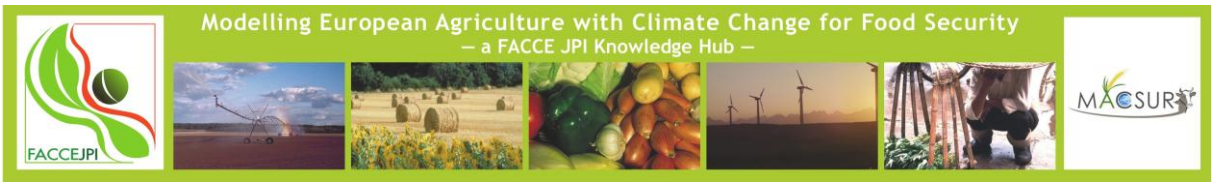

FACCE-MACSUR

D-L1.4.1: Appropriate Meta-data for Modellers

Richard Kipling[1]*, Kairsty Topp[2], Axel Don[3]

[1] Aberystwyth University, King Street, Aberystwyth, Ceredigion, SY23 2BH, United Kingdom
[2] SRUC, West Mains Road, Edinburgh, EH9 3JG
[3] Thünen Instiute of Climate Smart Agriculture, Bundesallee 50, 38116 Braunschweig, Germany

*rpk@aber.ac.uk

Instrument:	Joint Programming Initiative
Topic:	Agriculture, Food Security, and Climate Change
Project:	Modelling European Agriculture with Climate Change for Food Security (FACCE-MACSUR)
Start date of project:	1 June 2012
Duration:	36 months
Theme, Work Package:	LiveM, WP1
Deliverable reference num.:	D-L1.4.1
Deliverable lead partner:	SRUC
Due date of deliverable:	10/2014
Submission date:	24/08/2014
Confidential until:	Not relevant

Revision	Changes	Date
1.0	First Release	2014-05-31
1.1	Changes to wording for clarity and alteration of one part of recommendations	2014-09-29
1.1.1	Relabeled from 1.4.2 to 1.4.1, adjusted references in abstract	2015-01-19

1. Abstract/Executive summary

Report D-L1.4.2 provides an overview of the data and related resources available online and through EU funded projects, relating to soil organic carbon (SOC), and carbon sequestration in grasslands in particular. Building on D-L1.4.2, the report presented here discusses how meta-data describing these types of data (and experimental data more generally) can best be presented in an online resource useful to grassland modellers requiring data to use in their modelling work. Identifying the useful categories of meta-data is a necessary precursor to providing such a resource, which could facilitate better communication between modelling and experimental research groups, allowing researchers to more efficiently locate relevant data and to link up with other scientists working on similar topics. A survey among grassland modelling teams and an assessment of online meta-data resources was used to produce recommendations about the meta-data categories that should be included in an online resource. The categories are generic, so that the recommendations can be followed in the design of meta-data resources for the more general agricultural modelling community.

Table of Contents

1. Abstract/Executive summary 1
Table of Contents 1
2. Introduction 2
3. Methods 2
4. Results 3
4.1. Use of soil Carbon data 3
4.2. Value of proposed meta-data types 3
4.3. Additional meta-data types suggested 4
4.4. Additional meta-data types suggested for soil Carbon data 4
4.5. General Comments 4
5. Discussion 6
6. Acknowledgements 7
7. References 8
Appendix 1: Survey forms sent to grassland modellers 9

2. Introduction

In recent years there has been a broad recognition that multi-disciplinary approaches are vital to addressing key societal challenges such as climate change (EC, 2011, Raab et al., 2013). Knowledge exchange between disciplines is an integral part of multi-disciplinary working (Fazey et al., 2014) and can require the development of novel management structures and shared resources. The MACSUR knowledge hub seeks to increase the capacity to model the impacts of climate change in European agriculture (MACSUR, 2012); an aim that will require the formation of closer ties between different modelling groups and disciplines, as well as between modellers and experimental researchers. Data sharing between experimental researchers and modellers can be problematic. Dataset holders may be unaware of the data requirements of modellers, and may not see the benefits of making their results available for use in models. On the other hand, modellers may have quite specific data requirements and be frustrated that required data are not available or that available data is not of the type they need to run and develop their models. This type of communication problem represents a lost opportunity for modelling groups and dataset holders. Modellers would benefit from having access to the most appropriate and recent data, and from working with experimental research groups to ensure that data outputs from experimental work are of a quality and type suitable for modelling, while also meeting the original requirements of the experimental study. Two-way communication between modellers and experimental researchers is beneficial in the modelling phase as misinterpretation of data is less likely to occur. By using new findings and datasets to inform their modelling work, modelling groups can demonstrate how these experimental data might impact agricultural systems over time and at different scales. Experimental researchers would therefore profit from the wider use of their data in modelling, which provides a path for them to gain increased impact from their research outputs. Modelling may also reveal systemic relationships and sensitivities that can help direct the focus of future research. The potential synergy between modelling and experimental research activities can be a spur to improving the integration of these scientific approaches.

3. Methods

In order to gain an understanding of modeller's requirements for meta-data when searching for datasets relevant to their modelling activities, a short questionnaire (Appendix 1) was prepared and sent out to grassland modellers in MACSUR. Twelve grassland modelling teams are represented in the project. The questionnaire asked modellers to rank the usefulness of some basic meta-data types, and to add further categories if required. Different sections gathered this information for general datasets and for soil carbon datasets specifically, in order to ascertain whether meta-data requirements for soil carbon data could be more broadly applied to datasets used in other types of modelling (farm-scale models, livestock models etc.). Information was also collected as to the requirements of grassland models in MACSUR, to understand better what types of dataset would be most useful for experimental researchers within the consortium to make available to their modelling colleagues. Meta-data types were ranked according to their reported usefulness and the results of the survey were compared to existing online meta-data resources. Online resources considered were:

- Meta-data standards used for the World Soil Database (WOSIS), part of ISRIC (World Soil Information). WOSIS is the 'World Data Center' for soil datasets.
- Meta-data profiles used by the European Soil Data Centre (ESDAC).

Recommendations for an online meta-data resource for MACSUR were made on the basis of these investigations.

4. Results

Seven modelling groups returned the questionnaire for grassland modellers, providing information on their preferences for meta-data, and on their use of soil carbon data, which is a return rate of 64%

4.1. Use of soil carbon data

Five of the groups use data on soil carbon stocks and soil carbon fluxes (one of these models uses these types of data for validation). One modelling group requires neither type of data, and one group did not return information in this section of the questionnaire. Three models use carbon data as an input, while four do not. Five of the seven groups use soil carbon data in their general research, as well as in their modelling work.

4.2. Value of proposed meta-data types

Modellers were asked how valuable each of several types of meta-data would be to them when using a meta-data database (Figure 1). All suggested categories were rated as essential or useful by the majority of modelling groups, with a list of recorded variables being considered to be the most important piece of information in a meta-data database. One respondent stated that inclusion of measurement units in the meta-data was not necessary and might cause problems (e.g. data might not be measurement data but descriptions of climate or management).

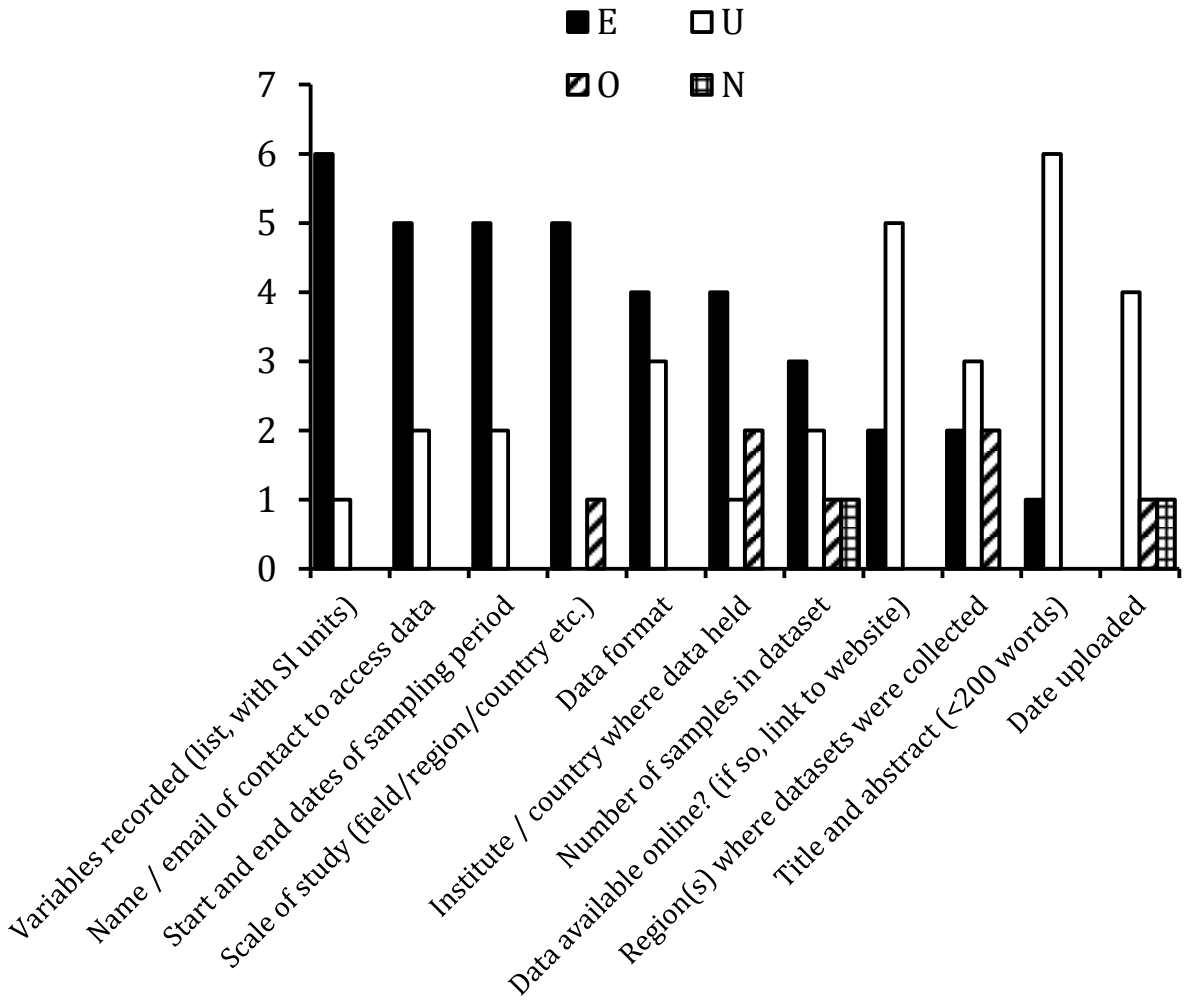

Figure 1: Types of meta-data and survey returns indicating their usefulness. E = essential, U = useful, O = optional, N = not useful.

4.3. Additional meta-data types suggested

Four additional types of meta-data were suggested and considered essential by one respondent: location, data ownership and use license, permanent data set identifier, and version of data set. One piece of meta-data (whether data integrity is checked) was suggested by one respondent but not rated.

4.4. Additional meta-data types suggested for soil carbon data

Respondents suggested that information on 26 additional individual variables (soil depth etc.) should be included in dataset meta-data. One respondent suggested that management information for surveyed sites should be included in meta-data:

> *'There is increasing interest in being able to assess mitigation measures on soil C stocks (e.g. tillage, management of crop residues, fertiliser rates, manures). Data that can be used to validate the models for these management techniques as well as responses to changes in management would be useful'*

One respondent suggested that information on the method for estimating bulk soil density in each layer should be included, while another suggested that information on the methods of measuring the coarse fraction of each layer should be added.

4.5. General Comments

One respondent provided information about two meta-data initiatives that provide useful standards for a modelling meta-data database to adhere to:

> *'The Open Archives Initiative (http://www.openarchives.org) is striving for the standardization of metadata access so that data sets may be aggregated with minimal effort. (Copied from: http://www.cocos-carbon.org/docs/D1.2_data_policy_report.pdf)'*

> *'The World Data System is a central portal and depository for data from national and international monitoring programs administrated by the International Council for Science in Paris, France. As of 2009 the WDS comprises the formerly thematically specialized World Data Centers, including the WDC for Biodiversity and Ecology at the Oak Ridge National Laboratory. Data hosted by the WDS are available for free via online access for non-commercial purposes. Submission of datasets or links to the portal for hosting or linking requires that metadata are supplied according the WDS standard. Metadata are collected according to the principles of the Directory Interchange Format (DIF, http://gcmd.nasa.gov/User/difguide) and the Service Entry Resource Format (SERF, http://gcmd.nasa.gov/User/serfguide). These formats are in line with the internationally agreed minimal metadata, the Dublin Core Set (http://dublincore.org).'*

The WOSIS meta-data service (one of the 'World Data Centers' referred to in the preceding quote) utilises a long list of meta-data categories. Such detail may not be necessary in relation to the requirements of experimental researchers and modelling groups. However, a debate is required concerning the compatibility of any MACSUR meta-data resource with the types of meta-data standards used by WOSIS (see discussion below).

Categories of meta-data requested via the current survey were compared with those used in ESDAC (Table 1). The exercise indicated a reassuring overlap between the meta-data types identified and those used in this European online resource. A revised list of meta-data types was then produced (Table 1) which (it is hoped) would align a MACSUR meta-data resource with ESDAC while meeting the demands of modellers.

Table 1: Meta-data types identified in the current survey compared to ESDAC, and the proposed categories for a MACSUR meta-data resource

MACSUR SURVEY CATEGORIES	ESDAC CATEGORIES	PROPOSED MACSUR RESOURCE
Title and abstract (<200 words)	Title Abstract Keywords Topic category (for searches)	Title Abstract (including site description) Keywords
Region(s) where datasets were collected Scale of study (field/region/country etc.)	Geographic extent (degree latitude and longitude)	Geographic location of survey (latitude and longitude) Scale of survey (field/region/country)
Date uploaded Start and end dates of sampling period	Date (publication, revision, creation)	Start and end dates of sampling period Date of publication
		Version of data
Data available online? (if so, link to website)	Online resource (web link)	Online resource (web links)
Name / email of contact to access data Institute / country where data held	Responsible party (institute name)	Contact details including country and institute Responsible party (owner of data if contact person unavailable) Restrictions on data use
Variables recorded (list, with SI units) Number of samples in dataset Data format	Quality (free text information on data)	List of measurement variables recorded with SI units List of management information recorded Data presentation (digital map, Excel file etc.)
	Language	Language

5. Discussion

With the exception of the request from modellers for a list of variables examined (which are specific to soil carbon datasets) the meta-data types identified in this survey are generic, and can be a template for wider meta-data resources (for animal disease or farm-scale modelling etc.). Here these meta-data types have been aligned with those used by the ESDAC to create a recommended list of meta-data types that should be included in any MACSUR online resource.

With regard to soil carbon meta-data, COCOS (Coordination Action Carbon Observation System) present a comprehensive list of ECCVs (Essential Carbon Cycle Variables) for carbon datasets of all types (COCOS, 2014). For the purposes of the MACSUR resource, the subset of variables for grassland soil carbon data (suggested by a respondent to the current survey) might be considered sufficient for inclusion. One approach would be to include a section for soil carbon data only, under which would appear a tick list of variables with an option to add more if required. Alternatively, a 'free text' area could be provided in which variables are listed along with the SI units used in the dataset. Because information may not always be in the form of measurement variables (e.g. management information) it will be important that instructions for meta-data entry clearly indicate that both measurement and other types of variable are important and should be included (perhaps through the use of separate sections for each meta-data type).

The aim of this report was to suggest how the meta-data types identified as important by grassland modellers could be incorporated into a MACSUR meta-data resource, using categories comparable to those of the international soil databases already in existence. Two further areas of investigation are important:

1) This report does not consider issues relating to ISO (International Standards Organisation) meta-data standards, which define for example the minimum meta-data types required for compatibility with meta-data services (ISO, 2014). A collective decision needs to be made on the choice of strategy for MACSUR in respect to these aspects of meta-data (specifically, is it sufficient to create a resource which displays the meta-data types required by modellers (described here) or do we also need to direct resources towards meeting meta-data standards to enable wider compatibility?)
   For those with an interest in this, the Open Archives Initiative site (mentioned above) considers these issues (OAI, 2014) and experts also share ideas and develop meta-data standards through groups in the Research Data Alliance (RDA). The RDA website includes details of how to sign up to online discussion groups on meta-data standards (RDA, 2014)
2) This report does not cover the issue of the design of a meta-data resource. Work at hub level is on-going in this respect, and this report is a contribution to that wider effort, including the development of the Agrimod resource (JHI, 2012)

Improving links between experimental researchers and modellers is a key aspect of efforts to increase modelling capacity, in that it ensures that modellers have access to the most recent and appropriate data in their field, while experimental researchers can understand better how the relationships revealed by their data might impact agricultural systems at different scales and over different timeframes. Online resources that allow modellers to identify datasets (and research groups) relevant to their work, and which allow experimental researchers to identify modelling efforts which they could utilise to add impact to their findings, can play an important role in increasing links between these groups of scientists, between disciplines and internationally. The findings above represent

one step in the development of such resources, and it is hoped that this report will help to develop an increased dialogue and understanding between modelling and experimental research scientists.

6. Acknowledgements

We thank Julien Minet, Martin Koechy, Susanne Rolinski, Gianni Bellocchi, Kairsty Topp, Raphaël Martin and Lianhai Wu for responding to the meta-data survey.
This paper is a contribution to the FACCE MACSUR knowledge hub.
The work was funded by BBSRC.

7. References

COCOS. 2014. *COCOS - Coordination Action Carbon Observing System* [Online]. COCOS. Available: http://www.cocos-carbon.org/ [Accessed 21/08/14 2014].

EC. Year. Horizon 2020 - The Framework Programme for Research and Innovation - Impact Assessment Report. *In*, 2011 Brussels.

FAZEY, I., BUNSE, L., MSIKA, J., PINKE, M., PREEDY, K., EVELY, A. C., LAMBERT, E., HASTINGS, E., SUE, M. & REED, M. S. 2014. Evaluating knowledge exchange in interdisciplinary and multi-stakeholder research. *Global Environmental Change,* 25**,** 204-220.

ISO. 2014. *International Standards Organisation: Standards* [Online]. ISO. Available: http://www.iso.org/iso/home/standards.htm [Accessed 21/08/14 2014].

JHI. 2012. *Agrimod: Agricultural modelling knowledge hub* [Online]. James Hutton Institute. Available: http://www.agrimod.org/ [Accessed].

MACSUR. 2012. *MACSUR Overview* [Online]. MACSUR. Available: http://macsur.eu/ [Accessed 12/08/14 2014].

OAI. 2014. *Open Archives Initiative* [Online]. Cornell University Library. Available: http://www.openarchives.org/ [Accessed 21/08/14 2014].

RAAB, J., MANNAK, R. S. & CAMBRÉ, B. 2013. Combining Structure, Governance, and Context: A Configurational Approach to Network Effectiveness. *Journal of Public Administration Research and Theory*.

RDA. 2014. *Research data sharing without barriers: Metadata IG* [Online]. Research Data Alliance. Available: https://rd-alliance.org/internal-groups/metadata-ig.html [Accessed 21/08/14 2014].

Appendix 1: Survey forms sent to grassland modellers

Questionnaire on meta-data required for dataset inventory

We would like to create an inventory of the datasets available for modelling, as a resource for modellers and other experimental research teams.
We would like to know what meta-data you would find most useful to be included in such an inventory (what information would be most effective in helping you decide whether to investigate the dataset further).
We would like to make the inventory simple but effective, by including only the most important types of meta-data for each dataset. In the table on page 2, please indicate for each type of meta-data whether the meta-data type is:
E (essential)
U (useful but not essential)
O (optional – of some limited use)
N (of no real value)

If you add further meta-data types of your own, please indicate their level of value as well. You can also comment on potential issues related to the provision of each type of meta-data (potential for variation in terminology etc.).
Thanks from the Task 1.4 team for your contribution to this survey

TABLE FOR YOUR RESPONSES

Please return this document, including your responses, to rpk@aber.ac.uk by Fri 13th June 2014

Type of Meta-data	**Indicate level of Usefulness (E, U, O, N)**	**Comments / potential issues**
Title and abstract (<200 words)		
Institute and country where data held		
Name and email of contact to access data		
Data available online? (if so, link to website)		
Data format		
Date uploaded		
Region(s) where datasets were collected (county/state and country)		
Variables recorded (list, with the SI units used shown for each variable)		
Scale of study (field/region/country etc.)		
Start and end dates of sampling period		
Number of samples in dataset		
Other meta-data types you think are important	**Indicate level of Usefulness (E, U, O, N)**	**Comments / potential issues**
A		
B		
C		
D		
E		
Meta-data specific to soil carbon datasets	**Indicate level of Usefulness (E, U, O, N)**	**Comments / potential issues**
Soil types surveyed		
Soil depths surveyed (range)		
Other soil carbon related meta-data types you think are important	**Indicate level of Usefulness (E, U, O, N)**	**Comments / potential issues**
A		
B		
C		
D		
Does your model use soil carbon data as an input?	YES/NO	
Do you make use of soil carbon data more generally during your research?	YES/NO	
Do you make use of data on soil carbon stocks (A), soil carbon fluxes (B), both (C)?	A B C	
General Comments:		