Implementing FAIR in the Agricultural Research Federation

Megan Wong[1], Kerry Levett[2]

Megan Wong[1], Kerry Levett[2], Helen Thompson[1], Paul Box[3], Bruce Simons[1], Andrew Macleod[1], Rakesh David[4], Derek Schneider[5], Daniel Watkins[5], Rachelle Hergenhan[5], Linda Gregory[3], Peter Wilson[3], Nicolas Taylor[6], Scott Limmer[1] Heath Gillett[1], Simon Cox[3],

1 Federation University Australia, Ballarat {mr.wong, h.thompson, b.simons, a.macleod, h.gillett}@federation.edu.au,
2 Australian Research Data Commons (ARDC), Adelaide, Australia kerry.levett@ardc.edu.au
3 CSIRO, Australia {paul.j.box, linda.gregory, peter.wilson, simon.cox}@csiro.au,
4 ARC CoE in Plant Energy Biology, University of Adelaide, Adelaide, Australia rakesh.david@adelaide.edu.au
5 University of New England, Armidale, Australia {dschnei5, dwatkin5, rhergen4}@une.edu.au
6 ARC CoE in Plant Energy Biology, The University of Western Australia, Perth, Australia nicolas.taylor@uwa.edu.au

BACKGROUND

The Agricultural Research Federation (AgReFed) aims to improve the sharing and reuse of agricultural data, including datasets, metadata and data related products. This is being achieved through bringing together independent partner organisations across research, government, and private sectors and enabling them to make their data Findable, Accessible, Interoperable, and Reusable (FAIR) through participation in, and alignment with a federated community and distributed socio-technical architecture (agrefed.org.au).

To improve the sharing and (re)usability of agricultural data, individual providers' heterogeneous data and provisioning arrangements need to be brought into alignment through AgReFed. One of the key alignment processes (i.e. the ways in which partners' data is harmonised) is the AgReFed FAIR policy proposed in the *Guidelines for the development of a Data Stewardship and Governance Framework for the Agricultural Research Federation (AgReFed)* [1], which is based on the FAIR principles [2]. The proposed AgReFed FAIR policy identifies minimum thresholds of "FAIRness" for data to be included in AgReFed as well as 'stretch' targets. These thresholds are reflected in the 14 question *AgReFed FAIR assessment* [3], which was adapted from the ARDC FAIR data self-assessment tool [4].

Inaugural partners worked together with the Australian Research Data Commons (ARDC) to increase the FAIRness of various agricultural research datasets, data products, and/or services. Data covered a wide range of types (point observation, spatial, temporal, on-ground, sensor and remote sensed) for plant (e.g. yield, crop rotation, metabolomic, proteomic, hyperspectral), soil and climatic variables. This presentation describes how the six inaugural data providers assessed and improved the FAIRness of their data products; and how the learnings from the process were used to improve the AgReFed FAIR assessment itself.

CLIMBING THE FAIR LADDER

Data providers used the ARDC FAIR tool [4] at the beginning of the project to get a baseline assessment of the FAIRness of AgReFed data products. This highlighted where focus is needed to make the data more FAIR. It also informed initial AgReFed FAIR thresholds and stretch targets for the development of the AgReFed FAIR assessment [3].

The exemplar data collections were at very different initial levels of FAIRness, with some achieving high findability and accessibility but low interoperability and reusability, while others were very interoperable and accessible, but not at all findable. Activities were targeted to the areas that could be made more FAIR by the data providers and their organisational support. Such activities (and the FAIR component addressed) included:

- F, A, R - All providers creating or updating metadata records (of data and services) to machine-readable AgReFed community acceptable standards discoverable through Research Data Australia and the AgReFed portal. This included licencing and any access conditions and protocols.
- I - Previously unstructured datasets were converted to structured, machine-readable formats.

- I - Partners used data exchange standards where feasible (e.g. Open Geospatial Consortium (OGC) Observations and Measurements, SensorThings API, ANZSoilML).
- F, I, R - Teams worked with ARDC, CeRDI and CSIRO to describe data within their databases with machine-readable controlled-vocabularies, hosted by Research Vocabularies Australia (RVA) or elsewhere. If required, providers developed machine-readable controlled vocabularies (e.g. http://registry.it.csiro.au/def/soil/au/asls) and worked to make these discoverable through RVA.
- A, R - Data providers worked with their organisational IT support to host and deliver data to AgReFed via an interface that suited their organisation's current arrangements and aspirations including OGC Standards (WFS, WMS, WCS) or other delivery solutions such as hybrid APIs (e.g. PostgREST/RESTful).
- F, A, R - The ability of repositories to deliver continued access to quality data was assessed using a subset of the CoreTrustSeal Data Repositories Requirements.

FINDINGS AND FUTURE

- Use cases for the data, articulated by data providers, were achieved through making significant gains in improving the FAIRness of data, metadata and services. These included increasing discoverability of their agricultural research data; accessibility of data in line with data providers' wishes; efficiency gains in data management; improving remote and dynamic student learning experience; the ability to search, select, combine and add to large data collections; and the ability to search and combine across data collections.
- 'Easy wins' in improving findability, accessibility and reusability were made through improving content and discoverability of metadata records. Representing the relationship between collections/multiple datasets and services through metadata records was a challenge.
- For those data providers storing their data collections across several spreadsheets or files, improving the structure of data for accessibility was an achievable step with project support. Storage and delivery options for data required multiple solutions to suit the research groups' everyday business and institutional support capability and requirements.
- Improving the interoperability of data through domain accepted delivery standards and machine-readable vocabularies was achieved to varying degrees by providers, and is a work in progress for others. Which standards to employ (including vocabularies), evaluating the governance of these standards and upskilling data providers as they moved from 'users' to 'creators' presented areas for future targeted education and upskilling.
- Data providers encountered several challenges evaluating their data products. These included understanding what data products needed to be assessed and why; for example, services, collections, or individual datasets? Also, what defines 'brief' and 'comprehensive' metadata description, and what defines, for example, acceptable 'machine readable' data formats? Definitions and specific examples were incorporated in the AgReFed FAIR assessment. This assessment will continue to be informed by AgReFed data providers' feedback for use as an educational, assessment and alignment tool for AgReFed FAIR data.
- The six agricultural data providers' exemplars demonstrated variations in FAIR starting points, incentives/disincentives for providing data, and institutional support and resourcing. Such variation is likely to be reflected across the diverse holdings of research data in the agricultural sector and presents a challenge in making agricultural data FAIR through a federated community. Therefore, as outlined by the AgReFed Framework [1], the FAIR Policy and thresholds will continue to be developed by data provider communities participating in the AgReFed through a representative Federation Council, Technical Committee and Data Steward and Standards roles.

REFERENCES

1 Box, Paul; Levett, Kerry; Simons, Bruce; Wong, Megan. Guidelines for the development of a Data Stewardship and Governance Framework for the Agricultural Research Federation (AgReFed). Sydney: CSIRO; 2019. https://doi.org/10.25919/5cf179ba35db9

2 Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. http://dx.doi.org/10.1038/sdata.2016.18

3 AgReFed FAIR assessment. https://www.agrefed.org.au/FAIRassessment (Accessed 6 June 2019).

4 ARDC FAIR data self-assessment tool. https://www.ands-nectar-rds.org.au/fair-tool (Accessed 6 June 2019).