International Congress on Environmental Modelling and Software

8th International Congress on Environmental Modelling and Software - Toulouse, France - July 2016

Jul 11th, 9:10 AM - 9:30 AM

Documenting Models for Interoperability and Reusability

Rajbir Parmar
U.S. Environmental Protection Agency, parmar.rajbir@epa.gov

Kurt Wolfe
U.S. Environmental Protection Agency, wolfe.kurt@epa.gov

Gene Whelan
U.S. Environmental Protection Agency, Whelan.Gene@epa.gov

Gerard F. Laniak
U.S. Environmental Protection Agency, laniak.gerry@epa.gov

Mike Galvin
U.S. Environmental Protection, galvin.mike@epa.gov

See next page for additional authors

Parmar, Rajbir; Wolfe, Kurt; Whelan, Gene; Laniak, Gerard F.; Galvin, Mike; and Goodall, Jonathan L., "Documenting Models for Interoperability and Reusability" (2016). *International Congress on Environmental Modelling and Software*. 2.
https://scholarsarchive.byu.edu/iemssconference/2016/Stream-A/2

Presenter/Author Information

Rajbir Parmar, Kurt Wolfe, Gene Whelan, Gerard F. Laniak, Mike Galvin, and Jonathan L. Goodall

International Environmental Modelling and Software Society (iEMSs)
8th International Congress on Environmental Modelling and Software
Toulouse, France, Sabine Sauvage, José-Miguel Sánchez-Pérez, Andrea Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2016-proceedings

Documenting Models for Interoperability and Reusability

Rajbir Parmar[a], Kurt Wolfe[a], Gene Whelan[a], Gerard F. Laniak[a], Mike Galvin[a], Jonathan L. Goodall[b]
[a]*U.S. Environmental Protection Agency, Athens, GA USA (Parmar.Rajbir@epa.gov, Wolfe.Kurt@epa.gov, Whelan.Gene@epa.gov, Laniak.Gerry@epa.gov, Galvin.Mike@epa.gov)*
[b]*University of Virginia, Charlottesville, VA USA (goodall@virginia.edu)*

Abstract: Many modeling frameworks compartmentalize science via individual models that link sets of small components to create larger modeling workflows. Developing integrated watershed models increasingly requires coupling multidisciplinary, independent models, as well as collaboration between scientific communities, since component-based modeling can integrate models from different disciplines. Integrated Environmental Modeling (IEM) systems focus on transferring information between components by capturing a conceptual site model; establishing local metadata standards for input/output of models and databases; managing data flow between models and throughout the system; facilitating quality control of data exchanges (e.g., checking units, unit conversions, transfers between software languages); warning and error handling; and coordinating sensitivity/uncertainty analyses. Although many computational software systems facilitate communication between, and execution of, components, there are no common approaches, protocols, or standards for turn-key linkages between software systems and models, especially if modifying components is not the intent. Using a standard ontology, this paper reviews how models can be described for discovery, understanding, evaluation, access, and implementation to facilitate interoperability and reusability.

Keywords*:* Integrated Environmental Modeling; Ontology; Metadata; Interoperability; Reusability

1 INTRODUCTION

Many modeling frameworks compartmentalize science via individual models that link sets of small components to create larger modeling workflows; this allows models from different disciplines to be integrated (Elag and Goodall, 2013). Integrated Environmental Modeling (IEM) is a systems analysis approach with inter-dependent, science-based components (models, data, modules, and assessment methods) that, together, are the basis for constructing an appropriate modeling system (Gaber et al., 2008). IEM is represented by software-based computational systems (platforms or frameworks) that describe how science-based components will be organized/linked and then used to address a specific problem coherently (Gaber et al., 2008). IEM systems focus on transferring information between components by capturing a conceptual site model (CSM); establishing local metadata standards for input/output of models and databases; managing data flow between models and throughout the system; facilitating quality control of data exchanges (e.g., units checking, units conversion, transfers between software languages); handling warnings and errors; and coordinating sensitivity/uncertainty analyses (Whelan et al., 2014). A CSM represents an environmental system of biological, physical, and chemical processes that determine transport of contaminants through environmental media to environmental receptors (ASTM, 2008), and identifies linkages of concern for a particular problem (Whelan et al., 2014). Although many computational software systems facilitate communication between, and execution of, components (Whelan et al., 2014; Laniak et al., 2013), there are no commonly accepted approaches, protocols, or standards for turn-key linkages between software systems and models, especially if modifying components is not the intent.

While there has been a notable increase in component-based modeling frameworks in recent years (Laniak et al., 2013; Whelan et al., 2014), there has been less work on creating standard vocabularies, metadata, semantics, and ontologies (see Table 1) to ensure proper technical and conceptual assemblage, although work on ontologies is gaining traction. SEAMLESS is an example of a framework that explicitly builds component ontologies (Van Ittersum et al., 2008), whereas Elag and Goodall (2012, 2013) and Morsey et al. (2014) designed an ontology for the water resources community using a skeletal methodology described by Uschold and Gruninger (1996). Titled the Water Resources Component (WRC) ontology, it advances applications of component-based modeling frameworks across water-related disciplines. Although the WRC ontology was designed for water resources, it can include other domains – air, ecosystems, socioeconomics, chemicals, microbes, etc. – to document individual modeling components for eventual inclusion in larger, disparate systems. It advances the conceptual integration of components from different but related disciplines by demonstrating how semantic and syntactic heterogeneities among models can be handled to be more easily reused, extended, and maintained by model developers and end users. The WRC has four ontological layers (Elag and Goodall, 2012, 2013):

- Resources: defines digital resources related to the component.
- Coupling: defines coupling standards used by the component, the framework in which the component can be coupled, and the system’s computational resolution.
- Scientific: describes equations, symbols, mathematical classification, and component purpose.
- Technical: defines required computer architecture to employ and edit the component.

The strength of this ontology, like others, is its structure for capturing and documenting key information that defines a component's vocabulary, metadata, and semantics for promoting interoperability of components across disciplines and modeling frameworks. This effort extends WRC and illustrates how an ontological framework can be represented in different formats (e.g., spreadsheet, RDF, etc.) and can be applied to disciplines outside the water resource domain, so models can be discovered, understood, evaluated, accessed, and implemented for interoperability and reusability.

2 METHODS AND MATERIALS

The Scientific Layer describes the component’s equations, Input and Output (I/O) variables, parameters, purpose, and mathematical classification (Elag and Goodall, 2013); it is divided into four classes: Domain, Mathematical Classification, Equation, and Symbol. Domain describes the category with which the model is affiliated (e.g., river model is surface water, aquifer model is groundwater, etc.). Mathematical Classification class defines how variables are treated in space and time and if they are deterministic or stochastic. Equation describes all equations and related assumptions used by the model; its purpose is to cross-correlate 1) input, output, and internal variables; 2) equations using or defining the variables; and 3) associated assumptions. Internal variables refer to those used within the mathematical formulations, not consumed as input or produced as output. Symbol classifies symbols as Independent or Dependent Variables, Parameters, or Constants, where each must have a unique and unambiguous name, and where the names themselves can represent the symbols (Elag and Goodall, 2013). A variable is an entity that changes with respect to another, and a parameter is an entity that connects variables. A variable is a real world entity with a measurable quantity, while a parameter is an entity that may or may not be measurable; therefore, the same set of variables can be described by different parameters (e.g., indices) (Difference Between, 2012). For example, in the equation of a straight line ($y = mx + b$), x and y are independent and dependent variables, respectively, and m and b are parameters. When modeling this equation, x, m, and b are typically inputs, and y is typically an output. The output of one model (which produces dependent variables), could be classified as independent variables or parameters of a downstream model that consumes the information as input.

The focus of this paper is an extension of the WRC ontological metadata format, which is part of the Symbol class. The ontology documents metadata, syntactics, and semantics (Table 1) of the model’s Input/Output (I/O) through expanded dictionaries (Whelan et al., 2014), mathematical formulations that define and/or use each I/O parameter, and constraints (assumptions) associated with each I/O parameter. The intent is to extend variable names and definitions to succinctly capture the vocabulary, metadata, syntactics, semantics, and ontology associated with input, output, and internal variables. As used here, an ontological dictionary groups like and related parameters and provides a single naming convention for variables and parameters shared by modeling components, specifically including ontological attributes outlined in Table 2 (Whelan et al., 2014):

Table 1. Definition of Terms Related to Interoperability and Reusability (after Whelan et al., 2015)

TERM	DEFINITION
Data	Information that is produced and consumed
Vocabulary	Terminological dictionary, which contains designations (e.g., names) and definitions from one or more specific subject fields (JCGM, 2008)
Taxonomy	Science of classification according to a pre-determined system with the resulting catalog used to provide a conceptual framework for discussion, analysis, or information retrieval (i.e., identifies, names, and classifies data,[1] so it can be standardized, shared, and re-used in multiple systems[2])
Metadata	Information about the data used to capture content (Kashyap and Sheth, 2000)
Syntactics	Data structure [i.e., how elements are sequenced to form valid conditions (e.g., keywords, object names, operators, delimiters, etc. in the correct places)]
Semantics	Data and their relationship to other data[3] by relating content and representation of information resources to entities and concepts in the real world (Meersman and Mark, 1997) and including not only the metadata about data but also data's intended use (i.e., application) (Sheth, 2001)
Ontology	Explicit specification of conceptualization, describing knowledge about the domain[4] and relationships between domain concepts[5]

[1] http://en.wikipedia.org/wiki/Taxonomy
[2] http://it.toolbox.com/blogs/irm-blog/the-benefits-of-a-data-taxonomy-4916
[3] http://en.wikipedia.org/wiki/Semantic_data_model
[4] http://www.obitko.com/tutorials/ontologies-semantic-web/ontologies.html
[5] http://en.wikipedia.org/wiki/Ontology_(information_science)

3 RESULTS AND DISCUSSION

Using ontological metadata dictionaries, a definitive relationship can be established between variables, components (e.g., different models) within a workflow, equations, metadata, and assumptions. Even when a variable is shared between multiple components, relationships can be established using an approach similar to a Resource Description Framework (RDF) triple (Price, 2004). An RDF format is the standard for encoding metadata and other knowledge on the semantic web (GitHub, 2014), and an RDF triple consists of the 1) subject, which identifies the object the triple describes; 2) predicate, which defines the piece of data in the object to which we are giving a value; and 3) object that is the actual value. In other words, a subject and an object are linked by a predicate. For our purposes, Elag and Goodall (2013) describe this as a "3-ary" because neither the equation, variable (symbol), nor component can be considered the primary subject.

An expanded example of a "3-ary" in Figure 1 illustrates the relationships between the variable "NumberOfAnimals" and the components (SDMProjectBuilder and Microbial Source Module) and equations that define and use it (Equations 37, and 6,7,9-17,32, respectively). The SDMProjectBuilder accesses, retrieves, analyzes, and caches web-based environmental data; provides geographic information system capabilities; automates watershed delineation; assigns automatically map-layer features (e.g., slope, soil); and automatically pre-populates input files of fate and transport models (Whelan et al., 2015). The Microbial Source Module (MSM) estimates microbial loading rates to land surfaces from non-point sources, and to streams from point sources (Whelan et al., 2015). The SDMProjectBuilder produces input data for consumption by MSM and consumes microbial loadings from MSM to serve up data for possible consumption by other models (Figure 2). In this particular example, the distribution of farms and number of animals (NumberOfAnimals) per farm are collected by the SDMProjectBuilder, which the MSM needs to estimate manure production and microbial loading rates to land surfaces and streams.

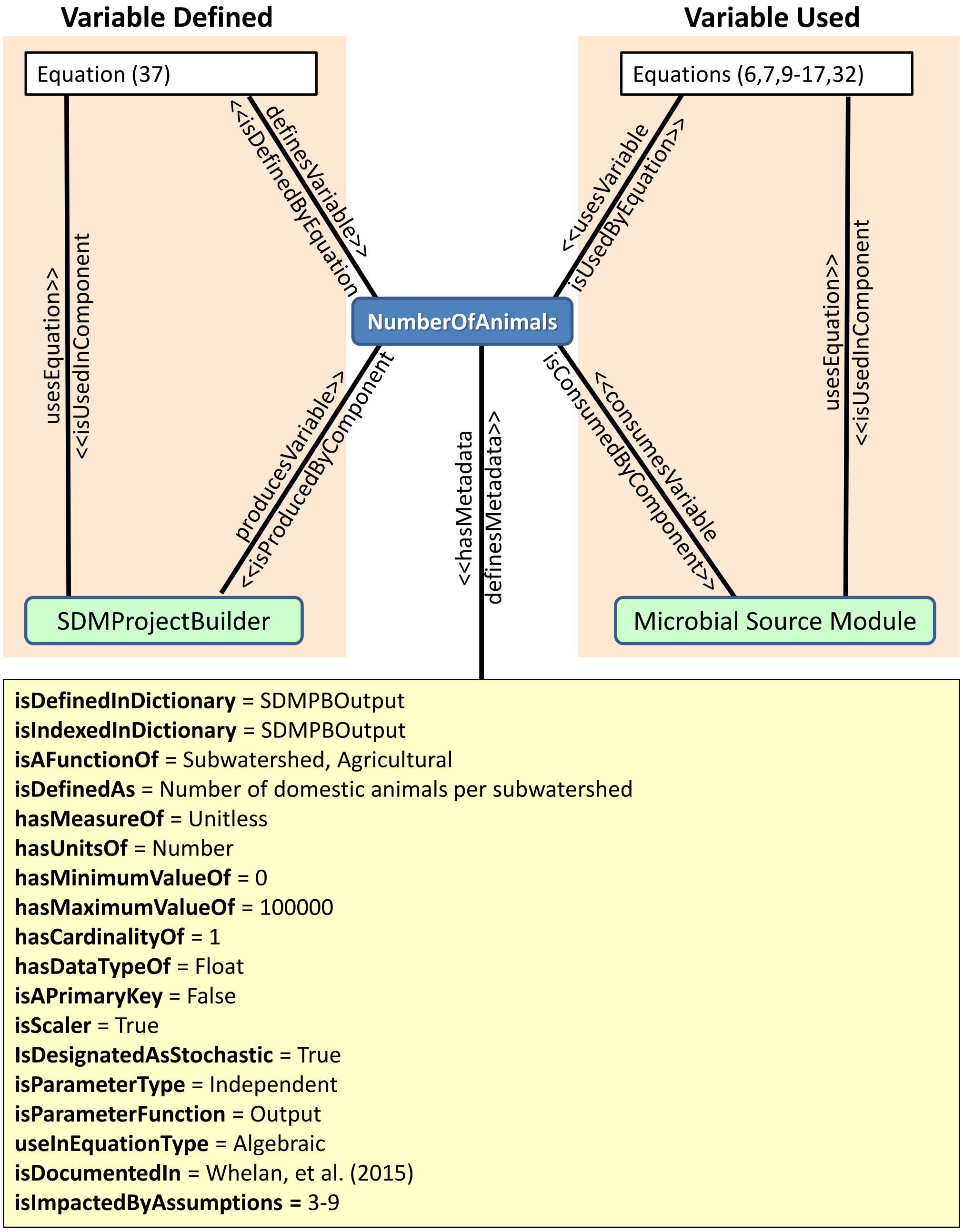
The ontological metadata associated with NumberOfAnimals, of which the second and third columns represent indices or in the metadata terms "isAFunctionOf" in Figure 1, is provided in Table 3. Figure 1 illustrates that both equations and components link to the same variable, through which the metadata and assumptions are described. Interesting features that are captured include 1) linkage of two different modules (SDMProjectBuilder and Microbial Source Module); 2) definition of the parameter in one module (SDMProjectBuilder) which registers it as output, and consumption of the same parameter in another module (MSM) as input; and 3) demonstration of how to account for input as "module-specific" (MSMInput ontological dictionary) and as output (e.g., a boundary condition) (SDMPBOutput ontological dictionary). Since multiple modules could be linked to this variable, this figure is not limited to two.

Table 2. Attributes of an Ontological Dictionary (after Whelan et al., 2014)

#	ATTRIBUTES
1	Parameter/Variable Name
2	Parameter/Variable Description (Definition of parameter/variable)
3	Data Type (String, Float, Integer, Logical)
4	Cardinality [Number of elements in a set or grouping, as a property of that parameter/variable (dimensions). For example, if the parameter "Concentration" is a function of its chemical name, location, and time, it has a cardinality of 3].
5	Primary Key [Parameters/Variables that can be identified and defined only once in a workflow ontology, so that the universal parameter/variable is equally recognized by all components within a workflow (e.g., when all components use the same time reference]
6	Scaler [If TRUE, the variable is not part of a list. If FALSE, it is part of a list and considered self-indexed (a function of itself) or self-enumerated (specified one after another). For example, a time series is typically self-enumerated, so the first time is indexed to 1, the second time to 2, etc. Self-indexing (being non-scaler) increases the parameter/variable cardinality by one.]
7	Parameter/Variable Range (Minimum and Maximum)
8	Measure (Categorizes a collection of units that inherit the same measuring properties; for example, meter, foot, and yard are units for the measure "length.")
9	Parameter/Variable Units (Scaling properties within the same measure.)
10	Stochastic (Boolean that identifies parameters/variables available for statistical manipulation, such as Monte Carlo)
11	Indices [Define elements implied by cardinality, elements in a set or grouping as a property of that parameter/variable). Indices organize the dimensionality of a system by providing hierarchical relationships (context) between variables and parameters, supporting the concept of semantics (see Table 1)].
12	Parameter/Variable Type (Independent, Dependent, Parameter, or Constant)
13	Parameter/Variable Function (Input, Output, Internal: whether the parameter/variable represents input, output, or is associated with linking input to output)
14	Component (Identifies component(s) that defines the parameter/variable)
15	Document (Identifies the document related to the parameter's/variable's descriptions, equations, and assumptions)
16	Equation in Document that defines Parameter/Variable
17	Equations in Document that use Parameter/Variable
18	Equation Type (Algebraic, Differential, or Integral)
19	Relevant Assumption (Assumptions that impact the parameter's/variable's use and/or value)

The concept of an ontological framework for documenting science software "products" (e.g., components, models, databases, assessments, etc.) lends itself to describing knowledge about the product and relationships between product concepts (see Table 1, for example). A science-based software product communicates science theory and software usability, where the traditional means were text-based, although technology is changing to digitized formats that facilitate product understanding as well as automated discovery, evaluation (for a purpose), and integration (with other products). To achieve digitized documentation for communication, discovery, evaluation, and integration, an ontological framework provides a structured, standardized approach to combine data, taxonomy, and relationships among concepts and data. Hence, the WRC ontology framework described by Elag and Goodall (2013, 2012) encompasses many elements (data, taxonomy, concepts, and relationships) in one format such as OWL, web ontology language.

Spreadsheets, illustrated by Table 3, combine essential variables with metadata and intra- and inter-parameter relationships between variables (see Figure 1). They have traditionally been used because they are intuitive and most environmental modelers are comfortable with them, although they are not the only format that can capture ontological metadata. By agreeing on a format to express data exchange, tools can facilitate the translation of this information into higher-level ontological frameworks such as OWL. With standardization, user-friendly, graphical user interfaces (GUIs) can be developed from a spreadsheet to capture ontological metadata, as illustrated by the FRAMES Dictionary Editor (Whelan et al., 2014). Likewise, spreadsheet-based ontological metadata can be converted easily to GUIs, as illustrated by the Dictionary Registration Tool (Pelton, 2009), resulting in interchangeable forms of the same ontological metadata (i.e., spreadsheet to GUI and vice versa).



isDefinedInDictionary = SDMPBOutput
isIndexedInDictionary = SDMPBOutput
isAFunctionOf = Subwatershed, Agricultural
isDefinedAs = Number of domestic animals per subwatershed
hasMeasureOf = Unitless
hasUnitsOf = Number
hasMinimumValueOf = 0
hasMaximumValueOf = 100000
hasCardinalityOf = 1
hasDataTypeOf = Float
isAPrimaryKey = False
isScaler = True
IsDesignatedAsStochastic = True
isParameterType = Independent
isParameterFunction = Output
useInEquationType = Algebraic
isDocumentedIn = Whelan, et al. (2015)
isImpactedByAssumptions = 3-9

Figure 1. Relationships between the variable “NumberOfAnimals” and components, equations, metadata, and assumptions (after Whelan et al., 2015)

Moving toward controlled vocabularies to name data elements and associated metadata, standardized tools can link controlled vocabularies (and definitions) to individual software product digitized formats. Coupled with taxonomy (classification of concepts) relative to science software, a more complete ontology can be documented and tools developed to compare, merge, and produce such files and formats, as illustrated by expression of the WRC ontology using Protégé (2014). Figure 3a illustrates interchangeable forms describing ontological metadata (or schema) between spreadsheets, GUIs, and ontology editors.

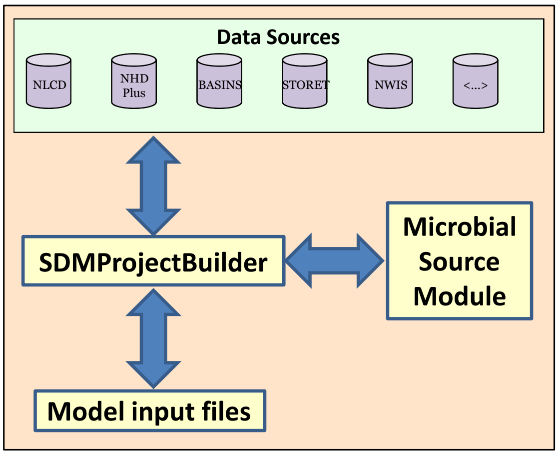


Figure 2. Workflow relationships between SDMProjectBuilder, data sources, Microbial Source Module and, downstream files

When coupled with input values, data transfer with metadata can ensure proper quality control within and between components – not only is the value known, but its metadata (description, units, range, relationships to other parameters/variables, etc.) accompanies it. Standardization facilitates multiple formats for expressing values with their ontological metadata (Figure 3b); for example, it allows standardized user interfaces to deliver input to or produce data from models (e.g., FRAMES's Data Client Editor; DCE, 2010). This information can be easily converted to a flat file (e.g., csv, txt) or expressed electronically, as illustrated by output captured in an Extensible Markup Language (XML) in Figure 3b. XML is 1) a standard or set of rules that governs encoding of documents into an electronic format (Difference Between, 2014) that is human- and machine-readable; 2) a textual data format with strong support via Unicode for different human languages; and 3) used widely to represent arbitrary data structures such as those in web services (XML Wikipedia, 2014). An XML document captures rules in a readable form and is compared to the XML schema (XSD) that executes the web service. Comparison verifies the syntax and validates the structure of an XML document.

4 SUMMARY

This paper focuses on an extension of the Water Resources Component (WRC) ontological metadata format. Ontology documents metadata, syntactics, and semantics of a model's Input/Output (I/O) through expanded dictionaries, mathematical formulations that define and/or use each I/O parameter/variable, and constraints (i.e., assumptions) associated with each I/O parameter/variable. The intent is to expand parameter/variable names and definitions to succinctly capture the vocabulary, metadata, syntactics, semantics, and ontology associated with input, output, and internal parameters/variables. Twenty attributes are described that can enhance a model's interoperability and reusability. An example application on a subset of parameters/variables is provided in which relationships are defined between a variable, its equations, and its use between and within multiple models, established through an approach similar to a Resource Description Framework (RDF) triple "3-ary", which captures the relationship between a variable and its components. Finally, there is a brief discussion on interchangeable forms describing ontological metadata and its relationship to input value assignments.

5 ACKNOWLEDGMENTS

The views expressed in these Proceedings are those of the individual authors and do not necessarily reflect the views and policies of the United States Environmental Protection Agency (EPA). Scientists in EPA have prepared the EPA sections, and those sections have been reviewed in accordance with EPA's peer and administrative review policies and approved for presentation and publication.

Table 3. Ontological metadata associated with NumberOfAnimals and its relationship to other parameters/variables and their ontological metadata (after Whelan et al., 2015)

Dictionary Name	SDMPBOutput	SDMPBOutput	SDMPBOutput	MSMInput
Name	Agriculture	Subwatershed	NumberOfAnimals	NumberOfAnimals
Description	Domestic Animal Name, self-indexed (i.e., self-enumerated)	Subwatershed Identification designation, self-indexed (i.e., self-enumerated)	Number of domestic animals (Agricultural) per subwatershed (Subwatershed)	Number of domestic animals (Agricultural) per subwatershed (Subwatershed)
Cardinality	1	1	2	2
Data Type (Float, Integer, etc.)	String	Integer	Float	Float
Primary Key (i.e., used as a Unversal parameter?)	TRUE	TRUE	FALSE	FALSE
Scaler [Not Self-Indexed (i.e., not self-enumerated) = True]	FALSE	FALSE	TRUE	TRUE
Minimum		0	0	0
Maximum		1000	1.0E+05	1.0E+05
Measure			Unitless	Unitless
Unit			Number	Number
Stochastic (Is it allowed to change in a Monte Carlo analysis?)	FALSE	FALSE	TRUE	TRUE
Index 1			SDMPBOutput.Subwatershed	SDMPBOutput.Subwatershed
Index 2			SDMPBOutput.Agricultural	SDMPBOutput.Agricultural
Type [Independent or Dependent Variable, Parameter (e.g., Index)]	Parameter	Parameter	Dependent	Independent
Function (Input, Output, Internal)	Input	Input	Output	Input
Component	Microbial Source Module	Microbial Source Module	SDMProjectBuilder	Microbial Source Module
Document (Reference number with reference)(a)	1. Whelan et al. (2015)	1. Whelan et al. (2015)	1. Whelan et al. (2015)	1. Whelan et al. (2015)
Equation in Document that Defines Parameter/Variable (Reference numbers with relevant equations in parentheses)(a)			1 (37)	1 (37)
Equations in Document that use Parameter/Variable (Reference numbers with relevant equations in parentheses)(a)	1 (1,2,6-17, 32,37)	1 (6,7,9-23, 27-40)	1 (6,7,9-17,32)	1 (6,7,9-17,32)
Equation Type			Algebraic	Algebraic
Relevant Assumption (Reference number with relevant assumptions in parentheses)(a)			1 (4-10)	1 (4-10)

(a) First number indicates relevant reference under "Document"

6 REFERENCES

ASTM (American Society for Testing and Materials), 2008. Standard Guide for Developing Conceptual Site Models for Contaminated Sites. ASTM Standard E1689 e 95. ASTM International, West Conshohocken, PA. http://www.astm.org/Standards/E1689.htm (last accessed 27.01.16).

DCE, 2010. FRAMES Downloads, EarthDomainDocumentation.Zip. Data Client Editor (DCE)_0710.ppt. http://iemhub.org/resources/133/supportingdocs (last accessed 21.03.15).

Difference Between, 2014. Difference Between XML and XSD. Difference Between.net <http://www.differencebetween.net/technology/difference-between-xml-and-xsd/#ixzz38KfkFB44> (last accessed 03.08.14).

Difference Between, 2012. Difference Between Variable and Parameter (last accessed 15.03.15), http://www.differencebetween.com/difference-between-variable-and-vs-parameter/.

Elag, M.M., Goodall, J.L., 2013. An ontology for component-based models of water resource systems. *Water Resour. Res.* 49(8), 5077–5091.

Elag, M.M., Goodall, J.L., 2012. Design and Application of an Ontology for Component-Based Modeling of Water Systems, Abstract IN11B-1470. Presented at 2012 Fall Meeting, AGU, San Francisco, Calif., December 3-7.

Gaber, N., Laniak, G.F., Linker, L., 2008. Integrated Modeling for Integrated Environmental Decision Making. EPA 100/R-08/010. U.S. Environmental Protection Agency, Washington, DC.

http://www.epa.gov/CREM/library/IM4IEDM_White_Paper_Final_(EPA100R08010).pdf (last accessed 27.01.16.)

GitHub, 2014. What is RDF and what is it good for? <https://github.com/JoshData/rdfabout/blob/gh-pages/intro-to-rdf.md#> (last accessed 02.08.14).

JCGM, 2008. International vocabulary of metrology — Basic and general concepts and associated terms (VIM), JCGM 200, Joint Committee for Guides in Metrology, Bureau International des Poids et Measures, Sevres Cedex, France. <http://www.bipm.org/en/publications/guides/vim.html> (last accessed 22.03.15.)

Kashyap, V., Sheth, A.P., 2000. Information Brokering across Heterogeneous Digital Data: A Metadata-based Approach, Kluwer Academic Publishers, Boston, MA.

Laniak, G.F., Olchin, G., Goodall, J.L., Voinov, A., Hill, M., Glynn, P., Whelan, G., Geller, G., Quinn, N., Blind, M., Peckham, S., Reaney, S., Gaber, N., Kennedy, R., Hughes, A., 2013. Integrated Environmental Modeling: A Vision and Roadmap for the Future. *Environ. Modell. Softw.* 39, 3-23.

Meersman R., Mark, L. (Eds.), 1997. Database Application Semantics. Chapman and Hall.

Pelton, M.A. 2009. Requirements, Design, and Specifications for Excel Dictionary Registration Tool (DRT), PNWD-4141, Battelle, Pacific Northwest Division, Richland, WA.

Price, R., 2004. What Is An RDF Triple? (last accessed 02.08.14) http://www.robertprice.co.uk/robblog/2004/10/what_is_an_rdf_triple_-shtml/.

Protégé, 2014. Protégé 4 User's Guide. <http://protegewiki.stanford.edu/wiki/Pr4_UG> (last accessed 22.03.15.)

Sheth, A.P., 2001. Changing Focus on Interoperability in Information Systems: From System, Syntax, Structure To Semantics. In: M.F. Goodchild, M.J. Egenhofer, R. Fegeas, and C.A. Kottman (eds). Interoperating Geographic Information Systems, Kluwer Academic Publishers, 25 pp.

Uschold, M., Gruninger, M., 1996. Ontologies: Principles, methods and applications, *Knowl. Eng. Rev.* 11(2), 93–136.

Van Ittersum, M.K., Ewert, F., Heckelei, T., Wery, J., Olsson, J.A., Andersen, E., Bezlepkina, I., Brouwer, F., Donatelli, M., Flichman, G., 2008. Integrated assessment of agricultural systems–A component-based framework for the European Union (SEAMLESS). *Agricultural systems* 96, 150-165.

Whelan, G., Parmar, R., Laniak, G.F., 2015. Microbial Source Module (MSM): Documenting the Science and Software for Discovery, Evaluation, and Integration. EPA/600/B-15/315. U.S. Environmental Protection Agency, Athens, GA.

Whelan, G., Kim, K., Pelton, M.A., Castleton, K.J., Laniak, G.F., Wolfe, K., Parmar, R., Galvin, M., Babendreier, J., 2014. Design of a Component-based Integrated Environmental Modeling Framework. *Environ. Modell. Softw*. 55, 1-24.

XML Wikipedia, 2014. XML. <http://en.wikipedia.org/wiki/XML> (last accessed 03.08.14).

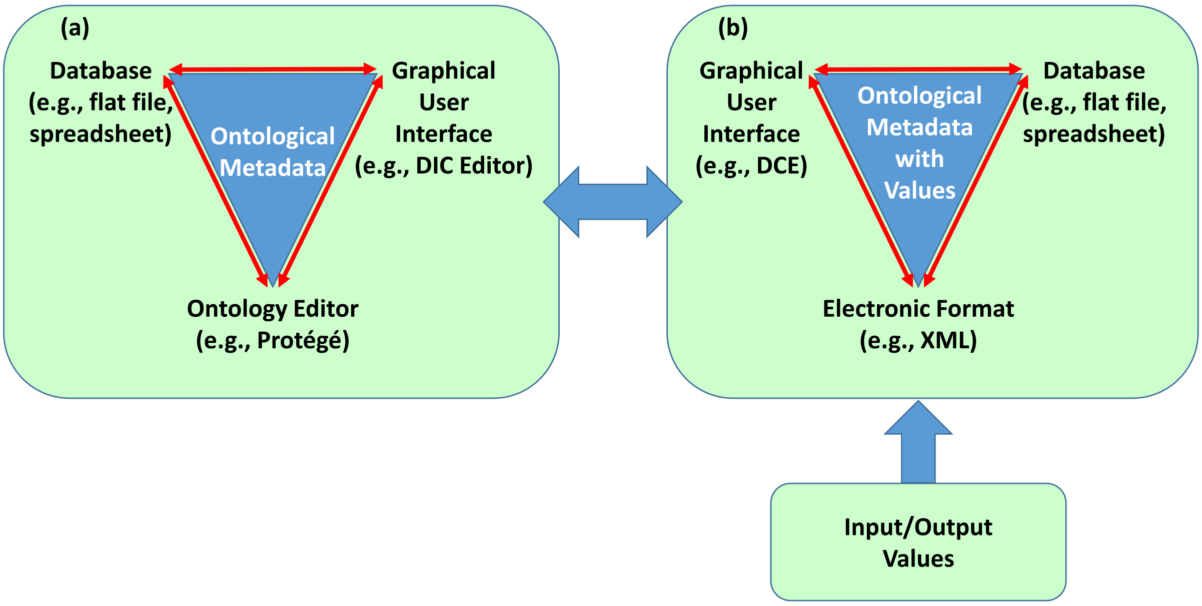


Figure 3. Interchangeable forms (a) describing ontological metadata (or schema) and (b) documenting instances of a dataset related to the schema (after Whelan et al., 2015)