

# Implementing FAIR in the Agricultural Research Federation

Dr Megan Wong

Federation University Australia



Kerry Levett

Australian Research Data Commons

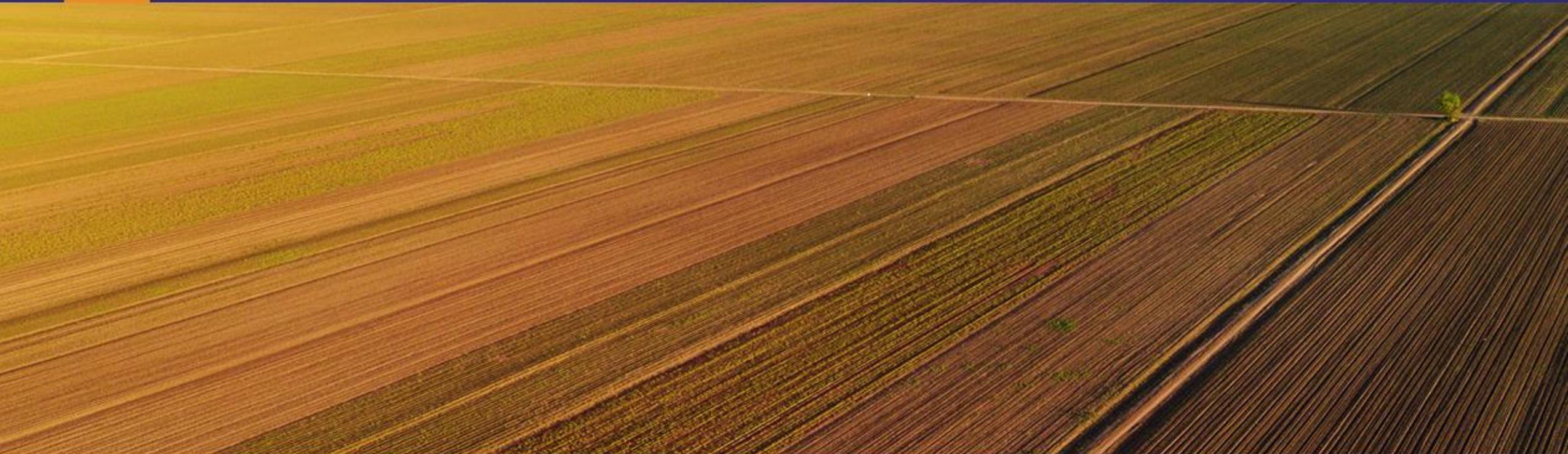


Australian Research Data Commons

Box, P., Simons, B., Thompson, H., Macleod, A., David, R., Schneider, D.,  
Watkins, D., Hergenhan, R., Gregory, L., Wilson, P., Taylor, N., Limmer, S.  
Gillett, H., Simon Cox



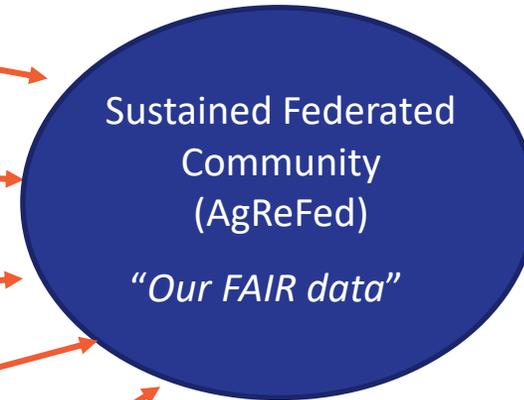
CC BY 4.0 International



Making the most of agricultural data for research

# 'My data' to 'Our FAIR data'

Autonomous data provider communities  
delivering data to an Agricultural Research Cloud  
'My data'



*Representation*

*Delegation*

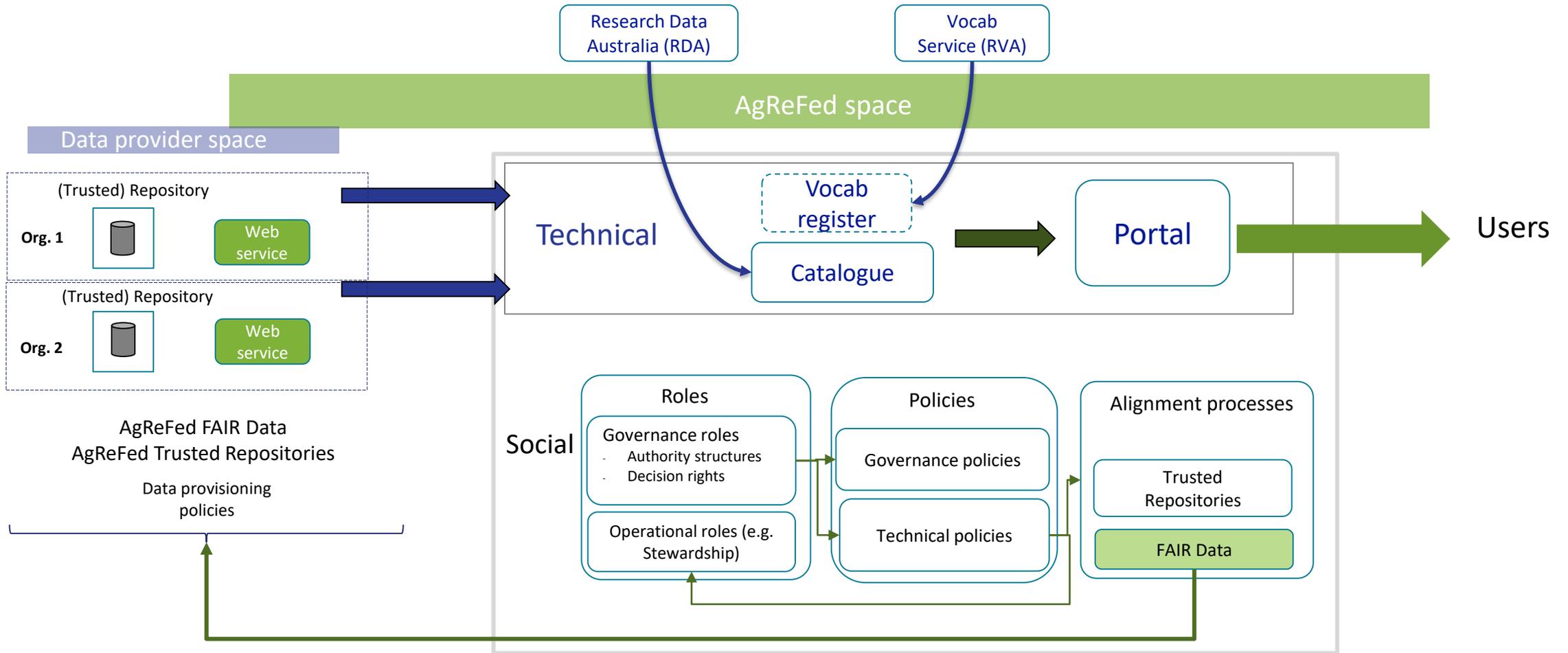


## The data (so far)

- Rotational crop trials
- Frost nurseley trials
- Soil moisture sensors and networks
- Weather station sensor network
- National soil data and services
- Soil sample dataset

# AgReFed – A socio-technical system

encompassing provider communities, roles, policies and alignment processes for enabling the discovery and (re)usability of agricultural data



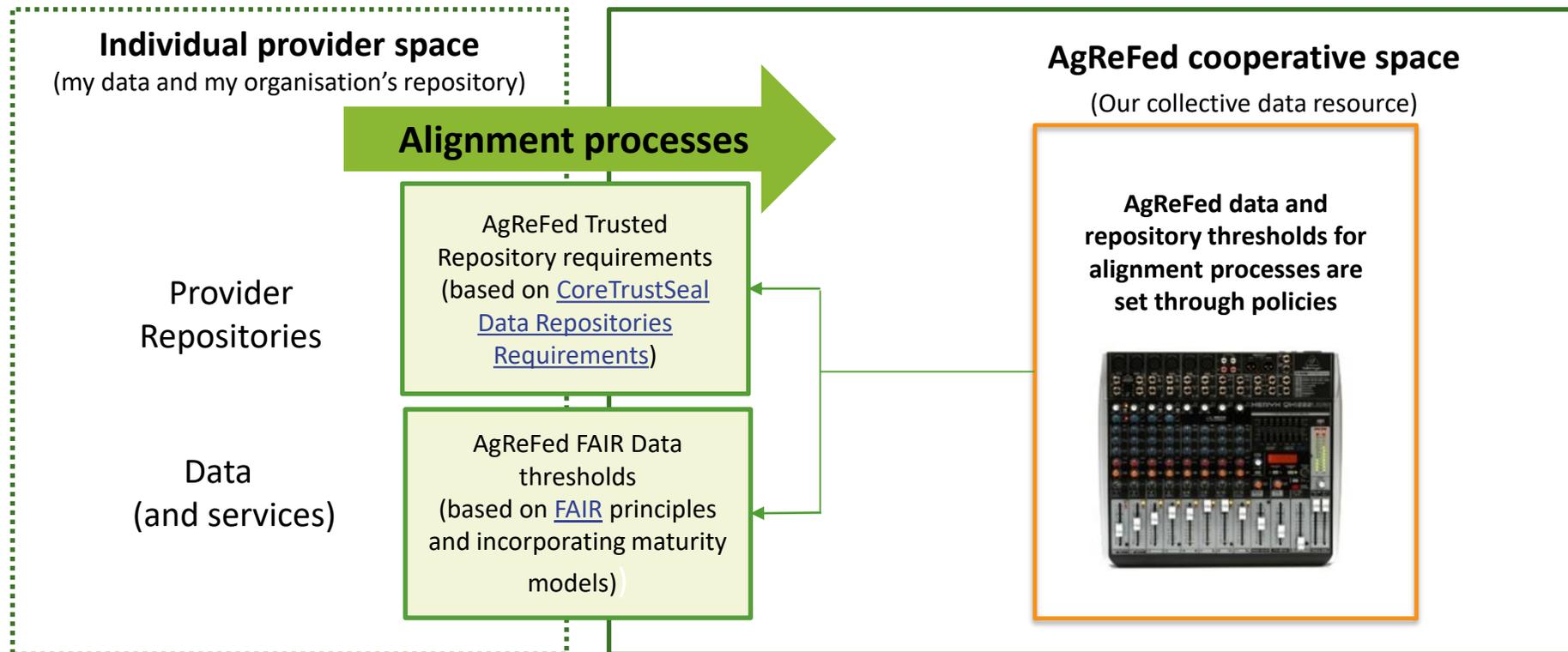
Guidelines for Governance and Data stewardship:

<https://doi.org/10.25919/5cf179ba35db9>

# Alignment processes and policies

From 'my data' to Our FAIR data

Individual providers' heterogeneous data and provisioning arrangements can be brought into **alignment** with agreed levels of FAIRness and repository trustedness



# AgReFed FAIR Data Policy

- AgReFed FAIR Data policy is based on the [FAIR<sup>1</sup>](https://www.go-fair.org/fair-principles/) principles.
- The AgReFed FAIR Data Assessment is based on the [ARDC FAIR Data Self-assessment Tool<sup>2</sup>](https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool/). Modified based on user feedback to improve usability.

Principle (for AgReFed)	Increasingly FAIR →					
<b>FINDABLE</b>						
Q1 The data product has been assigned (an) identifier(s)	No identifier	Local identifier	Web address (URL)	Globally unique, stable and persistent identifier (e.g. DOI, PURL, or Handle)		
Q2 The data product identifier is included in all metadata records/files describing the data	No	Yes				
Q3 The data product is described by a metadata record that facilitates discovery, access and reuse of the data.	The data is not described	Brief title and description	Brief title and description, and multiple other fields filled out, albeit briefly.	Comprehensively (including all AgReFed required fields*) using a formal machine-readable metadata schema.		
Q4 The data product is described by a metadata record that is indexed in a searchable registry or repository...	The data is not described in any registry or repository	Local institutional repository	Domain-specific repository	Generalist public repository	Data is in one place but discoverable through several places (i.e. other registries, RDA, Google Data Search)	
<b>ACCESSIBLE</b>						
Q5 How accessible is the data? The access method(s) must be explicitly stated in the metadata record, e.g. if any authentication is needed, or there are any restrictions to access.	No metadata record	Access to metadata only	Unspecified access conditions e.g. "contact the data custodian to discuss access"	Embargoed access after a specified date	A deidentified version of the data is publicly accessible	Fully accessible public, or to persons who meet and follow explicitly stated conditions and processes, e.g. ethics approval for sensitive data
Q6 Data are available for reuse via a standardised communication protocol, such as file download over https, or a web service.	No access to data	By individual arrangement	File download from online location	Non-standard web service (e.g. OpenAPI/Swagger/Informal API)	Standard web service API (e.g. OGC)	
Q7 The repository/registry agrees to maintain the persistence of the metadata record, even if the data product is no longer available.	No (or not applicable, if no metadata record exists)	Unsure	Yes			
<b>INTEROPERABLE</b>						
Q8 The data products are available in (an) open (file) format(s)	Data are mostly available only in a proprietary format	Data are available in an open format	Data are available in an open, documented, widely used standard format (i.e. NetCDF, CSV, JSON, XML, etc)			
Q9 The data is machine readable (see Glossary for definition)	The data are unstructured	The data are structured and machine readable (i.e. csv, JSON, XML, RDF, database files, etc)				
Q10 The data are semantically interoperable, because they use standard, accessible ontologies and/or vocabularies to describe the data elements/variables.	Data elements are not described (i.e. fields or objects are labelled with codes or not at all)	Data elements are described (so that a human user can correctly interpret the data), but no standards have been used in the description	Published vocabularies / ontologies / schema (without global identifiers) are used	Published vocabularies using resolvable global identifiers linking to explanations, are used, so that the data can be read and understood by machines as well as humans.		
Q11 The relationships to other data and resources (e.g. related datasets, services, publications, grants, etc) are described in the metadata or data, to provide context around the data.	There are no links to other metadata or data	The metadata record includes URL links to related metadata, data and definitions	Qualified links to other resources are recorded in a machine-readable format, e.g. a linked data format such as RDF			
<b>REUSABLE</b>						
Q12 Machine-readable data licenses are assigned to each data product, and are stated in the metadata record.	No license is applied	Non-standard license applied, without a license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record	Non-standard license applied, WITH the license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record	Standard license applied (e.g. Creative Commons), without a license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record	Standard license applied (e.g. Creative Commons), WITH the license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record	
Q13 The provenance of the data product is described in the metadata, i.e. project objectives, data generation/collection (including from external sources) and processing workflows.	No provenance information is recorded	Partially recorded	Comprehensively recorded in a text format (i.e. TXT or PDF)	Comprehensively recorded in a machine-readable format (i.e. in metadata record's schema or PROV, or in RDF, JSON, NetCDF, XML, etc)		
Q14 The preferred citation for the data product is provided in metadata record	No	Citation does not include identifier	Citation includes identifier			

## Qualifying thresholds:

The green cells indicate the proposed minimum acceptable level that data must comply with before it can be 'published' as AgReFed Data

- Where different shades of green are shown, the lightest green indicates minimum acceptable level, and the darkest green indicates stretch goal

1: <https://www.go-fair.org/fair-principles/>

2: <https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool/>

# AgReFed FAIR Assessment - Findable

Principle (for AgReFed)	Increasingly FAIR -->				
<b>FINDABLE</b>					
Q1 The data product has been assigned (an) identifier(s)	No identifier	Local identifier	Web address (URL)	Globally unique, citable and persistent identifier (e.g. DOI, PURL, or Handle)	
Q2 The data product identifier is included in all metadata records/files describing the data	No	Yes			
Q3 The data product is described by a metadata record	The data is not described	Brief title and description	Brief title and description, and multiple other fields filled out, albeit briefly.	Comprehensively (including all AgReFed required fields*) using a formal machine-readable metadata schema.	
Q4 The data product is described by a metadata record that is indexed in a searchable registry or repository	The data is not described in any registry or repository	Local institutional repository	Domain-specific repository	Generalist public repository	Data is in one place but discoverable through several places (i.e. other registries, RDA, Google Data Search)
<b>ACCESSIBLE</b>					
<b>INTEROPERABLE</b>					
<b>REUSABLE</b>					

\* Q3 - Minimum metadata requirements were specified. See <https://doi.org/10.25919/5cf179ba35db9>

# AgReFed FAIR Assessment - Accessible

Principle (for AgReFed)	Increasingly FAIR -->				
<b>FINDABLE</b>					
<b>ACCESSIBLE</b>					
Q5 How accessible is the data? The access method(s) must be explicitly stated in the metadata record, e.g. if any authentication is needed, or there are any restrictions to access.	No metadata record	Access to metadata only	Unspecified access conditions e.g. "contact the data custodian to discuss access"	Embargoed access after a specified date; or A deidentified version of the data is publicly accessible	Fully accessible public, or to persons who meet and follow explicitly stated conditions and processes, e.g. ethics approval for sensitive data
Q6 Data are available for reuse via a standardised communication protocol, such as file download over https, or a web service.	No access to data	By individual arrangement	File download from online location	Non-standard web service (e.g. OpenAPI/Swagger/informal API)	Standard web service API (e.g. OGC)
Q7 The repository/registry agrees to maintain the persistence of the metadata record, even if the data product is no longer available.	No (or not applicable, if no metadata record exists)	Unsure	Yes		
<b>INTEROPERABLE</b>					
<b>REUSABLE</b>					

# AgReFed FAIR Assessment - Interoperable

Principle (for AgReFed)	Increasingly FAIR -->			
<b>FINDABLE</b>				
<b>ACCESSIBLE</b>				
<b>INTEROPERABLE</b>				
Q8 The data products are available in (an) open (file) format(s)	Data are mostly available only in a proprietary format	Data are available in an open format	Data are available in an open, documented, widely-used standard format (i.e. NetCDF, CSV, JSON, XML, etc)	
Q9 The data is machine readable (see Glossary for definition)	The data are unstructured	The data are structured and machine-readable (i.e. csv, JSON, XML, RDF, database files, etc)		
Q10 The data are semantically interoperable, because they use standard, accessible ontologies and/or vocabularies to describe the data elements/variables.	Data elements are not described (i.e. fields or objects are labelled with codes or not at all)	Data elements are described (so that a human user can correctly interpret the data), but no standards have been used in the description	Recognised standards have been used in the description of data elements, but no published vocabularies with resolvable URIs	Published vocabularies using resolvable global identifiers linking to explanations are used, so that the data can be read and understood by machines as well as humans.
Q11 The relationships to other data and resources (e.g. related datasets, services, publications, grants, etc) are described in the metadata or data, to provide context around the data.	There are no links to other metadata or data	The metadata record includes URI links to related metadata, data and definitions	Qualified links to other resources are recorded in a machine readable format, e.g. a linked data format such as RDF	
<b>REUSABLE</b>				

# AgReFed FAIR Assessment - Reusable

Principle (for AgReFed)	Increasingly FAIR -->				
<b>FINDABLE</b>					
<b>ACCESSIBLE</b>					
<b>INTEROPERABLE</b>					
<b>REUSABLE</b>					
Q12 Machine-readable data licenses are assigned to each data product, and are stated in the metadata record.	No license is applied	Non-standard license applied, without a license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record	Non-standard license applied, WITH the license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record	Standard license applied (e.g. Creative Commons), without a license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record	Standard license applied (e.g. Creative Commons), WITH the license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record
Q13 The provenance of the data product is described in the metadata, i.e. project objectives, data generation/collection (including from external sources) and processing workflows.	No provenance information is recorded	Partially recorded	Comprehensively recorded in a text format (i.e. TXT or PDF)	Comprehensively recorded in a machine readable format (i.e. in metadata record's schema or PROV, or in RDF, JSON, NetCDF, XML, etc)	
Q14 The preferred citation for the data product is provided in metadata record	No	Citation does not include identifier	Citation includes identifier		

# Experience of making data FAIR

- **Easy wins** – creating and improving metadata records
- **A bit harder**
  - Structuring data for accessibility – needed some support
    - e.g. Unstructured spreadsheets into databases
    - e.g. Integrating multiple data types
  - Storage and delivery options - multiple solutions to suit the research groups' everyday business and institutional support
    - Data and metadata standards used where possible
- **Hardest** – Semantic interoperability!
  - Controlled vocabularies: Stretch goal for research groups, achievable for data providers in soils
- Variation in incentives/disincentives for providing data and making it FAIR
- Variations in institutional support and resourcing

**On the experience of making data and services FAIRer using the FAIR data assessment:**

*“Incredibly useful”*

*“Helped us understand what FAIR is and how important it is to data”*

*“At a conference I advocated for FAIR data to help advance (agricultural) research”*

*“The process has increased my knowledge of FAIR. How FAIR is FAIR and how FAIR you want it to be? The benchmarking and levels have been good. No tool is perfect”*

*“We were doing a good job around Interoperability, and this process showed us how we could easily improve Findability and Accessibility.”*

*“Improving accessibility to our services.....the FAIR tool certainly helped”*

*(re. institutional data plan) “We are not using a FAIR assessment tool, but probably could”*

*“Quite subjective. But reassessment was easier because FAIR tool had been updated, and made more nuanced.”*

## Suggestions for improvement

### Repeatability as an assessment tool

*“A little ambiguous. I’d give different answer on a different day. Sitting and doing it with you helped me dig a little deeper regarding how to answer the questions”*

*“Making the FAIR tool a bit FAIRer! Capturing the answers as metadata”*

### What is being assessed and why

*“Guidance for providers to decide and describe what data products needed to be assessed and why; for example, services, collections, or individual datasets”*

*“Helping partners understand what they are assessing and why is still a bit of a challenge”*

### Interpretation

*“It would be good for online tool to save results; and spit out suggestions for how to make the dataset FAIRer”*

*“Some more info around FAIR vs Open”*

### Support needed

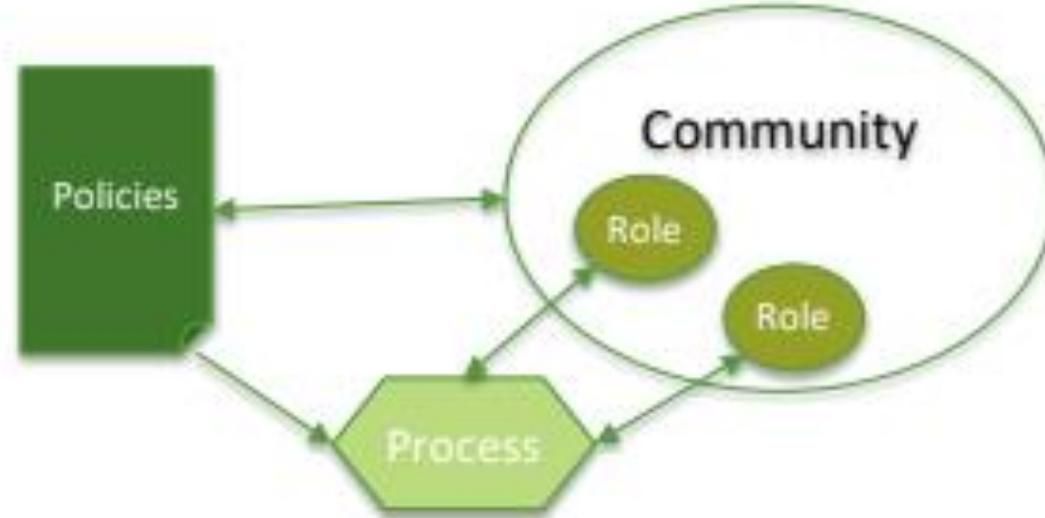
*“Definitely needed ARDC assistance to complete - needs examples of different levels of FAIRness to be clearer.”*

*“Essential to have one-on-one time with someone who really understands the FAIR process, and who also can understand the research process/topic.”*

*“Difficult when data scientists (or data managers) and researchers don't speak the same language, to get the right info out of each other.”*

***How FAIR is FAIR, How FAIR do we want it to be?.....***

# Incorporating FAIR settings into AgReFed policy and processes



*“How FAIR is FAIR?”*

- Tested and further developed within the socio-technical framework of AgReFed
- How it works for provider communities,
- How it is implemented and works with roles, processes and other policies (e.g. Membership Policies, Strategic Policies, Data Steward role)

Guidelines for Governance and Data stewardship:

<https://doi.org/10.25919/5cf179ba35db9>

This project is supported by the Australian Research Data Commons (ARDC). The ARDC is enabled by NCRIS.



Australian Research Data Commons



Project Support and Stewardship Leads:



and

Centre for eResearch and Digital Innovation



CSIRO Environmental informatics

Participating partners:



UNE Smart Farms



National soil data and information



THE UNIVERSITY OF WESTERN AUSTRALIA



Department of Primary Industries and Regional Development

ARC Centre of Excellence in Plant Energy Biology



THE UNIVERSITY of ADELAIDE



School of Agriculture Food and Wine

Waite Research Institute

Australian Plant Phenomics Facility

