

## Australian soil vocabularies: Availability and lessons learnt when using them to describe diverse soil data

M.Wong<sup>1</sup>, L. Gregory<sup>3</sup>, B.Simons<sup>1</sup>, A.MacLeod<sup>1</sup>, H.Gillett<sup>1</sup>, N.Robinson<sup>1</sup>, S.Cox<sup>2</sup>, J.Yu<sup>2</sup>,

<sup>1</sup> Centre for eResearch and Digital Innovation (CeRDI), Federation University Australia, Ballarat VIC 3350 Australia  
mr.wong@federation.edu.au

<sup>2</sup> CSIRO, Clayton VIC 3168 Australia

<sup>3</sup> CSIRO, Black Mountain, Canberra ACT 2601 Australia

### Introduction

Applying the 'FAIR' principles to data (making data Findable, Accessible, Interoperable, and Reusable) promotes the sharing and reuse of data, supports knowledge discovery and innovation, supports data integration and helps data to become 'machine' readable, supporting new discoveries through the harvest and analysis of multiple datasets (Australian Research Data Commons, 2018). As the number of public and private soil data custodians contributing soil data to our existing soil database at CeRDI grew, so too did the types of data including the tests being used, the properties being measured, units of measure, and variations in the ways the one soil measurement could be described. It became clear that improvements were required to our soil database that ensured that soil data are 'FAIR' for the data providers and custodians into the future. One of the steps we took was to describe the soil data using soil domain accepted standards available and overseen by the National Committee on Soil and Terrain. Ideally, these standards would be in a machine-readable format to facilitate interoperability. This involved: 1. Working with partners at CSIRO to make standard Australian soil testing procedures and codes available as published vocabulary services available in standard machine readable format with persistent URLs (web addresses) (see Australian Government Linked Data Working Group, 2018 and W3C, 2014 – Working Document) and 2. Mapping the existing data in our database to these standards.

Here we demonstrate the standard soil vocabularies that we have made available to the soils community. We focus on the challenges encountered when mapping our diverse soil dataset, with multiple contributors, to these vocabularies, as these learnings may be useful for defining the required governance (e.g. through the National Committee on Soil and Terrain) and future application of Australian soil vocabularies and code lists by the soil community.

### Methodology and Results

Strictly adhering to existing standard nomenclatures, we progressed making the following available as "Australian soil vocabularies and code lists" in CSIRO's Linked Data Registry:

1. Soil Chemical Methods, Australasia (SCM) (Rayment and Lyons, 2011);
2. Soil Chemical Methods currently used in CSIRO's SITES database, where the source of the chemical test were not clearly derived from SCM nomenclature, and
3. Soil Profile classifiers from Chapter 8 of the Australian Soil and Land Survey Field Handbook (National Committee on Soil and Terrain, 2009).

The Soil Profile classifiers vocabulary content was prepared by CSIRO using the CSIRO Linked Data Registry service. To upload the Soil Chemical Method vocabulary content to CSIRO's Linked Data Registry, CeRDI used the Excel2LDR in development by CSIRO to assist domain specialists in easily uploading content with the standard desktop tool Microsoft Excel (Yu J, Cox, S, 2018).

We then mapped the soil measurements in our existing soil database to these Australian soil vocabularies and codes made available in CSIRO's Linked Data Registry. Where standard soil vocabularies were not available, additional terms were added to a vocabulary register specifically for the project, adhering to the nomenclatures within Australian soil vocabularies and code lists where possible. Some of the challenges we encountered when mapping the soil data included:

- The various levels of detail that contributors used to describe their data needed to be catered, as examples 'soil pH' down to 'pH of a 1:5 soil/water suspension' and 'texture' (field or lab determined?)

- In some cases, the Australian soil vocabulary terms did not provide the level of resolution required, for example the different Mineral-N forms;
- The resolution to which we could (or needed to) describe data was something we needed to consider, with some Australian soil vocabulary soil test procedures having multiple underlying methods (for example nutrient ratios);
- Catering for superseded methods, lab-specific methods and the modification and extension of standard methods and
- How to represent more recent and emerging soil test procedures, including those with underlying models and validation methods such as soil property predictions by Mid Infrared (MIR) spectra.

### Discussion and Conclusions

Soil Chemical Methods, Australasia and sections of the Australian Soil and Land Survey Field Handbook are being made available as vocabulary services in the CSIRO Linked Data Registry. Since this information is encoded in a standard format that allows for information to be passed between computer applications in an interoperable way (W3C, 2014 – Working Document) describing data with these vocabularies has the potential to help identify and bring data together from heterogeneous data sources (including from contributing farmers, academia and government organisations) to add power to modelling and machine-based learning for the development of knowledge based products such as soil maps and on-farm tools. At the basic level, linking our data with published soil vocabularies and code lists are assisting CeRDI in making our soil services more manageable as the diversity and volume of soil data being contributed continues to grow. It has allowed us to ensure that the soil test data contributed by stakeholders is described with a standard method and unit of measure via a persistent URL (web address). This will assist with the interpretation of the data into the future and will help to ensure the robustness of value-add tools that are delivered as part of this soil service (for example comparisons of change in nutrient status of a farm over time or soil test benchmarking). We foresee that linking data with other related vocabularies, such as land-use and management, could also add value to the analysis and interpretation of soil data for management at the farm, catchment and landscape scale. Our experience and challenges as users suggests there is great value in the active governance (e.g. through the National Committee on Soil and Terrain) of Australian soil vocabularies and code lists by the soil community.

**Keywords:** data, standards, Linked Data, vocabulary

### References

Australian Government Linked Data Working Group (2018). *Showcase* at <http://linked.data.gov.au/showcase.html> . Last accessed 18/06/2018.

Australian Research Data Commons (2018). The FAIR data principles at <https://www.ands.org.au/working-with-data/airdata>. Last accessed 18/06/2018.

National Committee on Soil and Terrain (2009) Australian Soil and Land Survey Field Handbook, 3<sup>rd</sup> edition. (CSIRO Publishing: Collingwood).

Rayment GE, Lyons DJ (2011) Soil Chemical Methods – Australasia. (CSIRO Publishing: Collingwood).

WC3 (2014 – Working document). Best Practices for Publishing Linked Data. *WC3 Working Group Note 09 January 2014* at <https://www.w3.org/TR/ld-bp/> . Last accessed 18/06/2018.

Yu J, Cox S. (2018). Excel2LDR: Lowering the bar to entry for defining vocabularies as Linked Data at <http://www.c3dis.com/1965>. Last accessed 18/06/2018. C3DIS 2018 Conference. Melbourne. Australia.