

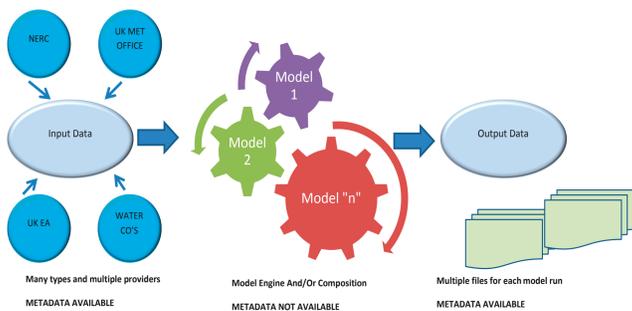
Improving the accessibility and re-use of environmental models through provision of model metadata — a scoping study

Riddick, A¹, Hughes, A¹, Harpham, Q², Royse, K¹ and Singh, A¹

¹ British Geological Survey, United Kingdom ² HR Wallingford, United Kingdom

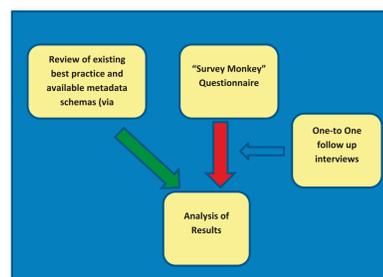
A. The problem

This poster presents the results of a scoping study funded under a recent Natural Environment Research Council (NERC) Environmental Data Call. The work was undertaken by the British Geological Survey (Nottinghamshire, UK) in collaboration with HR Wallingford (Oxfordshire, UK). This investigation was designed to better understand the problem that whilst the input data used for modelling frequently has metadata data available, and metadata is often routinely created for the datasets created by modelling, there was perceived to be a lack of schemes and systems to record metadata about the modelling process itself. From this analysis gaps in metadata provision were identified, and recommendations for further work to address these were identified.



B. Our approach

An online survey was constructed using Survey Monkey to capture the views of a wide spectrum of stakeholders concerning how they are currently managing metadata for integrated environmental modelling (IEM) and what gaps exist. A total of 108 responses to the survey were received over a four week period. The majority of the respondents held senior positions in their organisations giving weight to the findings of the study. In order to confirm and validate key trends, one to one interviews were conducted with selected individuals who had completed the Survey Monkey questionnaire. In parallel with this user consultation exercise an analysis of current best practice in the use of metadata for data and models was also undertaken.



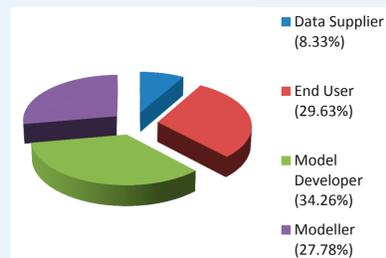
EXAMPLE SURVEY MONKEY MULTIPLE CHOICE QUESTION

3.4 When working with environmental models what metadata or other supporting information is most important to enable you to make effective use of the model(s)? Please rank the options below in relative importance (High, Moderate or Low) using the radio buttons

	High	Moderate	Low
The details of the software or model code used to create the model	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The datasets used as inputs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The models used as inputs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
File formats available (e.g. .dat, excel, oracle etc) for input datasets?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
File formats available for output datasets?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Compatible model coupling technologies (e.g. OpenMI)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Additional toolkits (e.g. for visualising results)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Assumptions made in building the model	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Descriptive information (title, abstract etc)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Parameters or phenomena represented (e.g. wind direction)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Provenance (e.g. how and why it was derived)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

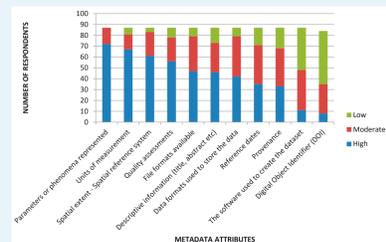
C. Some key results

C1 Scientific Roles Represented



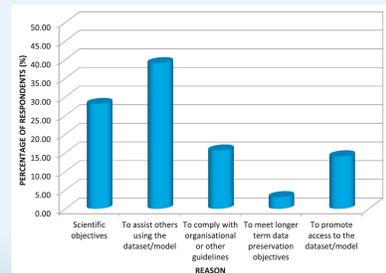
Overall the survey respondents included a small proportion of Data Suppliers, with the remainder split fairly equally between End Users, Model Developers, and Modellers. This suggests that the results represent the views of the main stakeholder categories involved in IEM.

C4 Making use of Data — Relative Importance of Metadata attributes



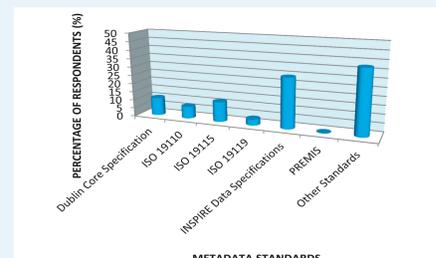
When making use of data the parameters represented and units of measurement together with spatial metadata are viewed as the most important attributes to record. Data and file formats are also seen as reasonably important. Interestingly attributes such as Reference Dates and Provenance (which are often regarded as important by data managers) are seen as less important by modellers.

C6 Primary reasons for providing metadata



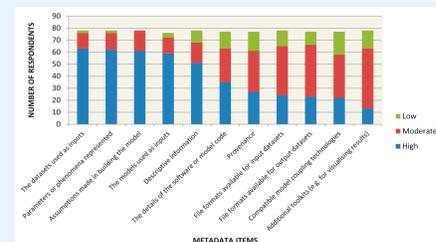
The results suggest that stakeholders view the provision of metadata to assist using models to be more important than for actually finding and accessing the model.

C2 Metadata Standards Applied to Data — and Models

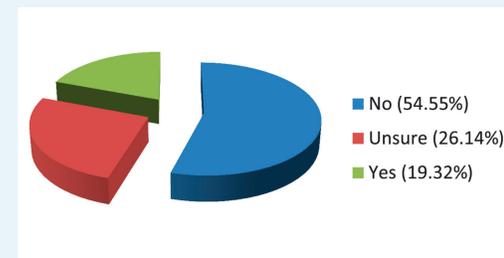


Our user consultation indicates that the usage of the ISO metadata standards (ISO 19110, ISO 19115, and ISO 19119) is relatively small. The 'other standards' category includes the metadata components of Water ML, the Climate and Forecasting Metadata convention and the MEDIN standard.

C5 Making use of Models — Relative Importance of Metadata attributes



C3 Is sufficient metadata data supplied with datasets?



There appears to be greater provision of metadata standards for data, compared to schemes for model metadata. However, the majority of respondents still felt that insufficient metadata is often supplied with datasets used in modelling.

Here information about the datasets used as inputs to the model (e.g. which datasets were used, parameters involved, as well as assumptions made in the modelling) were seen as most important to record in metadata.

D. Summary of current best practice

Metadata for Datasets

- Several metadata schemas available (ISO, Water ML, BUT ... these lack some key attributes e.g. relating to temporal data
- Low take up of ISO schemas by environmental modellers

Metadata for Models

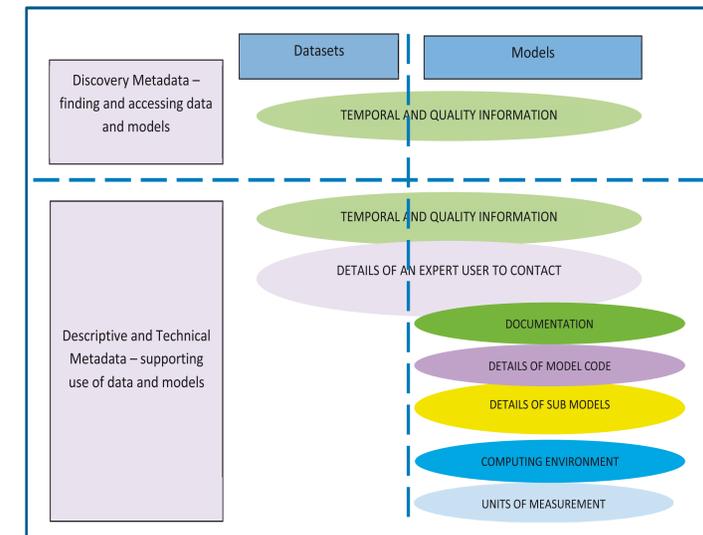
- No widely accepted scheme for model metadata
- Some evidence of locally used schemes which could be built on
- Need for a standard which includes model metadata

Tools for capturing and accessing metadata

- Some catalogues available for datasets - e.g. NERC data catalogue (<http://data-search.nerc.ac.uk>)
- Lack of capture tools linked to modelling workflows

E. Gaps in provision — additional metadata elements recommended

The user surveys indicate that a number of the attributes which feature commonly in discovery metadata schemes for data would also be relevant to creating a metadata profile for models, these include the spatial extent and spatial reference system. At the same time there are a number of attributes which modellers would like to record which do not feature in existing metadata standards for data. These include the facility to describe temporal resolution and scale, and the type of statistical information that might be available over a given time period. Other additional information required by a model metadata schema includes limitations and assumptions of models, as well as estimates of uncertainty.



F. Towards a solution — next steps

