, R. (2011) "Integrating Marine Modeling Data Into a SDI". In: Angela Schwering, Edzer Pebesma und Kai Behncke (Eds.): Geochange. Conference Proceedings. Heidelberg: AKA, Akad. Verlagsges., pp. 85–91.

Wosniok, C. and Lehfeldt, R. (2012) "A Metadata Profile for Computational Models", HIC 2012 - 10th International Conference on Hydroinformatics, Hamburg.