Towards standard metadata to support models and interfaces in a hydro-meteorological model chain

Quillon Harpham and Emanuele Danovaro

ABSTRACT

This paper seeks to move towards an un-encoded metadata standard supporting the description of environmental numerical models and their interfaces with other such models. Building on formal metadata standards and supported by the local standards applied by modelling frameworks, the desire is to produce a solution, which is as simple as possible yet meets the requirements to support model coupling processes. The purpose of this metadata is to allow environmental numerical models, with a first application for a hydro-meteorological model chain, to be discovered and then an initial evaluation made of their suitability for use, in particular for integrated model compositions. The method applied is to begin with the ISO19115 standard and add extensions suitable for environmental numerical models in general. Further extensions are considered pertaining to model interface parameters (or phenomena) together with spatial and temporal characteristics supported by feature types from climate science modelling language. Successful validation of parameters depends heavily on the existence of controlled vocabularies. The metadata structure formulated has been designed to strike the right balance between simplicity and supporting the purposes drawn out by interfacing the Real-time Interactive Basin Simulator hydrological model to meteorological and hydraulic models and, as such, successfully provides an initial level of information to the user.

Key words | controlled vocabulary, environmental numerical modelling, ISO19115, metadata standard, model interface

Quillon Harpham (corresponding author)
HR Wallingford,
Howbery Park, Wallingford,
Oxfordshire OX10 8BA,
UK
E-mail: *q.harpham@hrwallingford.co.uk*

Emanuele Danovaro
CNR-IMATI,
Via De Marini 6 (11th Floor),
16149 Genova,
Italy

ACRONYMS AND ABBREVIATIONS

DRIHM2US	Distributed Research Infrastructure for Hydro-Meteorology to the United States of America
DTM	Digital terrain model
EPA	Environmental Protection Agency
NetCDF	Network Common Data Form
WaterML	Water markup language

INTRODUCTION

It is common practice to pass data between environmental numerical models. A typical one-way connection would consist of part of the output of one model becoming part of the input to the next model down the chain. Building on early incarnations of this process supported by bespoke scripts and file types, many frameworks designed to reduce the effort in achieving such couplings now exist. Johnston *et al.* (2011) describe a US EPA integrated modelling framework for environmental assessment using the Framework for Risk Analysis of Multi-Media Environmental Systems (FRAMES) system; Weerts *et al.* (2010) demonstrate these processes in operational forecasting with the Delft – Flood Early Warning System (FEWS) forecasting platform using published interfaces between models encoded in extensible markup language (XML) and utilising adaptors to handle any differences between outputs produced and inputs required; the Earth System Modelling Framework (ESMF)

is building a flexible software infrastructure to increase inter-operability and reuse in numerical weather prediction and other environmental applications (Hill *et al.* 2004). Peckham *et al.* (2013) describe the design of a component-based approach to integrated modelling in the geosciences and Peckham & Goodall (2013) build on this further by demonstrating interoperability between two independently developed frameworks for models and data. Formal standards for model coupling are now also coming to the fore. Following the earlier open modelling interface (OpenMI) 1.4 (Gregersen *et al.* 2007), OpenMI 2.0 has been ratified by the Open Geospatial Consortium (OGC). OpenMI allows a two-way exchange of data between model components so that they may influence each other as they run (OpenMI Association Website 2014). OpenMI is itself supported by software tools allowing models to be adapted and coupled more easily. One such implementation is HR Wallingford's FluidEarth (Harpham *et al.* 2014) giving a software development kit (SDK) and graphical user interface (GUI) environment together with other supporting material and training.

By definition, the object interfaces defined within the OpenMI specification point the way to metadata describing the model components adapted to be OpenMI compatible. For example, 'output exchange items' are derived to pass data out of the model into another model's 'input exchange items'. Indeed, across all appropriate disciplines, metadata describing numerical models is clearly required to support any kind of automation or semi-automation of the model coupling process. Geller & Melton (2008) look forward to studying the impacts of climate change using a model web where data are passed between models using web services, which would, by definition, be supported by a set of such standards.

Nativi *et al.* (2013) emphasise the need for a clear information model for accommodating the components supporting environmental modelling including model engines and model services. This is supported by FluidEarth's model cataloguing component, configured to describe models as engines (core code) and instances (configured applications). Furthermore, Voinov *et al.* (2014) challenge the very basic processes underpinning common approaches to modelling and recommend a participatory approach, which challenges the traditional approach to modelling itself as a process beginning with a problem formulation and finishing with a product such

as a decision support system. Such thinking would surely demand greater flexibility and more accurate representation from a typical modelling framework.

Given these drivers and building on formal metadata standards supported by the local standards applied by modelling frameworks, this paper seeks to derive an un-encoded metadata structure supporting the description of environmental numerical models with particular attention to the construction of model compositions by interfacing independent model components. The desire is to produce a solution that is as simple as possible yet supports validation of model interfaces together with basic discovery and use requirements.

METHODS

Formulating model engine metadata

Beginning with the model engine, that is the core model code before it has been configured to apply to a particular use case, a number of formally ratified or community standards exist from which to build. In atmospheric science, Murphy *et al.* (2009) describe two such metadata structures incorporated in the Earth System Grid (ESG) and European Common Metadata for Climate Modelling Digital Repositories (METAFOR) projects and characterise a finite volume dynamical core as having 'Basic properties', 'Technical properties', 'Scientific properties', 'Components' and 'Outputs'. The Community Surface Dynamics Modelling System (CSDMS) focuses, as its name would suggest, on modelling earth's surface systems and includes a model repository supported by a metadata structure with 'Summary', 'Contact', 'Technical specs', 'Input/output', 'Process', 'Testing', 'Other' and 'Component info' elements. This community seeks to create metadata for cataloguing earth surface dynamics models in building a catalogue of those available (CSDMS Model Repository 2014). The result is a community standard derived from a sensible set of descriptive fields and implemented in an online repository. ISO19115 (2003) offers an ISO ratified metadata standard for describing spatial datasets, the typical input to and output from environmental models. This standard offers a formal definition covering many similar fields to those required by CSDMS. Another ISO standard, ISO15836 (2009) gives the Dublin

Core Metadata Element Set, a more generic set of elements describing cross-domain resources. Once again, there are many similarities to the more specific ISO19115 and CSDMS community standards. For example, each includes an element providing a general description of the resource ('Abstract' in ISO19115, 'Description' (including an abstract construct) in ISO15836 and 'Extended model description' in CSDMS). The desire in this case is to formulate a candidate metadata structure, which supports the assembly of environmental model chains or compositions. In addition to the usual discovery and to use metadata requirements, particular attention must be paid to the interfaces between the model components. Ideally (and increasingly typically), these interfaces are governed by standards such as OGC OpenMI 2.0 (2014) or OGC WaterML 2.0 (2012) (see, for example, D'Agostino et al. 2014). Users must be able to analyse outputs coming from one model for suitability to use as inputs into another. The attributes associated with these inputs and outputs take particular importance and need to refer, where relevant, to the standards governing the interfaces. As such, ISO19115 was chosen as the starting point for the metadata formulation due to its specific design supporting spatial datasets (Hughes et al. 2013). Drawing from ISO19115 also allows use of a mature set of flexible cataloguing tools implementing the standard together with bespoke extensions such as the FluidEarth Catalogue (2011).

Initially, the approach of CSDMS and Murphy et al. (2009) was followed in drawing together the typical metadata elements required to describe a model engine. It has already been observed that a good proportion of these (such as a title, an abstract, owning organisation or contact details) are present in ISO19115 and more generically in ISO15836. Table 1 gives a base set of model engine metadata elements, their ISO19115 representation and application of each to a hydrological model.

A principal driver for this metadata formulation is to logically extend this description of environmental numerical models to that of their results datasets. Again, elements similar to those adopted by CSDMS (CSDMS Model Repository 2014) and Murphy et al. (2009) are applied as an extension to formulate the complete set of model engine metadata elements and ISO15836 offers a more generic approach including 'format' and 'coverage'. This extension was first applied as part of the FluidEarth model catalogue (FluidEarth Catalogue 2011) in describing model engines. Table 2 documents the FluidEarth extension to ISO19115 with a continuation of the hydrological model example.

Formulating base model instance metadata

When an environmental numerical model engine is applied to a particular situation, a place and a time, it becomes a model instance, which is an instance of that model engine. There is a natural inheritance relationship here where the model instances inherit all of the metadata from their parent model engine. This approach is followed in HR Wallingford's FluidEarth catalogue (FluidEarth Catalogue, 2011) with each model instance being directly associated with just one model engine thereby inheriting all of its metadata.

A further extension to the metadata elements defined above is required to give all of the metadata needed as a minimum to reasonably describe such a model instance. We begin with the spatial aspects with a view to discovering the model instance through a search of spatial extents. Indeed, this is part of the base functionality of the GeoNetwork cataloguing tool for spatial metadata (GeoNetwork 2014). Again, since they have been defined to describe spatial datasets, ISO19115 can provide these spatial elements. Table 3 gives two additional spatial elements used in this extension and shows how they are applied to the hydrological model example used previously.

Formulating interface driven model instance metadata

Further metadata is required to describe model instance outputs and inputs if the metadata set is to have any value in assessing the validity of interfaces to other models. If this metadata is to take a structured form across a large set of models, then the nature of the interfaces will need to be characterised in some way. Three aspects of the model inputs and outputs are singled out as having particular importance in evaluating model interfaces: the spatial characteristics, the temporal characteristics and the environmental parameters (or phenomena) described. These must be defined for each input and output.

The climate science modelling language (CSML) gives a set of 10 spatial feature types describing environmental data (Lowe 2011). Given in Table 4, they have been defined to be

Table 1 | Model engine metadata elements taken from ISO19115

Title, ISO19115 representation and description	Hydrological model example
Title (CI_Citation.title): the title of the dataset (model engine)	RIBS
Dataset Reference Date (CI_Citation.date) and DateType: the date marking the 'creation' of the dataset describing the model engine	2011-05-04: CI_DateTypeCode = creation
Abstract (MD_DataIdentification.abstract): description of the model engine	The Real-time Interactive Basin Simulator (RIBS) model is a distributed hydrological rainfall–runoff model that simulates the basin response to an event of spatially distributed rainfall. This model was designed for real-time application in medium-size basins. The model follows the structure of the grid of a DTM in a matrix form. The data are stored in layers of raster-type information, which are combined to obtain the model parameters
Point of Contact (Organisation) (CI_ResponsibleParty. organisationName): the organisation responsible for the model engine	Technical University of Madrid
Point of Contact (Online Resource) (CI_Contact.onlineResource): URL where more information can be obtained	www.upm.es
Point of Contact (Role): the precise role that the point of contact organisation plays identified as 'custodian'	CI_RoleCode = custodian
Point of Contact (Individual) (CI_ResponsibleParty.individualName): a person who can be contacted regarding this model engine	Luis Garrote
Point of Contact (Organisation) (CI_ResponsibleParty. organisationName): the organisation the individual point of contact belongs to	Technical University of Madrid
Point of Contact (Position) (CI_ResponsibleParty.positionName): the role occupied by the individual point of contact	
Point of Contact (Address and Email) (CI_Contact.address): the postal address of the individual point of contact including their email address	l.garrote@upm.es
Descriptive Keywords (MD_DataIdentification.descriptiveKeywords): a list of keywords describing the model engine	Rainfall, runoff, model
Topic Category Code (MD_TopicCategoryCode): the topic category to which the model belongs, most commonly 'Environment'	Geoscientific information
Date Stamp (MD_Metadata.dateStamp): the date (and time) stamp when the metadata file was created	2011-12-02T12:11:08

specialisations of the observations and measurements (O&M) model (ISO19156 2011) with the exception of 'observation' which is a direct usage. Crucially, these feature types are not only spatial representations, but also incorporate a temporal aspect.

This set of feature types is derived principally from considering earth observations from sensors of various kinds. However, a strong subset can be applied directly to numerical model output: PointSeries, ProfileSeries and GridSeries in particular. As such, the CSML feature types are adopted here as a controlled vocabulary for describing

environmental numerical model inputs and outputs. In addition to this spatial and temporal description, a measure of the precise position of each input/output in space and time is required. The spatial aspect is given through a bounding box for each input and output (in addition to the bounding box representing the model instance as a whole); the temporal aspects are covered similarly by considering the time range covered by each input and output, as well as elements describing their associated timesteps.

Syvitski *et al.* (2014) highlight the need for precise description of model output and input parameter, units

Table 2 | FluidEarth extension to ISO19115 used to describe model engines

Title and description	Hydrological model example
Programming Language: the programming language(s) used to develop the model engine	C++
Supported Platforms: the technical platform(s) supported by the model engine	Windows
Spatial Dimension: the spatial dimension of the model results	2
Source Code URI: a URI from which the source code of the model can be obtained	None supplied
Executable URI: a URI from which the model executable can be obtained	None supplied
Documentation URI: a URI from which the model documentation can be obtained	None supplied
Supported Model Standard: description of the model engine's compatibility with standards such as OpenMI and BMI (Peckham et al. 2013)	None
Supported Model Standard Version: the version of the compatible supported model standard	None
Number of Processors: the number of processors needed to run the model	1
Typical Run Time (and Time Unit): an estimate of the elapsed time for a typical run of the model. Although this may vary, it is included to give a 'ballpark' estimate	100 s
Input: input(s) to the model (Name, Description, Format, whether it is mandatory)	Name: DTM Description: digital terrain model of the basin Format: ESRI shapefile Mandatory: true
Output: output(s) from the model (Name, Description, Format, whether it is mandatory)	Name: hydrograph Description: discharges in time at selected locations Format: WaterML2 Mandatory: false

Table 3 | Additional spatial model instance metadata elements from ISO19115

Title, ISO19115 representation and description	Hydrological model example
Reference System (MD_ReferenceSystem.referenceSystemIdentifier): the coordinate reference system used	urn:ogc:def:crs:EPSG::3857
Extent (EX_GeographicBoundingBox): a geographic two-dimensional bounding box describing the extent of the model instance. The coordinates of the north, south, east and west bounds are given	8.8,44.3;8.8,44.4;9.0,44.4;9.0,44.3

and other attributes at interfaces between models. A set of standard parameter names, CSDMS standard names (CSDMS Standard Names 2013), is given as an extension to the well-established climate and forecasting standard names (CF Standard Names 2003), itself an extension to the Cooperative Ocean/Atmospheric Research Data Service standards (COARDS Conventions 1995). The metadata described here simply uses such standard naming conventions (which often produce very long parameter names) giving space for the precise parameter name and the unit used against each input and output.

The additional metadata elements given to support model interfaces are given in Table 5 with application to the hydrological model.

RESULTS AND DISCUSSION

General applicability

Further to the snippets given as the full metadata structure outlined above, a full example metadata set is given in

Table 4 | Climate science modelling language feature types

CSML feature type	Description	Example
Point	A single observation at a point	A single raingauge measurement
PointSeries	A series of 'Point' observations, varying in time, but not space	A stream of raingauge measurements
Profile	An observation along a vertical line in space	Air temperature at a varying height above sea level
ProfileSeries	A time-series of 'Profile' measurements	A set of air temperature profiles taken at a set of timesteps
Grid	Results given across a set of defined points in space	Two-dimensional high frequency (HF) Radar current output at a single time instant
GridSeries	A time-series of 'Grid' measurements from the same defined grid	Two-dimensional HF Radar current outputs at multiple time instants against the same set of grid points
Trajectory	An observation along a discrete path, varying in time and space	Water quality measurements taken from a moving ship
Section	A series of 'Profiles' from a 'Trajectory'	Marine CTD measurements taken from a moving ship
Swath	A 'Trajectory' but with two spatial dimensions resulting in a 'Grid' output but varying also in time	AVHRR satellite imagery taken from a satellite fly-past
ScanningRadar	Backscatter profiles along a look direction at fixed elevation but rotating in azimuth	Weather radar output

Table 6. It represents the metadata given by the Technical University of Madrid for a hydrological model called RIBS, the Real-time Interactive Basin Simulator (Garrote & Bras 1995), as part of the Distributed Research Infrastructure for Hydro-Meteorology (DRIHM) project (Danovaro et al. 2014).

The result is a human readable metadata set giving the model engine elements together with the three inputs to the model and one output produced by it. The purpose of this metadata set is two-fold: (i) to allow the model to be found (discovery metadata) by potential users, and (ii) to allow potential users to evaluate whether the model is appropriate for their needs (use metadata). In general, the base ISO19115 metadata fields have been designed for these purposes for geospatial datasets, yet their extension into environmental models (in this case, a hydrological model) is equally as effective. The standard topic category code of 'Geoscientific Information' (itself from a keyword list) is generic and high level, but appropriate. Sensible search fields are present including abstract, keywords and point of contact details. The technical information added allows a rudimentary evaluation of the model yielding language and platform details together with a runtime estimate and uniform resource identifiers (URIs) where executables, documentation and source code can be found if they are available.

Evaluating interface feasibility using the RIBS model

We now consider whether it is possible to evaluate the feasibility of using output data from one model as input data to another using just the metadata for the two models. The RIBS model was selected, because it lies in the centre of a hydro-meteorological model chain. Precipitation predictions are provided as input to RIBS from meteorological models. RIBS calculates the catchment drainage and provides hydrographs into hydraulic models. These two file-based, one-way interfaces are denoted the 'P Interface' (or 'Precipitation Interface') and 'Q Interface' (or 'Flow Interface'). The P Interface is an example of passing gridded data between models where RIBS is the 'receiving model' and the Q Interface concerns point data where RIBS is the 'providing model'. This is illustrated in Figure 1. We consider each interface in turn.

The 'P' or 'Precipitation' Interface

The 'P' or 'Precipitation' Interface is the interface between the meteorological model and the hydrological model. The meteorological model produces a series of parameters, in particular precipitation, over the catchment to be drained. The meteorological model sequence can include

Table 5 | Additional model instance input and output metadata elements

Title and description	Hydrological model example
Feature Type: a description of the spatial/temporal structure of the data. Valid values from CSML feature type controlled vocabulary	GridSeries
Position: the two-dimensional geospatial position of the data given as a rectangular bounding polygon	8.8,44.3;8.8,44.4;9.0,44.4;9.0,44.3
Time Range: the timestamp of the first (earliest) and last (latest) reading in the time-series in ASCII format, i.e., YYYY-DD-MMThh:mm:ss + hh (e.g., 2014-01-31T15:46:51 + 01) defining the time interval of the data	2011-11-04T01:00:00 + 01, 2011-11-04T15:00:00 + 01
Timestep Type: indicator of 'regular' or 'irregular' timestep interval. Regular timestep types indicate a fixed interval or set of fixed intervals in the result dataset	Regular
Maximum Timestep Interval: the length of the largest timestep represented in the data and its unit of measurement. Used to allow validation of the temporal stability of interfaces	3,600 s
Minimum Timestep Interval: the length of the smallest timestep represented in the data and its unit of measurement. Used to allow validation of the temporal stability of interfaces	1,800 s
Parameter Name and Unit: the name and unit of measurement of the physical parameter/phenomenon represented	lwe_thickness_of_precipitation_amount m

downscaling routines and also the generation of ensembles. In all these cases, the interface to the hydrological drainage model is the same. The meteorological models produce results, which are usually represented as a three-dimensional terrain following GridSeries, as shown in Figure 2, with results being produced over a set of levels.

One of these three-dimensional results cubes is produced at each timestep. A wide variety of atmospheric parameters (or phenomena) are usually described, ranging from precipitation to wind to air pressure. Precipitation is applicable to the 'P Interface' and the parameter 'lwe_thickness_of_precipitation_amount' (CF Standard Names 2003), calculated at the surface only, is expected to be passed to the hydrological model as a two-dimensional GridSeries.

We now consider evaluating the feasibility of connecting a meteorological model (in this case, Weather Research and Forecasting – Advanced Research (WRF-ARW) model (Michalakes *et al.* 2004)) to RIBS using just metadata expressed using this structure. Table 7 shows the metadata element for an example output from WRF-ARW and Table 8 the counterpart input element, which describes what is expected by RIBS. Both model instances refer to a flash flood event that took place in Genoa, Italy in 2011 (Silvestro *et al.* 2012; Rebora *et al.* 2013; Fiori *et al.* 2014).

As previously discussed, the validation of this potential interface (i.e., whether it is valid to pass such data between the two models) should primarily concern the spatial characteristics, the temporal characteristics and the environmental parameters. The parameter matching is straightforward and depends on correct use of the controlled vocabulary used to describe the parameter and its unit of measurement. The output parameter 'Name' and 'Unit' needs to be compared to the input parameter 'Name' and 'Unit'. In this example, there is a direct match with 'lwe_thickness_of_precipitation_amount' in 'm' supplied by WRF-ARW as output and expected by RIBS as input. If there is not an exact match between the two, the interface may still be valid if there is a formula for translating between the different parameters or units, but it is suggested that such adaptation into common standards be applied within the model suite (albeit as a separate module) and reflected in the metadata in the standard forms.

The temporal characteristics are evaluated by a direct comparison of 'Feature Type' elements (in this example, both 'GridSeries'), 'Timestep Type' (in this example, both 'Regular' but with result data containing more than one interval), the maximum and minimum 'Timestep Interval' and the 'Time Range'. An interface may be deemed valid if the input Time Range does not fall outside the output

Table 6 | Hydrological model example of a model instance metadata set

Citation

    Title: RIBS

    Creation Date: 2011-05-04

Abstract: The Real-time Interactive Basin Simulator (RIBS) model is a distributed hydrological rainfall–runoff model that simulates the basin response to an event of spatially distributed rainfall. This model was designed for real-time application in medium-size basins. The model follows the structure of the grid of a DTM in a matrix form. The data are stored in layers of raster-type information, which are combined to obtain the model parameters

Point of Contact

    Custodian Organisation Name: Technical University of Madrid

    Custodian Online Resource: www.upm.es

    Responsible Individual

        Name: Luis Garrote

        Organisation: Technical University of Madrid

        Position:

        Address and Email: l.garrote@upm.es

Descriptive Keywords: rainfall, runoff, model

Topic Category Code: geoscientific information

Date Stamp: 2011-12-02T12:11:08

Reference System: urn:ogc:def:crs:EPSG::3857

Extent: 8.88,44.37; 8.88,44.50; 9.09,44.50; 9.09,44.37

Programming Language: C++

Supported Platforms: Windows

Spatial Dimension: 2

Source Code URI:

Executable URI:

Documentation URI:

Supported Model Standard: none

Supported Model Standard Version: none

Number of Processors: 1

Typical Run Time

    Duration: 100

    Unit: second

Input

    Name: DTM

    Description: digital terrain model of the basin

    Format: ESRI shapefile

    Mandatory: true

    Feature Type: Grid

    Position: 8.88,44.37; 8.88,44.50; 9.09,44.50; 9.09,44.37

    Parameter

        Name: height above sea level

        Unit: m

    Time Range: none

    Timestep Type: regular/irregular

(continued)

Table 6 | continued

Maximum Timestep Interval: none

Minimum Timestep Interval: none

Input

  Name: soil type

  Description: spatially distributed map of soil types, according to a local soil type categorisation

  Format: ESRI Shapefile

  Mandatory: true

  Feature Type: Grid

  Position: 8.88,44.37; 8.88,44.50; 9.09,44.50; 9.09,44.37

  Parameter

    Name: soil type

    Unit: local categorisation

  Time Range: none

  Timestep Type: regular/irregular

  Maximum Timestep Interval: none

  Minimum Timestep Interval: none

Input

  Name: precipitation

  Description: spatially distributed fields of rainfall

  Format: NetCDF 1.6

  Mandatory: true

  Feature Type: GridSeries

  Position: 8.88,44.37; 8.88,44.50; 9.09,44.50; 9.09,44.37

  Parameter

    Name: lwe_thickness_of_precipitation_amount

    Unit: m

  Time Range: 2011-11-04T01:00:00 + 01,2011-11-04T15:00:00 + 01

  Timestep Type: regular/irregular

  Minimum Timestep Interval: 1,800 s

 Maximum Timestep Interval: 3,600 s

Output

  Name: hydrograph

  Description: discharges in time at selected locations

  Format: WaterML2

  Mandatory: false

  Feature Type: PointSeries

  Position: 8.9538,44.4108; 8.9538,44.4109; 8.9539,44.4109; 8.9539,44.4108

  Parameter

    Name: River_Discharge

    Unit: m^3s^{-1}

  Time Range: 2011-11-04T01:00:00 + 01,2011-11-05T12:00:00 + 01

  Timestep Type: regular

  Maximum Timestep Interval: 300 s

  Minimum Timestep Interval: 300 s
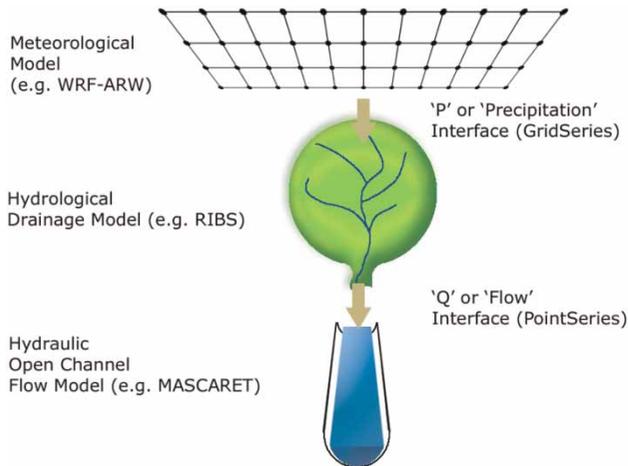
Figure 1 | The 'P' (Precipitation) and 'Q' (Flow) interfaces between the meteorological model, the hydrological drainage model and the hydraulic, open channel flow model.

Time Range and the Timestep Intervals between the two models are within a defined tolerance. These conditions may not always be necessary however, and this largely depends on how each model operates.

A comparison of spatial characteristics also depends on the Feature Type due to the dual spatial and temporal nature of this descriptor. Otherwise, the spatial validation consists solely of a comparison of 'Position'. Position consists of a bounding box (or polygon) expressed in the coordinate system defined once for the model instance. Usually, it would be expected that the input bounding box not lie outside that of the output model so that the spatial coverage required by the input model is guaranteed. If the bounding boxes are both rectangular, axis aligned and expressed in the same coordinate system then this comparison is

simple, otherwise spatial functions to compare polygons and transform coordinate systems are required. Assuming the same coordinate system, in this example, it can be seen that the RIBS input bounding box lies within the WRF-ARW output bounding box sitting on its northern boundary, both expressing the boundary of the model grid supporting their respective GridSeries.

There are two remaining metadata elements to be considered when validating model interfaces: 'Mandatory' and 'Format'. Clearly, if an output from one model is not mandatory then the input model cannot expect to receive it – any interface between the models must have such output guaranteed. Also, the Format element is largely informational giving certain technical information, in this case, a NetCDF 1.6 file is passed by WRF-ARW and expected by RIBS. However, a direct match of a loosely typed structure such as this does not guarantee that the interface will operate without the need for interpolation between the two files, and moreover, a controlled vocabulary does not exist to allow direct text matching in this field.

The 'Q' or 'Flow' Interface

The Q Interface (the letter Q given to represent flow, or discharge) is the interface between the hydrological drainage model, RIBS and the hydraulic open channel model. RIBS calculates the drainage into the river channel and produces a hydrograph giving the flow at a certain point on the river network. Wherever hydraulic modelling is required, a hydrograph needs to be present. That is, for every reach of the river that requires open channel modelling, a flow-time boundary
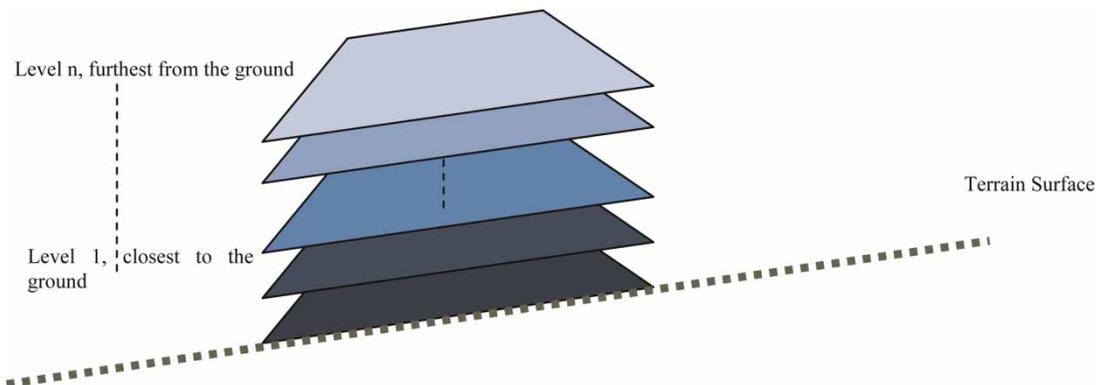
Figure 2 | Three-dimensional results from meteorological models.

Table 7 | Output metadata element from WRF-ARW

Output

Name: precipitation

Description: liquid water equivalent thickness of precipitation amount at the surface, defined as lwe_thickness_of_stratiform_precipitation_amount + lwe_thickness_of_convective_precipitation_amount

Format: NetCDF 1.6

Mandatory: false

Feature Type: GridSeries

Position: 8.50,44.25; 8.50,44.50; 9.25,44.50; 9.25,44.25

Parameter

Name: lwe_thickness_of_precipitation_amount

Unit: m

Time Range: 2011-11-04T01:00:00 + 01,2011-11-05T12:00:00 + 01

Timestep Type: regular

Minimum Timestep Interval: 900 s

Maximum Timestep Interval: 3,600 s

Table 8 | Input metadata element from RIBS

Input

Name: precipitation

Description: spatially distributed fields of rainfall

Format: NetCDF 1.6

Mandatory: true

Feature Type: GridSeries

Position: 8.88,44.37; 8.88,44.50; 9.09,44.50; 9.09,44.37

Parameter

Name: lwe_thickness_of_precipitation_amount

Unit: m

Time Range: 2011-11-04T01:00:00 + 01,2011-11-04T15:00:00 + 01

Timestep Type: regular

Minimum Timestep Interval: 900 s

Maximum |Timestep Interval: 1,800 s

condition must be supplied at the top of the stretch to be modelled. This information is passed to the hydraulic, open channel flow model, as illustrated in Figure 1.

We now consider evaluating the feasibility of passing the output from RIBS into a hydraulic open channel model (in this case, MASCARET (Goutal & Maurel 2002; Goutal et al. 2012)), using just metadata expressed in this structure.

Table 9 | Output metadata element from RIBS

Output

Name: hydrograph

Description: discharges in time at selected locations

Format: WaterML2

Mandatory: false

Feature Type: PointSeries

Position: 8.9538,44.4108; 8.9538,44.4109; 8.9539,44.4109; 8.9539,44.4108

Parameter

Name: River_Discharge

Unit: m^3s^{-1}

Time Range: 2011-11-04T01:00:00 + 01,2011-11-05T12:00:00 + 01

Timestep Type: regular

Maximum Timestep Interval: 300 s

Minimum Timestep Interval: 300 s

Table 9 shows the metadata element for an example output from RIBS and Table 10 the counterpart input element, which describes what is expected by MASCARET. Again, both model instances refer to the same Genoa flash flood from 2011 and together with the WRF-ARW model instance constitute a viable model chain.

The metadata design leads to performing the same validation of this potential interface as for the example P

Table 10 | Input metadata element from MASCARET

Input

Boundary Conditions

Description: discharge or level hydrograph, rating curve

Format: WaterML2

Mandatory: true

Feature Type: PointSeries

Position: 8.95388,44.41083; 8.95388,44.41084; 8.95389,44.41084; 8.95389,44.41083

Parameter

Name: River_Discharge

Unit: m^3s^{-1}

Time Range: 2011-11-04T01:00:00 + 01,2011-11-05T12:00:00 + 01

Timestep Type: regular

Maximum Timestep Interval: 300 s

Minimum Timestep Interval: 300 s

Interface, above. This time, the output parameters Name and Unit refer to a parameter called 'River_Discharge' measured in m^3s^{-1}. This parameter does not exist in CF Standard Names (CF Standard Names 2003). It has been defined as a candidate addition to such controlled vocabularies and corresponds to the 'Discharge, stream' item in the Consortium of Universities for the Advancement of Hydrological Science Incorporated – Hydrologic Information System (CUAHSI-HIS) ontology (Zaslavsky et al. 2012). A similar parameter, 'channel_outflow_end_water_discharge', also exists in the draft CSDMS standard names controlled vocabulary (CSDMS Standard Names 2013).

The temporal validation is the same as that explored above and gives the same outcome. A direct comparison of 'Feature Type' elements (in this example, both 'Point-Series') , 'Timestep Type' (both 'Regular'), the maximum and minimum 'Timestep Interval' and the 'Time Range' proceeds in the same way and yields the same uncertainty over validating timestep intervals and time ranges. However, the implications of a bounding box (or polygon) around a Grid-Series feature type are somewhat different to that of a PointSeries. In this example, RIBS produces data as a single PointSeries and MASCARET is expecting to receive a PointSeries. Geospatially, this is represented by a single point and sensible validation would ensure that the point used by RIBS is in the same place as that expected by MASCARET. It is reasonable to assume that there will be rounding errors in each representation or that each model has expressed the point in a slightly different position (the point given in this example is on the Bisagno river above Genoa (see Silvestro et al. 2012)). As such, a tight bounding box is given to represent the RIBS output (instead of a single point) and another for the MASCARET input. If the same validation is used as in the P Interface, then the MASCARET bounding box must lie inside the RIBS bounding box for the interface to pass this validation.

Table 11 | Candidate set of model interface validation conditions

Condition	Pseudo-code
Parameter Name and Unit: the providing model output parameter name and unit must match with the receiving model input parameter name and unit	receivingModel.input.parameterName = providingModel.output. parameterName AND receivingModel.input.parameterUnit = providingModel.output.parameterUnit
Feature Type: the providing model output feature type must match the receiving model input feature type	receivingModel.input.featureType = providingModel.output.featureType
Timestep Type: if the providing model output has an irregular timestep, check that the receiving model can accept it	If providingModel.output.timestepType = 'irregular' then receivingModel. input.timestepType must = 'irregular'
Time Range: warn if the time range of the receiving model input lies outside the time range of the providing model output	receivingModel.input.timeRange.minimumTime > =providingModel. output.timeRange.minimumTime AND receivingModel.input. timeRange.maximumTime < =providingModel.output.timeRange. maximumTime
Timestep Interval: warn if the minimum timestep interval of the receiving model input is less than a defined multiplier of the maximum timestep interval of the providing model output	receivingModel.input.maximumTimestepInterval < = tolerance*providingModel.output.minimumTimestepInterval 'for an appropriate tolerance'
Position: the bounding box of the receiving model input has to be contained entirely within the bounding box of the providing model output	providingModel.output.position contains receivingModel.input.position 'or if geospatial functionality is not available, for rectangular bounded grids only and ignoring wrapping from 0 to 360 (or −180 to 180)': greatest providingModel.output.y-coordinate > =greatest receivingModel.input.y-coordinate AND smallest providingModel.output.y-coordinate < = smallest receivingModel.input.y-coordinate AND greatest providingModel.output.x-coordinate > =greatest receivingModel.input.x-coordinate AND smallest providingModel.output.x-coordinate < = smallest receivingModel.input.x-coordinate
Mandatory: warn if the providing model output is not mandatory	providingModel.output.Mandatory = false

As with the P Interface, above, if the output from RIBS is not mandatory, then MASCARET is not guaranteed to receive any data and the same issues arise with a comparison of the 'Format' element.

'P' and 'Q' Interface validation summary

Accordingly, a candidate set of validation conditions with pseudo-code supporting both the P and Q Interfaces (as examples of a typical file-based GridSeries-to-GridSeries and PointSeries-to-PointSeries interfaces) can be summarised in Table 11.

CONCLUSIONS

The purpose of metadata is to provide supporting information to allow what it is describing to be found, correctly interpreted and utilised. In environmental modelling use cases such as the hydro-meteorological model chain discussed here, the utilisation aspects increasingly depend on the ability to interface models with each other (and, indeed, other supporting datasets). Standards such as ISO19115 and ISO15836 provide formal patterns for establishing such metadata sets. The effectiveness of any metadata structure and its resulting encoding lies in achieving the right level of complexity for the common requirements to be placed on it. If the metadata is too comprehensive, then there is a risk that suppliers will not provide it, or that provided metadata sets will be of low quality and not maintained. If the metadata is not comprehensive enough, then it will not be fit for its intended purpose.

The purpose of the metadata outlined here is to allow environmental numerical models to be discovered (discovery metadata) and then an initial evaluation made of their suitability for use (use metadata), in particular with reference to interfacing with other numerical models, with a first application for a hydro-meteorological model chain. As such, ISO19115 provides the important base elements as constructed for geospatial datasets, and a small number of additions extend its usage into environmental numerical models. Further extensions describing environmental parameters (or phenomena), temporal and spatial attributes

have been added to allow analysis of potential interfaces using inputs and outputs as follows:

- Successful validation of parameters depends heavily on the existence of controlled vocabularies. The interfaces to and from the hydrological RIBS model example demonstrate that these controlled vocabularies are more mature when interfacing to meteorological models than to hydraulic models.
- A level of temporal validation can be achieved by considering a limited number of attributes, most importantly the time range covered by the model.
- Use of a bounding box (or polygon) to describe spatial coverage is satisfactory for all of the CSML defined feature types and is particularly simple to apply if rectangular and in a common coordinate system. Precise validation is not possible without providing metadata including complete and comprehensive descriptions of the geo-temporal structures supporting the data.

The metadata structure formulated has been designed to strike the right balance between simplicity and supporting the purposes drawn out by the hydro-meteorological model chain and, as such, successfully provides an initial level of validation. It is easy to establish a base knowledge of the model functions and technology, the temporal and spatial coverage and the environmental parameters handled. This extends to individual interfaces with metadata attribution added to model inputs and outputs. However, a more comprehensive analysis and, in particular, precise confirmation that a model interface is valid would only be possible with considerably more information. Attempting to provide this with metadata, which must be available before the datasets are produced by the models, risks construction of an unwieldy metadataset, which would unnecessarily duplicate supplementary and essential model documentation and subsequent results datasets represented in self-describing file types such as NetCDF (OGC NetCDF 2011) and WaterML2 (OGC WaterML 2.0 2012).

ACKNOWLEDGEMENTS

Number 283568 and DRIHM2US Project, Grant Number 313122. Metadata for the RIBS model was adapted from that supplied by Luis Garrote of the Universidad Politecnica de Madrid. Examples of models catalogued using a version of the metadata structure outlined in this paper can be found in the DRIHM Model Catalogue (2014).

REFERENCES

CF Standard Names 2003 CF Metadata NetCDF CF Metadata Conventions. http://cf-convention.github.io/index.html (accessed 29 April 2014).

COARDS Conventions 1995 Conventions for the Standardization of NetCDF Files. http://ferret.wrc.noaa.gov/noaa_coop/coop_cdf_profile.html (accessed 29 April 2014).

CSDMS Model Repository 2014 All Models. http://csdms.colorado.edu/wiki/Model_download_portal (accessed 28 February 2014).

CSDMS Standard Names 2013 Standard Name Examples (Version 0.7.1). http://csdms.colorado.edu/wiki/CSN_Searchable_List (accessed 29 April 2014).

D'Agostino, D., Clematis, A., Galizia, A., Quarati, A., Danovaro, E., Roverelli, L., Zereik, G., Kranzlmuller, D., Schiffers, M., Gentschen Felde, N., Straube, C., Parodi, A., Fiori, E., Delogu, F., Caumont, O., Richard, E., Garrote, L., Harpham, Q., Jagers, B., Dimitrijević, V. & Dekic, L. 2014 The DRIHM project: a flexible approach to integrate HPC, grid and cloud resources for hydro-meteorological research. To be published in SC '14: In: *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, November 16–21, 2014* New Orleans, LA, USA.

Danovaro, E., Roverelli, L., Zereik, G., Galizia, A., D'Agostino, D., Quarati, A., Clematis, A., Delogu, F., Fiori, E., Parodi, A., Straube, C., Felde, N., Harpham, Q., Jagers, B., Garrote, L., Dekic, L., Ivkovic, M., Richard, E. & Caumont, O. 2014 Setup an hydro-meteo experiment in minutes: the DRIHM e-infrastructure for hydro-meteorology research. To be published in the proceedings of e-Science 2014: In: *10th IEEE International Conference on e-Science, October 20–24, 2014* Guarujá, SP, Brazil.

DRIHM Model Catalogue 2014 DRIHM Model Catalogue. http://drihmcatalogue.fluidearth.net/, in publication.

Fiori, E., Comellas, A., Molini, L., Rebora, N., Siccardi, F., Gochis, D., Tanelli, S. & Parodi, A. 2014 Analysis and hindcast simulations of an extreme rainfall event in the Mediterranean area: the Genoa 2011 case. *Atmos. Res.* **138**, 13–29.

FluidEarth Catalogue 2011 Welcome to the FluidEarth Catalogue. http://catalogue.fluidearth.net (accessed 2 May 2014).

Garrote, L. & Bras, R. L. 1995 A distributed model for real-time forecasting using digital elevation models. *J. Hydrol.* **167**, 279–306.

Geller, G. N. & Melton, F. 2008 Looking forward: applying an ecological model web to assess impacts of climate change. *Biodiversity* **9** (3–4), 79–93.

GeoNetwork 2014 GeoNetwork Open Source. http://geonetwork-opensource.org/ (accessed 2 May 2014).

Goutal, N. & Maurel, F. 2002 A finite volume solver for 1D shallow-water equations applied to an actual river. *Int. J. Numer. Methods Fluids* **38**, 1–19.

Goutal, N., Lacombe, J.-M., Zaoui, F. & El-Kadi-Abderrezzak, K. 2012 MASCARET: a 1-D open-source software for flow hydrodynamic and water quality in open channel networks. In: *River Flow* (Rafael Murillo Muñoz ed.). Taylor & Francis Group, London, pp. 1169–1174.

Gregersen, J. B., Gijsbers, P. J. A. & Westen, S. J. P. 2007 OpenMI: open modelling interface. *J. Hydroinform.* **9** (3), 175–191.

Harpham, Q. K., Cleverley, P. & Kelly, D. 2014 The Fluid Earth 2 implementation of OpenMI 2.0. *J. Hydroinform.* **16** (4), 890–906.

Hill, C., DeLuca, C., Balaji Suarez, M. & da Silva, A. 2004 The architecture of the earth system modeling framework. *Comput. Sci. Eng.* **6**, 18–28.

Hughes, A. G., Harpham, Q. K., Riddick, A. T., Royse, K. R. & Singh, A. 2013 *Meta-Model: Ensuring the Widespread Access to Metadata and Data for Environmental Models: Scoping Report.* British Geological Survey, Nottingham, UK, 39 (OR/13/042).

ISO15836 2009 ISO15836:2009 Information and Documentation – The Dublin Core Metadata Element Set. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=52142 (accessed 28 August 2014).

ISO19115 2003 ISO 19115:2003 Geographic Information – Metadata. http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020 (accessed 2 May 2014).

ISO19156 2011 ISO 19156:2011 Geographic Information – Observations and Measurements. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=32574 (accessed 2 May 2014).

Johnston, J. M., McGarvey, D. J., Barber, M. C., Laniak, G., Babendreier, J., Parmar, R., Wolfe, K., Kraemer, S. R., Cyterski, M., Knightes, C., Rashleigh, B., Suarez, L. & Ambrose, R. 2011 An integrated modeling framework for performing environmental assessments: application to ecosystem services in the Albemarle-Pamlico basins (NC and VA, USA). *Ecol. Modell.* **222**, 2471–2484.

Lowe, D. 2011 Climate Science Modelling Language v3.0 British Atmospheric Data Centre. http://csml.badc.rl.ac.uk/ (accessed 2 May 2014).

Michalakes, J., Dudhia, J., Gill, D., Henderson, T., Klemp, J., Skamarock, W. & Wang, W. 2004 The weather research and forecast model: software architecture and performance. In: *Proceedings of the 11th ECMWF Workshop on the Use of High Performance Computing In Meteorology* (Vol. 25, p. 29). World Scientific.

Murphy, S., Cinquini, L., Chastang, J., DeLuca, C., Middleton, D. & Balaji, V. 2009 Collaborative Model Metadata Development with ESG/Curator/Metafor. White paper.

ESMF website. http://www.earthsystemmodeling.org/publications/ (accessed 28 February 2014).

Nativi, S., Mazzetti, P. & Geller, G. N. 2013 Environmental model access and interoperability: the GEO ModelWeb initiative. *Environ. Modell. Softw.* **39**, 214–228.

OGC NetCDF 2011 OGC Network Common Data Form (NetCDF) Core Encoding Standard version 1.0. Open Geospatial Consortium. http://www.opengeospatial.org/standards/netcdf (accessed 2 May 2014).

OGC OpenMI 2.0 2014 OGC Open Modelling Interface (OpenMI) Interface Standard. Open Geospatial Consortium Interface Standard. http://www.opengeospatial.org/standards/openmi (accessed 28 August 2014).

OGC WaterML 2.0 2012 OGC WaterML 2.0 Part 1 – Timeseries. Open Geospatial Consortium Implementation Standard. http://www.opengeospatial.org/standards/waterml (accessed 2 May 2014).

OpenMI Association Website 2014 What is OpenMI? https://sites.google.com/a/openmi.org/home/new-to-openmi#TOC-What-is-OpenMI- (accessed 28 February 2014).

Peckham, S. & Goodall, J. 2013 Driving plug-and-play models with data from web services: a demonstration of interoperability between CSDMS and CUAHSI-HIS. *Comput. Geosci.* **53**, 154–161.

Peckham, S., Hutton, E. & Norris, D. 2013 A component-based approach to integrated modeling in the geosciences: the design of CSDMS. *Comput. Geosci.* **53**, 3–12.

Rebora, N., Molini, L., Casella, E., Comellas, A., Fiori, E., Pignone, F., Siccardi, F., Silvestro, F., Tanelli, S. & Parodi, A. 2013 Extreme rainfall in the Mediterranean: what can we learn from observations? *J. Hydrometeorol.* **14**, 906–922.

Silvestro, F., Gabellani, S., Giannoni, F., Parodi, A., Rebora, N., Rudari, R. & Siccardi, F. 2012 A hydrological analysis of the 4 November 2011 event in Genoa. *Natl. Hazard. Earth Syst. Sci.* **12**, 2743–2752.

Syvitski, J. P., Hutton, E., Piper, M., Overeem, I., Kettner, A. & Peckham, S. 2014 Plug and play component modeling – the CSDMS2.0 approach. In: *International Environmental Modelling and Software Society (iEMSs), 7th International Congress on Environmental Modelling and Software* (D. P. Ames & N. Quinn, eds), San Diego, CA, USA. http://www.iemss.org/society/index.php/iemss-2014-proceedings.

Voinov, A., Seppelt, R., Reis, S., Nabel, J. E. M. S. & Shokravi, S. 2014 Values in socio-environmental modelling: persuasion for action or excuse for inaction. *Environ. Modell. Softw.* **53**, 207–212.

Weerts, A. H., Schellekens, J. & Weiland, F. S. 2010 Real-time geospatial data handling and forecasting: examples from Delft-FEWS forecasting platform/system. *IEEE J. Sel. Top. Appl.* **3** (3), 386–394.

Zaslavsky, I., Valentine, D., Hooper, R., Piasecki, M., Couch, A. & Bedig, A. 2012 Community practices for naming and managing hydrologic variables. In: AWRA 2012 Spring Speciality Conference, New Orleans, LA.