The Agricultural Research Federation (AgReFed) Technical and Information Policy Suite

Version 1.0. Endorsed by the AgReFed Council at 13 May 2020.

Citation

MacLeod, A., Wong, M., Gregory, L., Schneider, D., Williams, A., Castleden, I, Simons, B., Levett, K., Box, P., (2020). The Agricultural Research Federation (AgReFed) Technical and Information Policy Suite. Version 1.0. Endorsed by the AgReFed Council at 13 May 2020. The Agricultural Research Federation. DOI: 10.5281/zenodo.3993784

Acknowledgements

This research was supported by the Australian Research Data Commons (ARDC). ARDC is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy Program (NCRIS).

Enactment project leads: Federation University Australia and the Commonwealth Scientific and Industrial Research Organisation. Partner organisations provided valuable input into this document: the University of Adelaide and Australian Plant Phonemics Facility, the University of Western Australia, the Western Australian Department of Primary Industries and Regional Development (DPIRD), the University of New England.

Version Control

AgReFed Technical Policy Suite	Policy	Date	Action
v1.0	Full	10/10/2019	Presented to Technical Committee for development by AgReFed enactment project team
v1.0	Full	17/02/2020	Tabled to Council by Technical Committee
v1.0	1.1.1. Acceptable Data	17/2/2020	'Agriculturally-relevant research data' changed to 'any data that is relevant to agricultural research'.
v1.0	Full	13/5/2020	v1.0 release - Approved by Council

Contents

Table of Contents

Purpose

Provider Communities that are members of AgReFed are independent, autonomous entities who retain control over their own data, data services, repositories and operations. This suite of Technical Policies specifies the technical requirements for Data Provider Communities to follow for membership to AgReFed. These policies cover the technology choices, data policy and standards, including the AgReFed Trusted Repository Policy and AgReFed FAIR[1] Data Policy and the initial qualifying thresholds, for data to qualify as AgReFed data.

The Technical Committee provides recommendations or decisions on the policies herein to the Federation Council in line with the Technical Committee's Terms of Reference. How these Technology Policies are implemented through AgReFed is the business of 'Steering' policies and processes to achieve the collective AgReFed vision and is beyond the scope of this Technical Policy Suite.

Background and Scope

AgReFed was conceived as a federated socio-technical system that enables participating organisations to collaborate to make their data Findable Accessible Interoperable and Reusable (FAIR) in a persistent sustainable manner, from trusted points of truth operated by them as outlined in the Data Stewardship and Governance framework guidelines[2].

The governance model requires a set of technical policies for the alignment of individual providers' heterogeneous data and provisioning arrangements with community agreed AgReFed levels of FAIRness and repository trustedness to achieve the collective vision of AgReFed.

This technical policy suite is based on recommendations of the Guidelines for the development of a Data Stewardship and Governance Framework for AgReFed. The guidelines specify a set of technical guidelines for the exemplar data providers based on 'Reference Model of Open Distributed Processing' (RM-ODP), a reference model that provides five architectural 'Viewpoints' for specifying distributed information systems[3]. The **Enterprise Viewpoint** focuses on the purpose, scope and policies of AgReFed and describes the business requirements and how to meet them. Whilst related to the Viewpoints within the Technical Policy suite, they are outside of the scope of the technical domain. The Viewpoints described by the Technical Policy are:

Information Viewpoint – the scope and nature of the data content (the information architecture). The initial documentation focused on the type of data (agricultural research) and the requirements for metadata content, repository trustworthiness and 'FAIRness'; that is, that data is made more findable, accessible, interoperable and re-usable[4].

Technical Viewpoints – covering the **Computational**, **Engineering** and **Technology Viewpoints**, it specifies the configuration of the system (the technical architecture).

a. Information Viewpoint

The Information Viewpoint specifies:

1. What constitutes acceptable data for sharing through AgReFed;
2. Metadata requirements, including recommended and mandatory requirements;
3. What the acceptable levels of AgReFed Trusted Repository requirements are;
4. What the acceptable levels of FAIRness for publication are;

[1] The FAIR Data Principles.
[2] Box *et al.* (2019)
[3] https://infogalactic.com/info/RM-ODP
[4] Stall, S., et al. (2018).

5. What common information models (data structures) will be supported by AgReFed;
6. What the agreed semantics are, that is, the provider specific and agreed community vocabularies and ontologies.

b. Technical Viewpoints

Computational Viewpoint

The Computational Viewpoint specifies:

1. What the common infrastructure elements are (such as resource discovery mechanisms and portals);
2. What computational interfaces are supported;
3. Expected functionality of common interfaces (discovery portal).

Engineering and Technology Viewpoints

The Engineering and Technology Viewpoints specify:

1. What components (services, software) are to be deployed;
2. The technology which has been used to build and support the existing infrastructure.

1. Policies

1.1. Information Policies

These policies are designed to define what the acceptable data, FAIRness, metadata levels, repository types, structures and semantics are as outlined above.

1.1.1 Acceptable Data

Acceptable data are the digital agricultural data collection(s) that Data Provider Communities wish to share with the Federation. Non-agricultural data, or data that a Provider Collection Custodian does not manage and represent, is not within the scope of the information policy.

The initial focus was on the AgRDC project exemplar datasets, which include crop yield, rotation information, weather and climate, hyper- and multispectral imagery, molecular analysis, and soil measurements (from sensors and other sources). Appendix 1 specifies the data content (features and properties of features) currently available through and of interest to AgReFed. Levels of quality regarding the data itself have not been mandated beyond ensuring it is well described in metadata. Ongoing "fit for purpose" assessments of data quality may occur (and should be encouraged) but this does not preclude initial inclusion.

The full scope of data to be made available is much broader, and could be **any data from Agriculture production, or any data that is relevant to Agricultural research**, as defined by the Federation Council.

The data must be findable as outlined in the Metadata Policy. The Trusted Repository Policy and FAIR Data Acceptability Policy define the qualifying levels required for data to be provided, and data must meet any requirements specified in the Information Model (data structure) and Vocabulary and Ontology Policies. This policy will be reviewed quarterly or at any time as appropriate by the Technical Committee for endorsement or otherwise by the Council.

1.1.2. Acceptable Metadata

The FAIR principles indicate that data must be Findable. In creating FAIR digital resources, metadata should be extensive, including descriptive information about the context, quality and condition, or characteristics of the data. Rich metadata underpins data discoverability, accessibility and reuse, it provides confidence to both human and artificial data consumers. Data should be able to be found based on the information provided by its metadata.

Metadata must be entered into, or be harvestable by Research Data Australia (RDA). The collection record must include a link to the AgReFed Portal Service record (key: www.agrefed.org.au) (see Appendix 2 for details).

The content and quality of metadata that is accepted by the Federation for collections and services is specified in Appendix 2.1 and Appendix 2.2 respectively.

This policy will be reviewed quarterly or at any time as appropriate by the Technical Committee for endorsement or otherwise by the Council.

1.1.3. Trusted Repository

In a distributed information system, the data providers' repositories need to be reliable and trustworthy. The CoreTrustSeal[5]certification provides a process whereby custodians can measure repository compliance levels against sixteen identified characteristics of trustworthy repositories, the Core Trustworthy Data Repositories Requirements[6].

The original CoreTrustSeal requirements reflect the characteristics of trustworthy repositories and they therefore specify that all their Requirements are mandatory and equally weighted. In addition, CoreTrustSeal assess each of the requirements against five levels:

0 – Not applicable
1 – The repository has not considered this yet
2 – The repository has a theoretical concept
3 – The repository is in the implementation phase
4 – The guideline has been fully implemented in the repository.

The Federation Trusted Repository Requirements Policy specified in Appendix 3 simplifies the CoreTrustSeal process. It only identifies those requirements the data repository must meet. As such, the Trusted Repository Policy does not conform to or meet CoreTrustSeal specifications.

This policy will be reviewed quarterly or at any time as appropriate by the Technical Committee for endorsement or otherwise by the Council.

1.1.4. FAIR Data Acceptability

The FAIR principles are a set of guiding principles for rendering data and services Findable, Accessible, and Interoperable, with the ultimate goal of ensuring that research objects are Reusable[7,8,9]. They provide for a continuum of increasing Reusability or 'FAIRness', based on community standards and best-practices rather than a prescriptive standard which can be failed.

[5] CoreTrustSeal
[6] Core Trustworthy Data Repository Requirements v01.00
[7] Wilkinson, M., D. *et al.* (2016)
[8] GoFAIR. *Fair Principles*
[9] Mons, *et al.* (2017)

FAIR does not necessarily equal 'Open' or free (i.e. no cost), as there are legitimate reasons why not all data should be openly shared; rather 'FAIRness' requires clear and transparent conditions for access and reuse, including a data licence.

Policies set by the Federation Technical Committee will be used to determine the acceptable level of FAIRness (e.g. requisite level of metadata and data content standardisation) that will need to be met for sharing of that data through the AgReFed. These levels may vary between data types (e.g. genomics, yield, etc.) and will change over time.

The current FAIR Data Assessment is described in (Appendix 4). Data providers can use the AgReFed FAIR Data Assessment to assess data products, and inform how to improve each findability, accessibility, interoperability and reusability component. This tool is an adaptation of the ARDC FAIR Self-assessment Tool[10] to accommodate the complexity of data and data-service relationships identified in AgReFed.

This policy will be reviewed quarterly or at any time as appropriate by the Technical Committee for endorsement or otherwise by the Council.

1.1.5. Supported Information Models (Data Structure for delivery)

Data should be interoperable through the adoption of agreed data structures or information models (schematic interoperability) for data delivery. Information models may be domain specific, such as GroundwaterML2[11], GeoSciML[12] and ANZSoilML[13]. Where relevant and available, the data providers should use community agreed models.

In some cases, a generic information model may be appropriate. For example, a lot of agriculture data is either observational, such as laboratory soil tests, or machine monitoring data. Standard generic information models have the added advantage that they are more easily used by systems outside the agriculture domain. Examples of standard generic models include the ISO/OGC Observations & Measurements model[14] and TimeSeriesML[15].

The Technical Committee will need to assess information models used by providers as to its appropriateness. As such, the Technical Committee may further develop this policy to include best practice requirements for the information models used to make data interoperable through AgReFed. Appendix 5.5 specifies the information models used for structuring Federation data.

This policy, including the community information models currently accepted by AgReFed, will be reviewed quarterly or at any time as appropriate by the Technical Committee for endorsement or otherwise by the Council.

1.1.6. Defined Data Semantics

Where possible, data should be made interoperable through the adoption of agreed vocabularies and ontologies via vocabulary services (semantic interoperability). Vocabulary terms may form part of a flat list of terms, or may form part of a more complex ontology.

Where possible and appropriate, external authoritative governed and managed vocabularies should be used, such as the Quantities, Units, Dimensions and Data Types (QUDT) ontologies and the AGROVOC controlled vocabulary.

---

[10] FAIR self-assessment tool, ARDC
[11] Brodaric, *et al.* (2018)
[12] Richard, S. M and the CGI Interoperability Working Group (2007)
[13] Simons *et al.* (2013)
[14] Cox *et al.* (2010)
[15] Tomkins, J., Lowe, D. (editors) (2016)

Where a required vocabulary does not exist, a vocabulary can be created, stored and managed by a data provider community through, as examples, Research Vocabularies Australia technology stack (RVA/SiSSVOC/PoolParty) and/or CSIRO's Linked Data Registry (LDR).

The Technical Committee will need to assess each vocabulary used by providers as to its appropriateness and governance arrangements. All vocabularies used should be well-governed and persistent. Appendix 6.1 lists the controlled vocabularies used by AgReFed, and Appendix 6.2 outlines the best practice policy for vocabulary and ontology adoption and creation of vocabulary used by AgReFed.

Modifications and further best practice requirements for the contents, display, construction, testing, maintenance, governance and relationships between controlled vocabularies and ontologies used to make data interoperable through AgReFed may be defined in this policy.

This technical policy will be reviewed quarterly or at any time as appropriate by the Technical Committee for endorsement or otherwise by the Council.

1.2. Computational Policies

The Computational Viewpoint specifies:

1. What the common infrastructure elements are (such as resource discovery mechanisms and portals);
2. What computational interfaces are supported;
3. Current functionality of common interfaces (discovery portal).

This computational policy specifies the common infrastructure elements, supported computational interfaces and the required level of functionality which common interfaces should provide. The use cases of data prosumers and how they can be addressed through the AgReFed user experience is provided in Appendix 11.

1.2.1. Approved Common Infrastructure

These include resource discovery mechanisms and portals for accessing the catalogued data.

Appendix 9 lists the currently accepted computing infrastructure. The list is not exhaustive or compulsory, in fact any combination of infrastructure components may be used if they provide data in accepted formats/protocols and interfaces with required functionality.

This policy will be reviewed quarterly or at any time as appropriate by the Technical Committee for endorsement or otherwise by the Council.

1.2.2. Approved Computational Interfaces

Computational interfaces or Application Program Interface (APIs) allow machine-to-machine interoperability between datasets and applications.

Appendix 7 lists the required computational interfaces.

This policy will be reviewed quarterly or at any time as appropriate by the Technical Committee for endorsement or otherwise by the Council.

1.2.3. Portal functionality

Portals include web accessible dashboards and other interactive visualisations of data which allow users to discover and explore data.

Appendix 7 lists the portal functionality (Discovery layer behaviour)

This policy will be reviewed quarterly or at any time as appropriate by the Technical Committee for endorsement or otherwise by the Council.

1.3. Engineering and Technology Policies

1.3.1. Services

The Federation is to use a **federated** approach to data supply. However, where it makes integration easier to achieve, elements of either brokering and/or aggregation patterns will also be used.

- For federated and brokered data, source data resides with the data provider system. Both solutions encourage currency and validity of data.
- A standards based Service Oriented Architecture (https://en.wikipedia.org/wiki/Service-oriented_architecture) is utilised, including metadata cataloguing and vocabulary linking. The metadata catalogues and vocabularies will provide information about the data using standardised terms.
- Data is transformed from services developed using a community application schema, or in the case of aggregation, using a respected standards based aggregation platform which has broad appeal.

Appendix 8 lists the interface transfer standards used by the Federation. Appendix 10 lists the active service end-points currently available via the Federation.

This policy will be reviewed quarterly or at any time as appropriate by the Technical Committee for endorsement or otherwise by the Council.

1.3.2. Technology

The Technical Policy does not specify the repositories, data, or services of individual providers. It does specify the alignment processes required for any assets or collections that are to be contributed into the Federation.

Appendix 9 lists the technology used in the platform chosen to provide the processing, functionality and presentation of information.

This policy will be reviewed quarterly or at any time as appropriate by the Technical Committee for endorsement or otherwise by the Council.

4. References

Box, P., Levett, K., Simons, B., Wong, M. (2019). *Guidelines for the development of a Data Stewardship and Governance Framework for the Agricultural Research Federation (AgReFed)*. CSIRO, Sydney. https://doi.org/10.25919/5cf179ba35db9

Brodaric, B., Boisvert, E., Chery, L. *et al*. (2018). Enabling global exchange of groundwater data: GroundWaterML2 (GWML2). https://doi.org/10.1007/s10040-018-1747-9

CoreTrustSeal. https://www.coretrustseal.org/ . Date of access 10/10/2019.

Core Trustworthy Data Repository Requirements v 01.00 https://www.coretrustseal.org/wp-content/uploads/2017/01/Core_Trustworthy_Data_Repositories_Requirements_01_00.pdf . Date of access 10/10/2019.

Cox, S. (2010). *Geographic Information: Observations and Measurements.* Version 2.0. Copyright Open Geospatial Consortium Inc. 2013. http://www.opengis.net/doc/om/2.0

FAIR Self-assessment Tool. Australian Research Data Commons. https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool/ . Date of access 10/10/2019

GoFAIR. *Fair Principles.* Creative Commons Attribution 4.0 License by GO FAIR https://www.go-fair.org/fair-principles/ . Date accessed 10/10/2019.

Mons, B., *et al.* (2017). *Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud.* Information Services & Use 37.1. https://doi.org/10.3233/ISU-170824 .

Richard, S. M and the CGI Interoperability Working Group (2007). GeoSciML – A GML Application forGeoscience Information Interchange. Conference Paper: digital Mapping Techniques 2006 - Workshop Proceedings, USHS Open File Report 2008. pp 47 - 59.. https://pubs.usgs.gov/of/2007/1285/pdf/Richard.pdf . Date of access 10/10/2019.

RM-ODP: The Reference Model for Open Distributed Processing. http://www.lcc.uma.es/%7Eav/RM-ODP/index-original.html . Date of access 10/10/2019.

Simons, B., Wilson, P., Ritchie, A., Cox, Simon. (2013). *ANZSoilML: An Australian - New Zealand standard for exchange of soil data.* 20th World Congress of Soil Science, Jeju, Korea, 8-13 June 2014. Permalink: http://hdl.handle.net/102.100.100/94776?index=1

Stall, S., et al. (2018). *Advancing FAIR data in Earth, space, and environmental science.* Nature. 520. https://doi.org/10.1029/2018EO109301

Tomkins, J., Lowe, D. (editors). (2016). TimeseriesML 1.0 – XML Encoding of the Timeseries Profile of Observations and Measurements. http://www.opengis.net/doc/is/tsml/1.0 . Copyright 2016 Open Geospatial Consortium. Date of access 10.10.2019.

Wilkinson, M., D. *et al.* (2016). *The FAIR Guiding Principles for scientific data management and stewardship*. Scientific data 3. https://doi.org/10.1038/sdata.2016.18

5. Appendices

Appendix 1: Data currently accepted as AgReFed data.

This table lists the data currently available through and of interest to AgReFed, including the features and the properties of features

The data available through AgReFed		Example Data collection and or service identifier
Features	Properties (attributes)	
Soil, SoilProfile, SoilSpecimen, Observation, SamplingFeature, SpatialSamplingFeature	Relevant soil feature property observations. Example services include: pH, texture, bulk density, Electrical Conductivity, nutrients (P,K,N,S), exchangeable cations, trace elements and Carbon fractions (Mid infrared corrected), Bulk Density (Whole Earth), Organic Carbon, Clay, Silt, Sand, pH Soil Water, pH CaCl2, Available Water Capacity, Effective Cation Exchange Capacity, Depth of Regolith, Depth of Soil, Coarse Fragments	DOI : 10.25955/5c1c6b8f4d8d2 DOI : 10.25955/5cdcff6168a76 DOI : 10.4225/08/55E5165EC0D29
Soil Moisture Probes	soil moisture, soil temperature, soil electrical conductivity and/or air temperature	DOI : 10.25955/5cdcff6168a76
Climate - atmospheric	Soil moisture and soil temperature, precipitation, wind speed, wind direction, barometric pressure, relative humidity and air temperature	DOI : http://dx.doi.org/10.4226/95/5b10d5ca18aef (data soon to be accessible)
Plant variables	year, rotation cereal, grain yield, dry matter, pasture cuts, pasture composition, grass species by cuts,	DOI : 10.4225/08/55E5165EC0D29
	Ground based plant hyper-spectra; UAV Based plant Multispectral Data; plant Proteomic and Metabolomic Data; on-site temperature; on-site climate; plant anethesis; plant harvest index; plant grain yield; grain protein; grain weight	DOI : http://doi.org/10.26182/5cedf001186f3

Appendix 2.1. AgReFed metadata requirements: Collections^

Information type (Links to RIF-CS guide)	**Meaning**
Metadata publisher	The organisation that is contributing the metadata record
Identifier	A unique identifier for the resource, i.e. DOI
Metadata source	The primary/authoritative source of truth for the metadata record, as represented by a URI.
Collection Type	The type of collection being described, i.e. collection, dataset, software, etc
Title	The name or title of the collection, should be descriptive and unique, avoid acronyms.
Parties	A related person or organisation linked to the collection (include ORCID if possible) e.g. creator, owner, manager.
Location	Online location (DOI, Handle or URL) of the **metadata record** OR to **download** the resource
Related Service	Include a link to the AgReFed portal RDA record (workflow TBA^); or to other Services.
Citation	The preferred form for citing a collection to enable data to be referenced.
Access Rights	Collection access conditions. Specify one of: **open**, **conditional** or **restricted**.
Licence	License conditions associated with the collection; a standard licence, e.g. creative commons is preferred.
Description	A summary description of the collection. Provide sufficient information to enable a user to assess suitability of the data for reuse for their purpose.
Subject	Keywords or terms to describe the topic of the resource. Include at least one ANZSRC-FOR code. Additionally, AGROVOC terms should be used.
Spatial coverage*	The geometry for the location the resource relates to.
Temporal coverage*	The time period the resource relates to, in W3C Date/Time Format.
Related information*	Related resources such as publications (via DOIs), websites (via URLs), funding info, etc

* Required if relevant

^ To be harvested into the AgReFed metadata catalogue, collection records must be linked to the AgReFed Portal Service record in Research Data Australia. The Federation Data Steward will work with providers to determine the best method to enable this, which can be done either by:

1) the Collection record provider, by providing a Collection record to the ARDC Registry that links to the AgReFed Portal Service record (key: www.agrefed.org.au) in the <relatedObject> field; or
2) CeRDI (who administer the AgReFed Portal Service record), by linking to the Collection from the AgReFed Portal Service record (in <relatedObject> or <relatedInfo>).

For more information see providing records to Research Data Australia.

Appendix 2.2. AgReFed metadata requirements: Services^

Information type (Links to RIF-CS guide)	**Meaning**
Metadata publisher	The organisation that is contributing the metadata record
Identifier	A unique identifier for the resource, i.e. DOI
Metadata source	The primary/authoritative source of truth for the metadata record, as represented by a URI.
Service Type	The type of service being described, from this list.
Service name	The name or title of the service, should be descriptive and unique, avoid acronyms.
Parties related to this service	A related person or organisation linked to the service (include ORCID if possible) e.g. owner, manager.
Service location	An electronic address (e.g. access URL) where the service may be accessed.
Related Collections	All collections that are related to, or may be accessed by, the AgReFed portal.
Related Service	Include a link to the AgReFed portal RDA record (workflow TBA^); or to other Services.
Access Rights	Service access conditions. Specify one of: **open**, **conditional** or **restricted**.
Description	A summary description of the collection. Provide sufficient information to enable a user to assess suitability of the data for reuse for their purpose.
Subject	Keywords or terms to describe the research focus of the service. Include at least one ANZSRC-FOR code. Additionally, AGROVOC terms should be used.
Spatial coverage*	The geometry for the location the resource relates to.
Temporal coverage*	The time period the resource relates to, in W3C Date/Time Format.
Related information*	Related resources such as publications (via DOIs), websites (via URLs), funding info, etc

* Required if relevant

^ To be harvested into the AgReFed metadata catalogue, service records must be linked to the AgReFed Portal Service record in Research Data Australia. The Federation Data Steward will work with providers to determine the best method to enable this, which can be done either by:

3) the Service record provider, by providing a Service record to the ARDC Registry that links to the AgReFed Portal Service record (key: www.agrefed.org.au) in the <relatedObject> field; or
4) CeRDI (who administer the AgReFed Portal Service record), by linking to the Service from the AgReFed Portal Service record (in <relatedObject> or <relatedInfo>).

For more information see providing records to Research Data Australia.

Appendix 3: Trusted Repository Requirements

Requirements the data repository must meet (R). Each of these requirements are assessed against five levels. For each requirement, level three and four are required to meet AgReFed trusted repository requirements (0 - 4):

0 – Not applicable
1 – The repository has not considered this yet
2 – The repository has a theoretical concept
3 – The repository is in the implementation phase (AgReFed compliant)
4 – The guideline has been fully implemented in the repository (AgReFed compliant)

Note: These Federation Trusted Repository Policy settings are a simplification of the CoreTrustSeal process. It only identifies those requirements the data repository must meet, selected from a subset of the characteristics of trustworthy repositories identified by Core Trustworthy Data Repositories Requirements v01.00. As such, these Trusted Repository Policy settings do not conform to or meet CoreTrustSeal specifications.

These are the settings of AgReFed trusted assessment v1.0.

Any changes to policy must be reflected in the AgReFed trusted assessment v1.0.xlsx assessment template (this resource should be updated with URI)

R2 Licenses: The repository maintains all applicable licenses covering data access and use and monitors compliance
R3 Continuity of Access: The repository has a continuity plan to ensure ongoing access to and preservation of its holdings
R4 Confidentiality/Ethics: The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms
R11 Data Quality: The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations
R13 Data Discovery and Identification: The repository enables users to discover the data and refer to them in a persistent way through proper citation
R14 Data Reuse: The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data
R15 Technical infrastructure: The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its users.
R16 Security: The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users

Appendix 4: FAIR Data Assessment

1. The AgReFed FAIR Data Assessment will be used for assessing the FAIR data alignment
2. Current thresholds required for qualification as FAIR data are shown in the table below
3. The Federation Data Steward will review and approve (or reject) data providers' FAIR Data Self-assessments
4. Any disputes in relation to the validation of assessment will be escalated to the Federation Technical Committee for review and decision
5. The FAIR assessment process and settings (including qualifying threshold levels) may be modified by the Federation Technical Committee

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

These are the settings of AgReFed FAIR Data Assessment v1.0.

Any changes to policy must be reflected in the AgReFed FAIR Data Assessment v1.0.xlsx assessment template (this resource should be updated with URI).

Principle (for AgReFed)	**Increasingly FAIR -->**				
FINDABLE					
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)
ACCESSIBLE					
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 de identified 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	No (or not applicable, if no metadata record exists)	Unsure	Yes		

available.					
INTEROPERABLE					
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		
REUSABLE					
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		

* Question 3 specifies minimum metadata requirements for collections and services (see Appendix 2)

Appendix 5: Schematic Interoperability

Data should be delivered in a structure that is specified using appropriate domain specific or generic information models.

Currently accepted domain specific information models are:

GroundWaterML2 (https://www.opengeospatial.org/standards/gwml2): This OGC standard specifies a conceptual model, a logical model and a GML (XML) implementation. It represents key hydrogeological entities such as aquifers and water wells, as well as related measurements and groundwater flows.

ANZSoilML (http://anzsoil.org/anzsoilml/): This Australian/New Zealand standard is a GML application schema that specifies a set of feature-types and supporting structures for information used in the Australian and New Zealand soil sciences.

Generic information models are:

Observational data: The preferred implementation for observational data, such as crop yield or soil moisture is the ISO/OGC Observations & Measurements (O&M) information model (http://www.opengis.net/doc/om/2.0).

The Spatial Data on the Web Working Group (SDWWG), a joint W3C-OGC project, has modified and upgraded O&M. The Semantic Sensor Network (SSN) ontology describes sensors and their observations, the involved procedures, the studied features of interest, the samples used to do so, and the observed properties, as well as actuators. SSN follows a horizontal and vertical modularization architecture by including a lightweight but self-contained core ontology called SOSA (Sensor, Observation, Sample, and Actuator) for its elementary classes and properties.
The Federation data providers should preferentially use SSN (http://www.w3.org/ns/ssn/), SOSA (http://www.w3.org/ns/sosa/) and agreed extensions to these, such as ssn-ext (https://www.w3.org/TR/vocab-ssn-ext/), as they become available.

SensorThingsAPI https://www.opengeospatial.org/standards/sensorthings

Physical sample data:
The ISO/OGC Observations & Measurements (O&M) information model (http://www.opengis.net/doc/om/2.0) specification or SOSA can be used for physical sample and specimen data.

Depending on the information model, the data may be transferred in various formats, such XML, RDF, JSON. The accepted formats are specified in the Computational Viewpoint.

Appendix 6.1. Controlled Vocabularies

Authoritative generalist or domain specific vocabularies currently used in AgReFed include:

Domain	Description	collection or example from collection (URI)	Status at 09/10/2019
Soil	Australian soil and land survey field handbook codes	http://registry.it.csiro.au/def/soil/au/asls	Stable collection, concepts experimental
	Soil chemical methods - Australasia	http://registry.it.csiro.au/def/soil/au/scma	Stable collection, concepts experimental
	Procedures	http://registry.it.csiro.au/def/soil/au/edu/fed-uni/procedure	Stable
	observed properties	http://environment.data.gov.au/def/property/sulfur_concentration	Stable
	observed properties	http://registry.it.csiro.au/sandbox/soil/soil-property/exchangeable_potassium	Experimental
	substance or species	http://registry.it.csiro.au/sandbox/soil/soil-object/soil_particles	Experimental
	substance or species	http://environment.data.gov.au/def/object/magnesium	Stable
general	units of measure	http://registry.it.csiro.au/def/environment/unit/MilliEquivalentPerHectoGram	Stable
general		http://www.opengis.net/def/nil/OGC/0/inapplicable	Item currently valid

Non-authoritative vocabularies currently used in AgReFed include:

Domain	Description	URI	Status
Soil	A collection of some procedures used to describe the measurement of soil observable properties. Local vocabs in use by Federation University	http://registry.it.csiro.au/def/soil/au/edu/fed-uni/procedure	Stable
Agriculture	AGROVOC - FAO/AIMS	http://aims.fao.org/vest-registry/vocabularies/agrovoc	Stable

Appendix 6.2. Best practice for finding and creating vocabularies for use in AgReFed

Data on the web Best Practice (WC3) should be used[16], that is to "Reuse vocabularies, preferably standardized ones: Use terms from shared vocabularies, preferably standardized ones, to encode data and metadata."

Best practice on finding vocabulary should be applied.

- Vocabularies MUST have a persistent identifier (URI)
- Vocabularies MUST be accessible for a long period
- Vocabularies MUST be documented. This includes the liberal use of labels and comments, as well as appropriate language tags. The publisher must provide human-readable pages that describe the vocabulary, along with its constituent classes and properties.
- Vocabularies SHOULD be published by a trusted group or organization
- Vocabularies SHOULD be self-descriptive
- Vocabularies SHOULD be described in more than one language
- Vocabularies SHOULD be used by other datasets
- Vocabularies SHOULD provide a versioning policy

The above is and adaptation of WC3[17] best practice on finding a vocabulary. The adaption is that vocabularies MUST have persistent URIs and MUST be accessible for a long period.

Where data providers decide they must develop their own vocabularies, it is recommended that best practice on vocabulary creation be applied (See WC3 "vocabulary creation"[15])

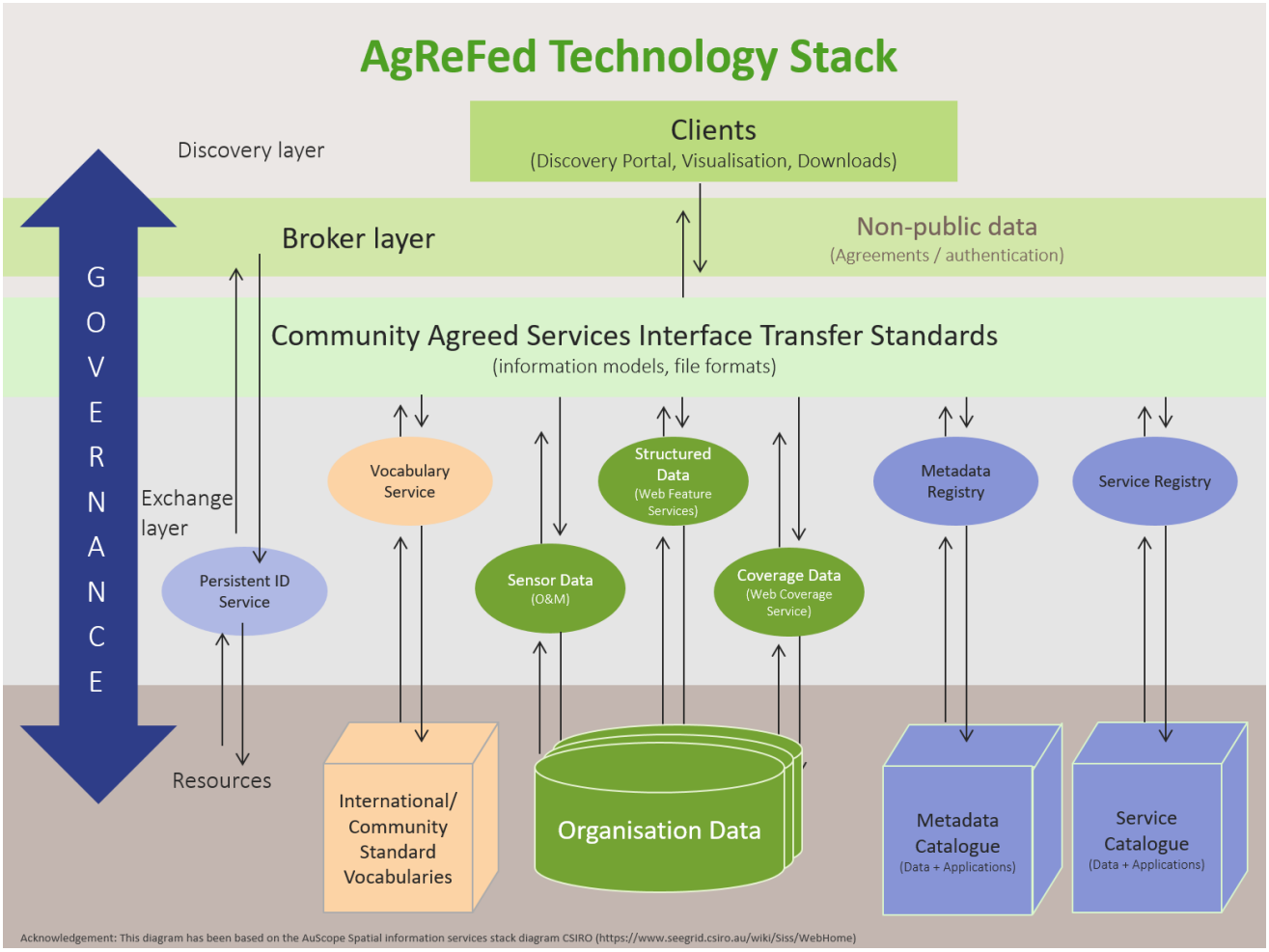
[16] https://www.w3.org/TR/dwbp/#dataVocabularies
[17] https://www.w3.org/TR/ld-bp/#VOCABULARIES

Appendix 7 Computational policy

The *computational policy* describes the functionality provided by the system.
This appendix covers the discovery and brokering layers such as:

- user access conditions/controls
- discovery functionality
- portal behavior,
- visualisation behavior
- download formats
- API bindings/functionality

Appendix 9 lists the accepted computing infrastructure.



High level functionality (computational interface) stack

7.1. Portal functionality

Primary discovery mechanism provided via RDA catalogue. The AgReFed portal has its own RDA service record which is used to create formal linkages to related data, services, organisations and people.

https://researchdata.ands.org.au/agrefed-portal/

The assumption is that RDA catalogue is (and will continue to be) well indexed by search engines and other catalogues.

Data providers create and maintain metadata within their institutional repositories which are harvested by RDA.

Discovery portal

The AgReFed discovery portal provides a data discovery and visualisation interface for AgReFed data. Data discovery should make use of ARDC services (RDA API). Centralised functionality should be kept to a minimum and be open/redistributable as appropriate. The discovery portal should contain at a minimum:

> **Spatial mapping portal** based on OpenLayers (https://openlayers.org/) providing visualisation and interrogation of data.
>
> **Data exploration** and demonstration user interfaces for exemplar datasets.
>
> **Export/Download** to common formats where appropriate (licencing)

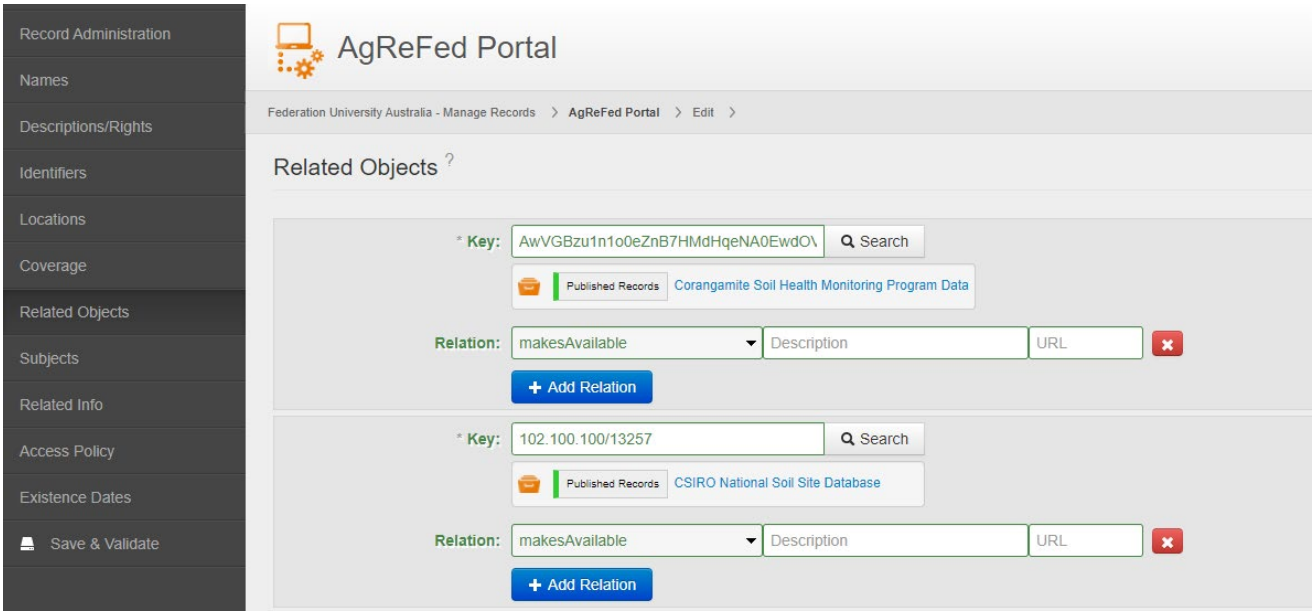
Developed codebase should be opensource and available via GitHub or similar.

Portal Behavior

Data discovery is exposed via the 'explore the data' section of the AgReFed website https://www.agrefed.org.au/ExploretheData

This page features the exemplar datasets and various ways to explore them via metadata records, download options, spatial and other data visualisation interfaces.

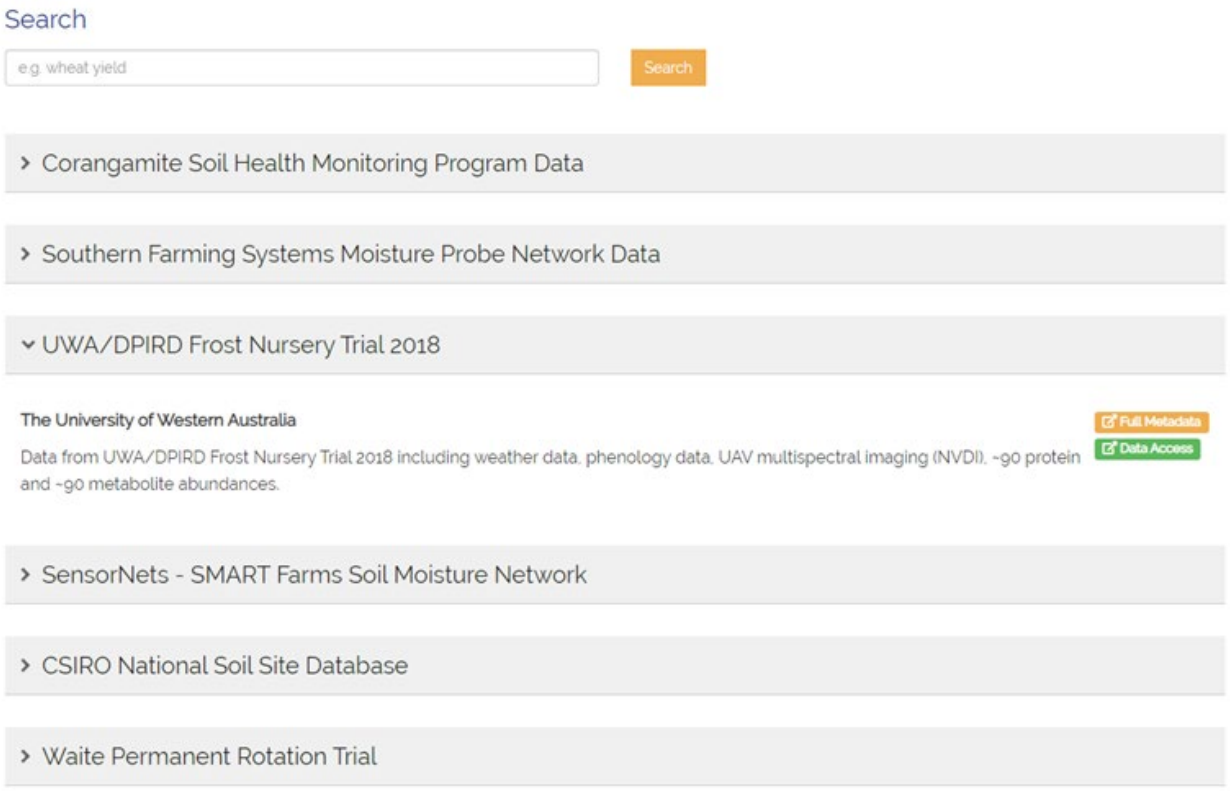
A search interface is provided which performs a query to the RDA registry via its getMetadata API. Results are limited to records which are 'related' to the AgReFed service record within RDA.



RDA Registry interface showing AgReFed service record and 'related' objects

https://documentation.ands.org.au/display/DOC/getMetadata+API

Results display provider organisation, dataset title, description and links to full metadata (RDA registry) and data access (DOI).

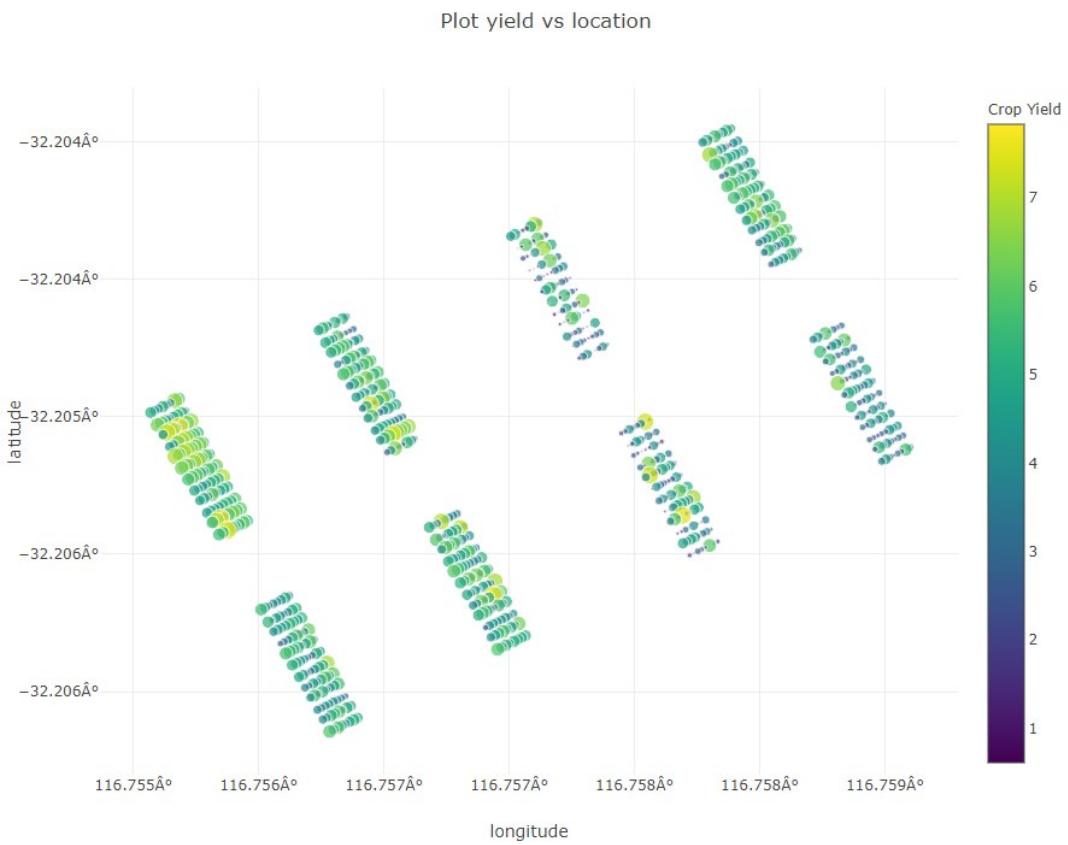


Explore the data section of AgReFed website showing dataset search results.

Data visualisation and exploration examples

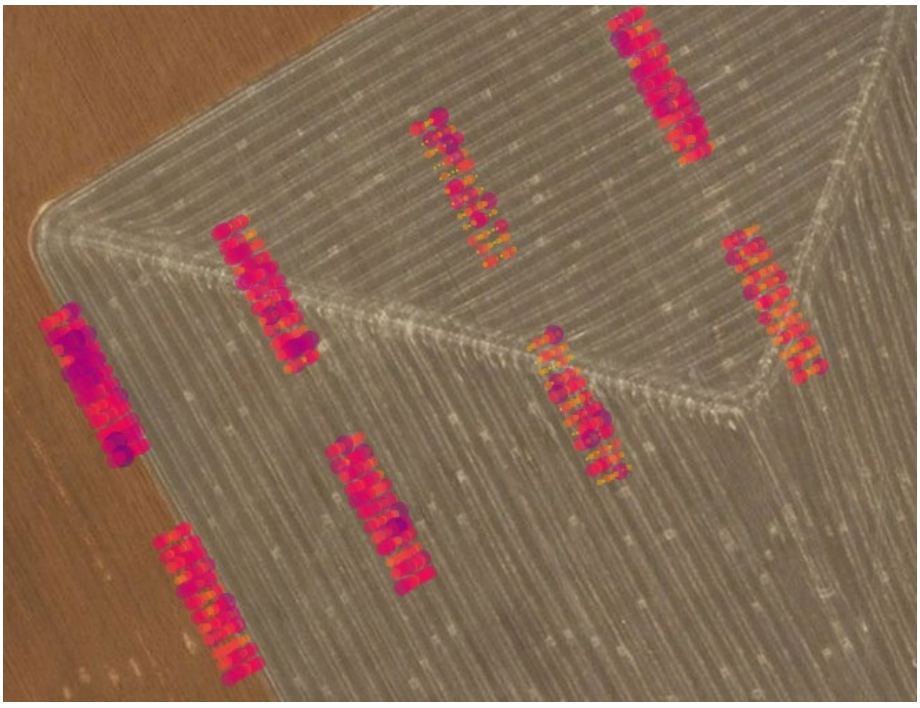
The following figures show data visualisations which have been deployed on the AgReFed portal as use-case demonstrations. The examples shown are all transferrable, making use of open-source software and retrieving data via web services.

The following two examples demonstrate two different visualisation options for the UWA / DPIRD Frost Trial dataset with data being accessed via the AgReFed@dale API hosted at UWA.



UWA visualisation of plot yield vs location on a Plotly.js chart

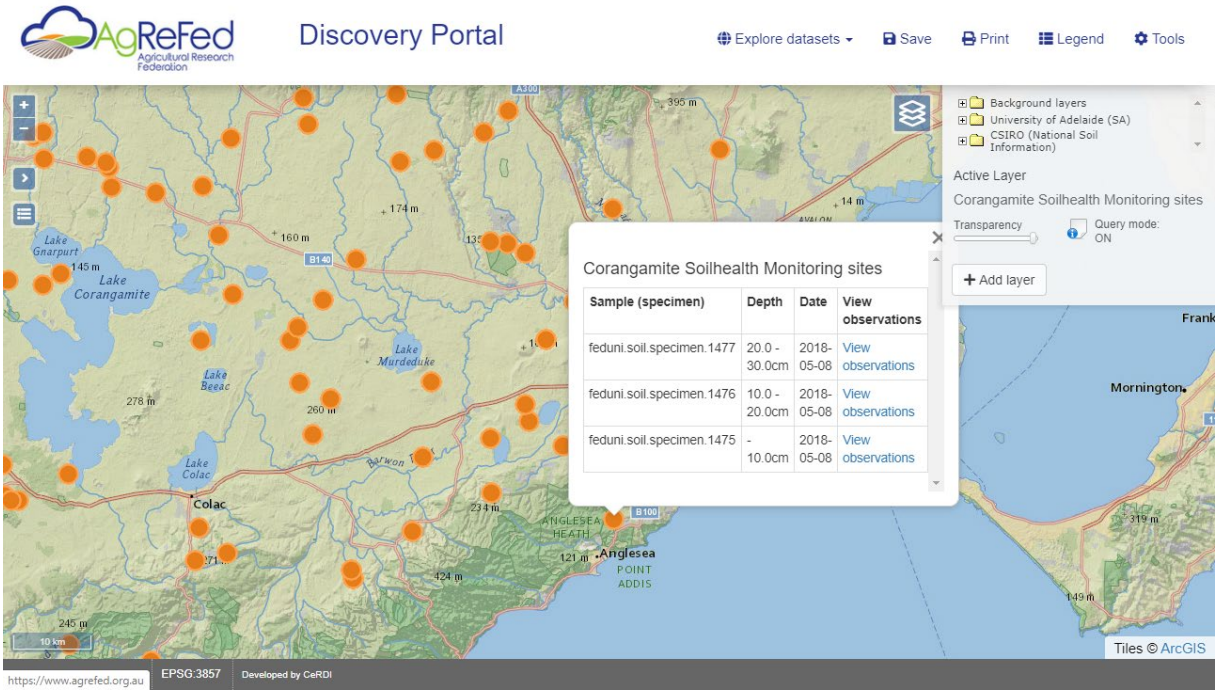
https://plot.ly/python/scatter-plots-on-maps/



UWA visualisation of plot yield vs location on a Leaflet.js map

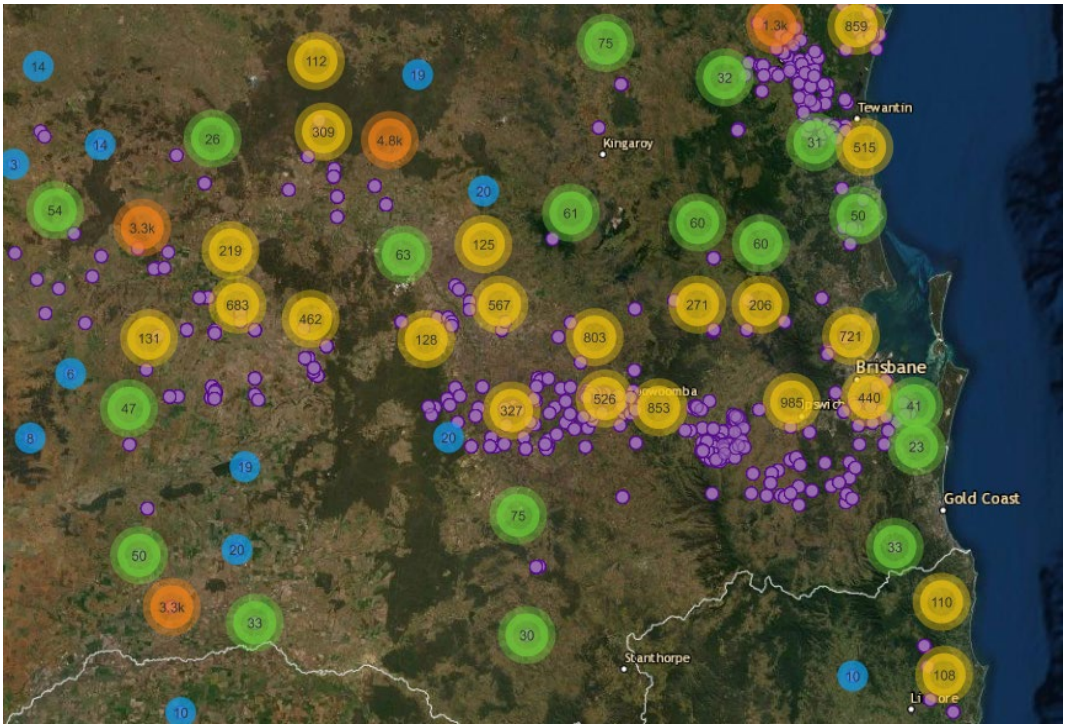
The next two examples demonstrate the spatial interface available at https://www.agrefed.org.au/agrefed_map.php The mapping interface is based on the most recent version of OpenLayers https://openlayers.org/ . OpenLayers is able to display (and query) a wide range of spatial data formats in both vector, and tiled (raster) formats.

In the first example the data layer of points is directly accessing a GeoJSON web service of soil site locations. When the points are clicked, the interface submits a query for soil test data to the ANZSoilML services deployed by Federation University Australia.



Spatial mapping portal showing Corangamite CMA soil health monitoring sites being queried

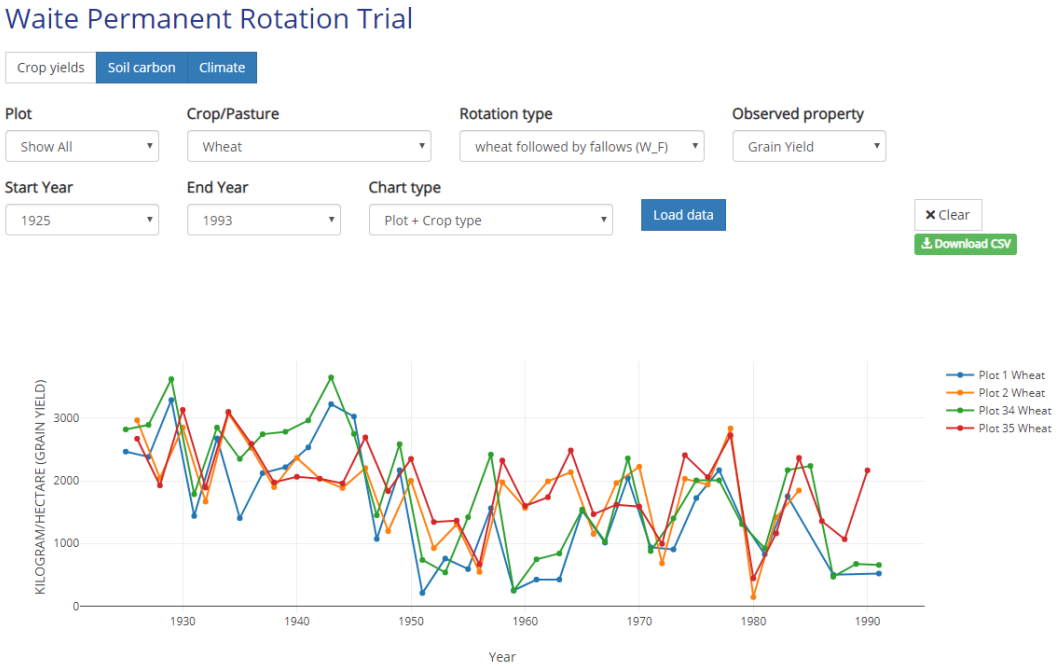
Similarly this second spatial example shows a ‘point cluster’ visualisation of the CSIRO national soil sites database. In this case the processing of the visualisation is provided by a Geoserver ‘rendering transformation’ layer which is cached at FedUni. When the user clicks a point, the interface switches to access the live ANZSoilML services deployed by CSIRO.



Mapping portal showing CSIRO national soil sites (clustered)

The final example shows a custom interface developed by CeRDI to demonstrate the API deployed for the UWA Waite trial data.

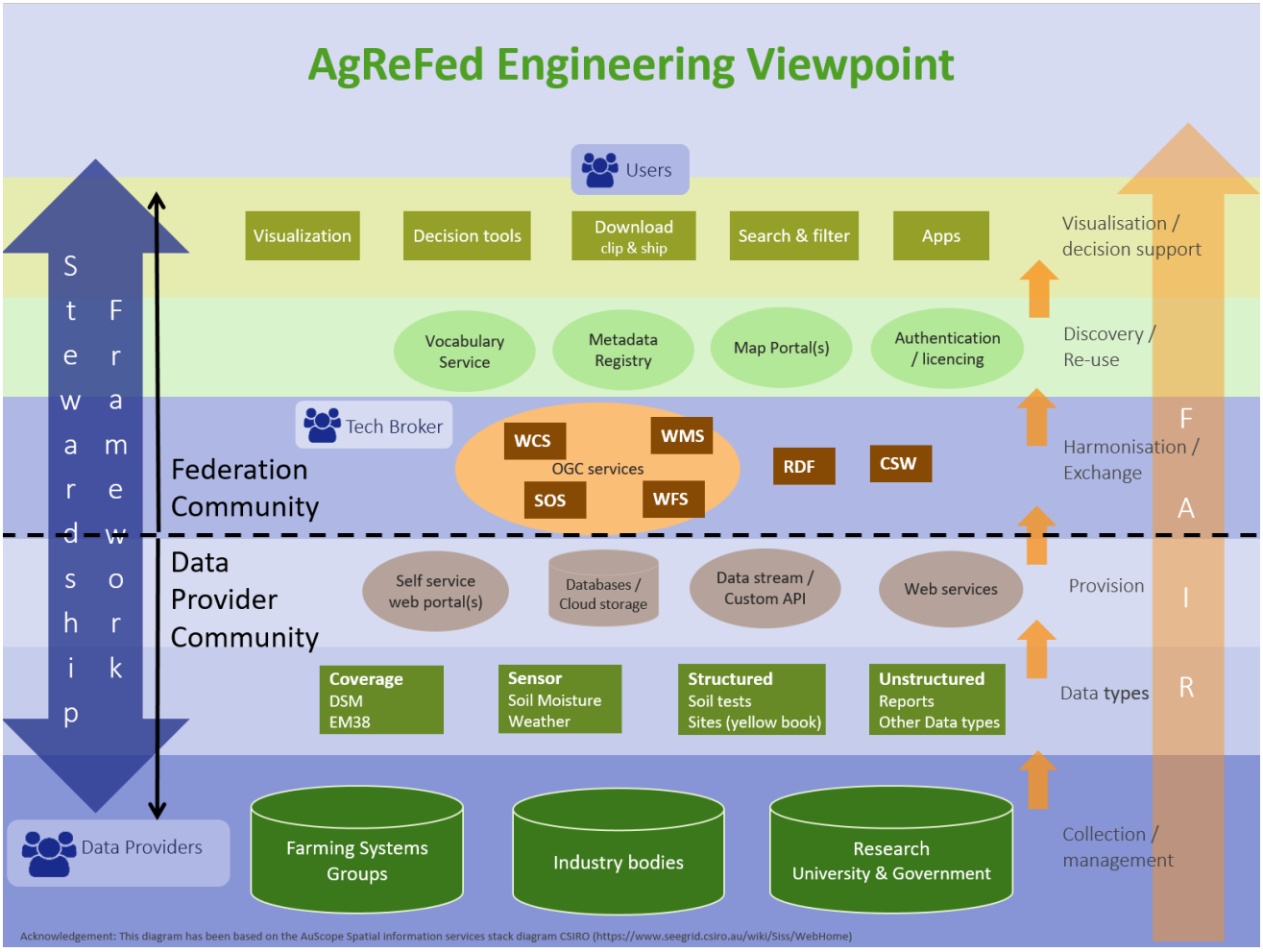
https://agrefed.org.au/waite_demo.php



University of Adelaide Waite trial - dynamic charting and download

Appendix 8: Engineering policy- interface transfer standards

The *engineering policy* describes the distribution of processing performed by the system to manage the information and provide the functionality.



Engineering viewpoint diagram

WFS (complex features): If using OGC web feature services (WFS) then the appropriate GML XML implementation should be used. Examples:

- O&M then use the OGC O&M v2.0 XML implementation (http://schemas.opengis.net/om/2.0/);
- ANZSoilML then use the ANZSoilML v2.0 XML implementation (http://anzsoil.org/def/schema/)
- TimeSeriesML encoding of O&M then use the TimeSeriesML v1.2 XML implementation (http://schemas.opengis.net/tsml/1.2/)

WFS (simple features): Acceptable, accompanying metadata and well-documented data schema required.

WMS: Use the current standard (https://www.opengeospatial.org/standards/wms) where possible. Deprecated versions are also acceptable.

WCS: Use the current standard https://www.opengeospatial.org/standards/wcs where possible. Deprecated versions are also acceptable.

RDF: For services delivering observation data in RDF, they should use one of:

- O&M RDF encoding (http://def.seegrid.csiro.au/static/isotc211/iso19156/2011/);
- SSN and SOSA (https://www.w3.org/TR/vocab-ssn/)

Specific APIs:

The use of standardised REST APIs is accepted:

- OASIS ODATA - https://www.odata.org/documentation/
- OpenAPI - https://swagger.io/docs/specification/about/

Alternative file formats, such as JSON, JSON-LD, GeoJSON may be appropriate for some data exchange standards.

Vocabulary Services

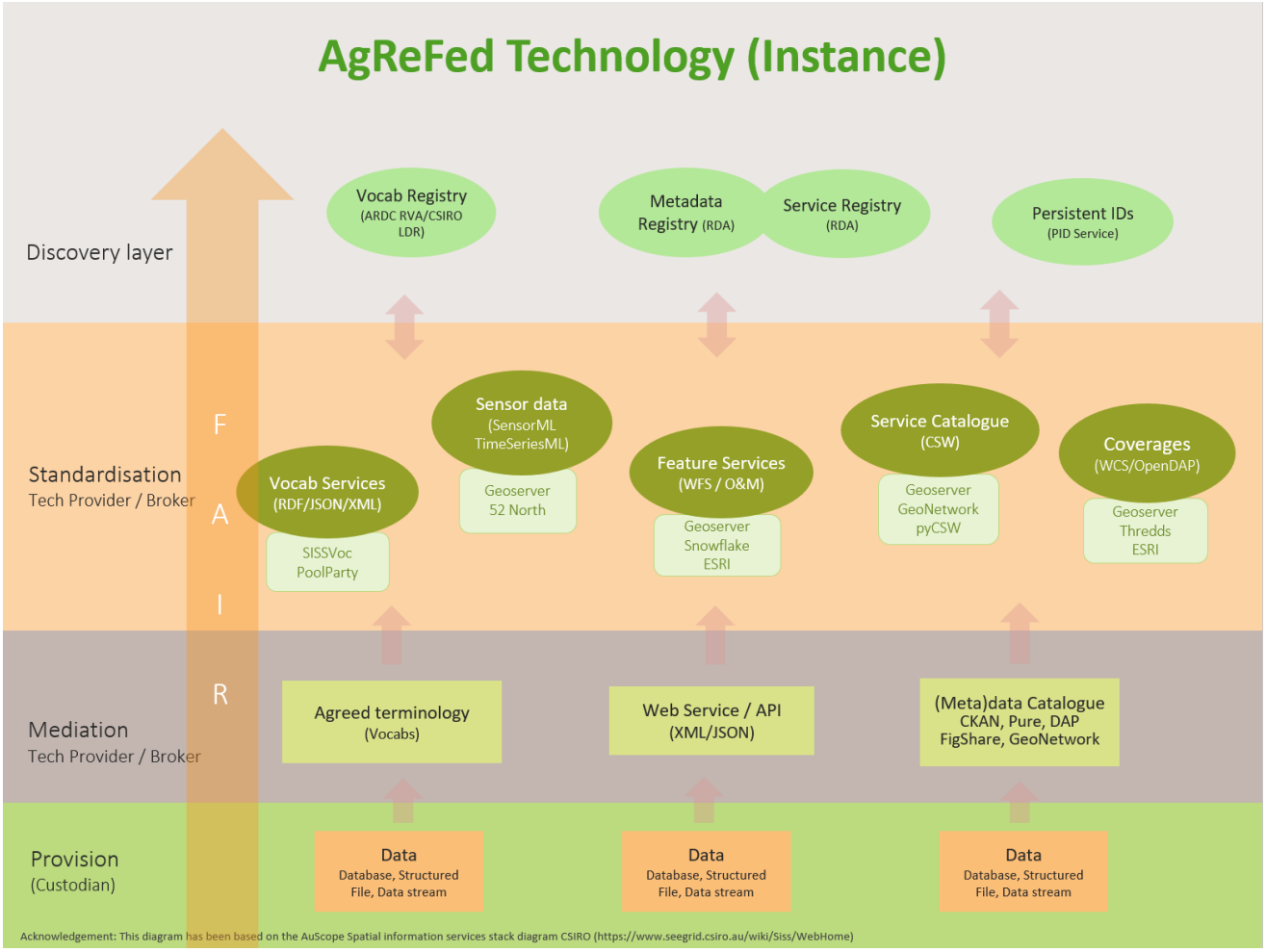
- **SISSVoc** - developed internally, originally hosted by SEEGrid, subsequently adopted as one element of Research Vocabularies Australia (RVA) hosted at ANDS
- Linked Data Registry (**LDR**) - originally developed for WMO and UK-DEFRA, subsequently used by Bioregional Assessments through the http://environment.data.gov.au/def service, and at http://registry.it.csiro.au/def as part of a semantic registry

Other semantic-web/linked-data compatible vocabulary services include:

- SKOSMOS
- PoolParty

Appendix 9: Technology Policy

The *technology policy* describes the current technologies used to provide the processing, functionality and presentation of information. Other technologies may be used if they provide the functionality required by AgReFed.



Technology stack (instance) with typical software, protocols and standards

Technologies currently used by AgReFed include:

Portal technology:
- OpenLayers

Metadata catalog technology:
- CSIRO Data Access Portal (DAP)
- Research Data Australia (RDA)
- GeoNetwork
- CKAN
- DKAN
- Figshare/Pure

Vocabulary technology:
- CSIRO Linked Data Registry
- PoolParty
- SKOSMOS

URI technology(s):
- Digital Object Identifier (DOI) service

- http-URI
- Handle service

WFS technology:

- Geoserver
- Snowflake
- ESRI ArcGIS

WMS technology:

- Mapserver
- Geoserver
- ESRI ArcGIS

WCS technology:

- Thredds
- Geoserver
- ArgGIS

RDF access technology:

- Apache Jena

SensorThingsAPI:

- Frost Server

Appendix 10. The services used by the Federation

Active service end-points currently available via the Federation.

Federation University

Geoserver WFS (getCapabilities)

https://services.cerdi.edu.au/anzsoilml/wfs?request=getCapabilities

Frost Server (SensorThingsAPI)

http://services.cerdi.edu.au/sfs/v1.0

CSIRO

Geoserver WFS (getCapabilities)

http://www.asris.csiro.au/natsoil/anzsoilml2/wfs?SERVICE=WFS&VERSION=1.1.0&REQUEST=GetCapabilities

ArcGIS WCS (getCapabilities)

http://www.asris.csiro.au/arcgis/services/TERN/CLY_ACLEP_AU_NAT_C/MapServer/WCSServer?request=GetCapabilities&service=WCS

University of Western Australia (UWA)

OpenAPI (Swagger)

https://webapps.plantenergy.uwa.edu.au/agrefed_dale

https://webapps.plantenergy.uwa.edu.au/agrefed_dale/swagger.yaml

University of Adelaide (UofA)

PostgREST ODATA API

https://prt.waite.adelaide.edu.au/

UNE

Currently trialling ODATA for institution wide implementation

https://www.odata.org/

Appendix 11. Use cases and their key data interoperability requirements

The use cases of data prosumers and how they could be addressed through the AgReFed user experience

Data product	https://doi.org/10.4225/08/55E5165EC0D29
Target user groups	Researchers
The user challenge	Currently, Waite Permanent Rotation Trial datasets (1925 – 1993) are stored within tables on separate worksheets within an excel file. This excel file is accessible as a single download through the CSIRO Data Access Portal. The researchers at the University of Adelaide would like to be able to quantify how changes in pasture composition and crop rotation histories affect the subsequent yield of wheat in a long-term rotation field trial, as well as investigate relationships with soil physicochemical properties from the soil data set. Additionally, they would like to be able to combine the Waite Trial datasets with other datasets, such as climate data available from the Bureau of Meteorology. The analysis would provide researchers an opportunity to study differences in the interaction of genotype, environment and management over the past decades with the possibility of using this historic data in predictive crop modelling. To be able to do this, they need to be able to access each individual Waite Trial dataset, search the parameters of interest within these datasets (including across a specified period of time), combine this data with other available datasets that match the parameters of interest (such as climate data across a specified period of time), and download this data in a form ready for statistical analysis.
How the challenge could be solved through AgReFed	The Waite Permanent Rotation Trial legacy data can be data delivered from UoA's MySQL database in a consistent and standard format so that user defined parameters can be searched, combined and downloaded in a structured format. Additionally, other agricultural and environmental datasets delivered in standard format through AgReFed can also be searched by these parameters and viewed, selected, combined and downloaded with the Waite Permanent Rotation Trial data ready for statistical analysis.
The user experience	The user searches by a parameter of interest for their research question, for example by a particular crop (such as 'wheat') and the wheat measurement of interest (such as 'yield', 'dry matter'). They then select, from a drop-down menu, covariates of interest that could affect wheat yield such as time periods, rotation types and soil variables. The user then searches and views data from other datasets accessible through AgReFed that meets these search parameters, such as climate data from the Bureau of Meteorology at the same locality and time. The researcher can then view this combined data as a table or graph, choose to refine the search (such as filtering by 'soil carbon' 'rotation history' 'rainfall' and 'temperature') and download the combined dataset in a standard format amenable to statistical analysis (such as .csv) to investigate the interactions of climate and rotation history with soil carbon sequestration and wheat yield between (for example) 1950 – 1993.
Key data interoperability requirements	• Allow the datasets in each of the data sheets to be discovered, and data providing services to be discovered (data and service metadata); • For this user case, as well as spatial (plot) and temporal properties (date-month-year), other parameters include what is being measured (e.g. wheat), the observed property being measured (e.g. 'yield'), how it is being measured ('dry matter biomass cuts'), the units of measurement (kg/ha-1) and the observation (here the kg/ha-1 of dry matter yield of wheat). • Allow the subset of the data that meets the query to be downloaded (note this is not a complete file download, so requires some kind of web service); • Allow the downloaded data to be reformatted to meet an externally defined structure (schema); • Allow mapping of terms in the data to external vocabularies and semantics (e.g. replacing local terms with RVA or external vocabularies such as "m" or "metre" in the data with "http://pid.geoscience.gov.au/def/voc/ga/uom/m") • Allow locality to be interoperable between datasets when the locality is specified in the metadata set (for example lat long with coordinate reference system)
Any other requirement – functional?	• Ability to graph parameters of interest • Single download of combined dataset (.csv)

Data Product	https://rune.une.edu.au/web/handle/1959.11/23173
Target user groups	Researchers, the general public and undergraduate students
The user challenge	UNE collect an extensive range of IoT, sensor and other agricultural data across a series of research and commercial farming enterprises. Currently, data from their SMART Farms Soil Moisture Network (real-time soil moisture data) is stored in an ORACLE database and whilst the dataset can be found through searchable repositories online, access to the dataset itself mediated. That is, UNE staff need to provide data to the user upon individual request. The researchers at the University of New England would like their students, researchers and the general public to have a way of visualising and downloading their SMART Farms Soil Moisture Network data in both current/real-time and across the specified timeframes or from a sensor selected by its location. This would make the data more useful for UNE students as an educational resource: students could remotely access, visualise, download the data required to investigate short term trends and use for scenario-based learning activities. Additionally, researchers at UNE would like to be able to enhance the learning experience of students by providing them the ability to combine the SMART Farms Soil Moisture Network data with additional datasets (from UNE or from other providers) as they become available.
How the challenge could be solved through AgReFed	The UNE SMART Farms Soil Moisture Network can be delivered from their ORACLE database in a consistent and standard format so that the user defined parameters can be searched, combined and downloaded in a user-defined structured format. In time, new datasets will also be able to be added to the ORACLE database and similarly delivered in the consistent and standard format. Additionally, other datasets delivered in standard format through AgReFed can also be searched by user-defined parameters and viewed, selected, combined and downloaded with the UNE SMART Farms data ready for statistical analysis.
The user experience	The user (researchers, the general public or students) can search for real-time UNE SMART Farms Soil Moisture Network data, or across a specified timeframe. They select the weather station/s of interest by selecting it on a map and the associated parameter/s of interest (soil moisture, soil temperature, soil electrical conductivity and/or air temperature), and then visualise the data on a dashboard and download it by .csv. Over time, researchers at UNE make other datasets in their database discoverable and accessible through both AgReFed and other services as desired – this data is then combined with Soil Moisture Network data by parameters of interest, and downloaded as a single dataset ready for analysis. Real-time data and trends across time are investigated, and remote on-demand scenario-based learning is enabled providing a flexible exploratory tool with little burden on data providers.
Key data interoperability requirements	• Allow the data held within ORACLE database to be discovered, and data providing services to be discovered (data and service metadata); • Allow the data to be queried on user-defined parameters including temporal properties (date-month-year), the soil moisture probe ID and georeferenced, the object of interest (e.g. water), the observed property (e.g. soil moisture), the procedure and the units of measure. The observed properties of interest include soil moisture, soil temperature, soil electrical conductivity and air temperature. • Allow the subset of the data that meets the query to be downloaded (note this is not a complete file download, so requires some kind of web service), particularly of interest in this user case is the ability to download a subset of data from a specified time period (real time or range) and probe ID (georeferenced) ; • Allow the downloaded data to be reformatted to meet an externally defined structure (schema); • Allow mapping of terms in the data to external vocabularies and semantics (e.g. replacing local terms with RVA or external vocabularies such as "m" or "metre" in the data with "http://pid.geoscience.gov.au/def/voc/ga/uom/m") • Allow locality to be interoperable between datasets when the locality is specified in the metadata set (for example lat long with coordinate reference system)
Any other requirement – functional?	• Ability to graph parameters of interest • Single download of combined dataset (.csv)

Data product	10.26182/5cedf001186f3
Target user groups	Researchers, the general public and undergraduate students
The user challenge	Researchers at UWA and the DPRID collect a wide range of datasets to investigate the desirable plant traits and growth conditions for crop yield. Currently, the raw datasets are all processed and stored in separate locations usually on the users specific computer for a range of data sets including hyper- and multispectral data, proteomics and metabolomic data), weather information and plant and grain properties. These data set are often only retained upto publication or report writing and then the data effectively becomes "lost" to other researchers. Researchers would like to be able to search all of the processed data from these datasets simultaneously by factors such as sowing time, variety type and replicates, view the data graphically by covariates of interest, and download the parameter of interest in a single dataset ready for analysis. Into the future, researchers need to be able to add further data and measurements to these datasets and have available well documented workflow of each dataset.
How the challenge could be solved through AgReFed	Processed data from spectral datasets, plant Proteomics and Metabolomics datasets, plant and grain growth measurement datasets, and on-site weather datasets can be delivered from their MySQL database in a consistent and standard format so that the user defined parameters can be searched, combined and downloaded in a user-defined structured format. As new data and parameters are collected, the UWA/DPRID will be able to add these in a consistent and standard format. Documented workflows will be available for each dataset within the metadata for each dataset. (Additionally, other datasets delivered in standard format through AgReFed can also be searched by user-defined parameters and viewed, selected, combined and downloaded with the UWA/DPRID trial data ready for statistical analysis)
The user experience	Researchers will be able to upload their processed trial datasets into their local MySQL database. This data will then be delivered and discoverable in a standard format through AgReFEd - The researchers will be able to compare their field spectral data with other spectral data to find the closest match (ground based hyperspectral and UAV based multispectral). They will also be able to select, combine and download by other covariates of interest including include plant growth/grain yield, ProteMetabolite data and on-site weather data, as well as factors such as sowing time, variety type and replicates. The selected data will viewable graphically, selected, combined and downloadable in a form ready for analysis to help identify plant traits (including Proteomics and Metabolomics markers) and their easily measurable proxies (such as spectral data). New data added to their datasets will also be discoverable through AgReFed, and for reproducibility the workflows of each of the accessed datasets will be easily accessible via the metadata. This will assist the research team to help identify new traits for plant breeding programs and engineering, for example frost-hardiness through frost trials.
Key data interoperability requirements	● Allow the data held within a MySQL database to be discovered, and data providing services to be discovered (data and service metadata); ● Allow the data to be queried on user-defined parameters including temporal properties (date-month-year), the object of interest, the observed property, the procedure and the units of measure (eg "variety type", "plant growth", "zaddock's growth stage" and unit of measure = none (1 – 8)). ● Allow the subset of the data that meets the query to be downloaded (note this is not a complete file download, so requires some kind of web service), specifically of interest in this user case is the ability to download a subset of data that matches a UAV spectra ● Allow the downloaded data to be reformatted to meet an externally defined structure (schema); ● Allow mapping of terms in the data to external vocabularies and semantics (e.g. replacing local terms with RVA or external vocabularies such as "m" or "metre" in the data with "http://pid.geoscience.gov.au/def/voc/ga/uom/m")
Any other requirement – functional?	● Ability to graph parameters of interest ● Single download of combined dataset (.csv) ● Ability to embargo some data until publication or submission of thesis