Guidelines for the development of a Data Stewardship and Governance Framework for the Agricultural Research Federation (AgReFed)

Paul Box[1], Kerry Levett[2], Bruce Simons[3], Megan Wong[3]

[1] CSIRO
[2] Australian Research Data Commons (ARDC)
[3] Federation University Australia

Version 1.2 2019-05-30

AgReFed Agricultural Research Federation

Citation

Box, Paul; Levett, Kerry; Simons, Bruce; Wong, Megan (2019) Guidelines for the development of a Data Stewardship and Governance Framework for the Agricultural Research Federation (AgReFed) Version 1.1. CSIRO.

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).

Partners in the project were the University of Adelaide, the University of Western Australia, the Western Australian Department of Primary Industries and Regional Development (DPIRD), the University of New England, Federation University Australia and the Commonwealth Scientific and Industrial Research Organisation. The authors would like to acknowledge all individuals from these partner organisation that provided valuable input into this document, including technical input from Andrew MacLeod, significant feedback on drafts by Helen Thompson and Peter Wilson and early version input by Nigel Prior and Kay Steel.

Table of contents

AgReFed
Agricultural Research Federation

Executive summary

This document is a product of the Agricultural Research Data Cloud (AgRDC) project led by Federation University Australia. The intent of the document is to enable the transition from an AgRDC project to a sustained federated community infrastructure - The Agricultural Research Federation(AgReFed) - with the common goal of improving the sharing and reuse of agricultural data including datasets, metadata and data related products.

This document provides high level design and guidelines for the implementation of the AgReFed Data Stewardship and Governance Framework (the Stewardship Framework), a socio-technical system that (once enacted):

- Brings **independent organisations** together to make strategic and technical decisions about data sharing and
- Will **guide** and be **implemented** by agricultural data providers.

The document in intended primarily for use by the AgRDC stakeholder community. However, the Stewardship Framework and the guidelines for its implementation have been designed with reuse in mind. In this context it can be used by other information communities as a design pattern for networked governance and data stewardship.

This Stewardship Framework includes a **suggested**:

- Technical and Information architecture to improve the findability, discoverability, accessibility and reusability (FAIRness of) agricultural data through AgReFed and
- Social architecture, including:
  - Organisational form, being a **Federation** comprised of autonomous **Data Provider Communities;**
  - A governance model comprising a **Federation Council** and **Technical Committee**, whose membership is made up of representatives of the Data Provider Communities;
  - **Policies** that will be produced by the Federation Council and Technical Committee, to guide both Data Providers and the Federation;
  - **Roles** performed by community members to **govern** and **operate** AgReFed; and
  - **Processes** to ensure long term availability of FAIR agricultural data through AgReFed.

This document is presented in two parts:

1. An overview of the Guidelines for the development of a Data Stewardship and Governance Framework, including what you need to know and do as a provider community/potential provider community
2. Further details on the proposed Social, Technical and Information architecture to improve the findability, discoverability, accessibility and reusability (FAIRness of) agricultural data through AgReFed.

To finalise the AgReFed Data Stewardship and Governance Framework and **establish the Agricultural Research Federation**, the following needs to occur:

1. Approval of this AgReFed Data Stewardship and Governance Framework document by AgRDC project Steering Committee
2. Terms of Reference (ToRs) for the Federation Council and Federation Technical Committee developed by AgRDC project Stewardship team, and reviewed by Steering Committee
3. Establishment of the Federation Council and Federation Technical Committee by Data Provider Communities nominating a representative to sit on these bodies
4. At the inaugural meeting of the Council: formally establish the Council, Technical Committee, and key AgReFed roles; and sign off on ToRs and an initial set of Establishment and Membership policies (over time, the Funding and Financial policies and the Strategic and Business Policies would be developed and promulgated by the Council)
5. At a subsequent inaugural Technical Committee meeting, an initial core set of rowing policies would be promulgated (policies that ensure FAIR data and trusted repositories)

Overview of the Stewardship Framework: AgReFed as a socio-technical system

Principles that underpin the Stewardship Framework

The Agricultural Research Federation (AgReFed) aims to improve the sharing and reuse of agricultural data. The AgReFed Data Stewardship and Governance Framework (the Stewardship Framework) describes a socio-technical system that:

- Brings **independent organisations** together to make strategic and technical decisions about data sharing to meet identified user needs;
- Will **guide** agricultural data providers; and
- Be **implemented** by data providers using common technical infrastructure.

The Stewardship Framework is described using the Reference Model for Open Distributed Processing (RM-ODP), which provides five architectural 'Viewpoints' for specifying distributed information systems; and is mainly focussed on the first:

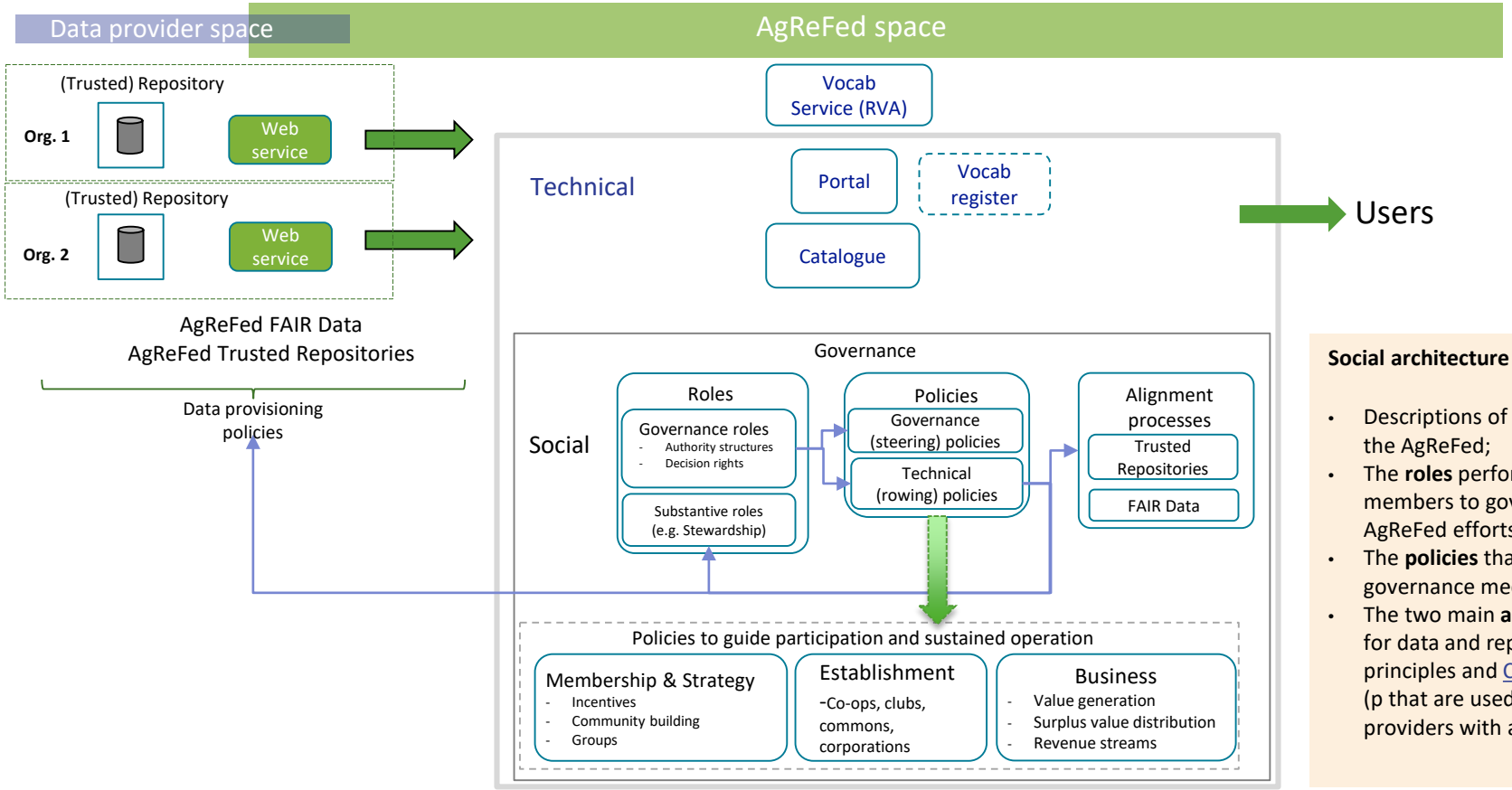
- **Enterprise Viewpoint - purpose, scope and policies of a system, and socio-economic and institutional environment within which it operates (the social architecture)**
- Information Viewpoint - scope and nature of the data content within AgReFed (the information architecture)
- Technical Viewpoints - the Computational, Engineering and Technology Viewpoint configuration of the system (the technical architecture)

The 'social architecture' is based around the following key concepts:

- Independent and autonomous **Data Provider Communities** and the collective **AgReFed Community** within which they participate;
- The **roles** performed by community members to **govern (steer)** and **contribute to (row)** AgReFed;
- The **policies** that are produced through governance mechanisms (decision making bodies and process) that guide collective and individual actions; and
- Two **processes to align** individual data providers' data and repositories with agreed collective direction. These alignment processes are adapted from FAIR Guiding Principles for scientific data management and stewardship (FAIR) and CoreTrustSeal Data Repositories Requirement (CoreTrustSeal).

AgReFed – A proposed socio-technical system

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

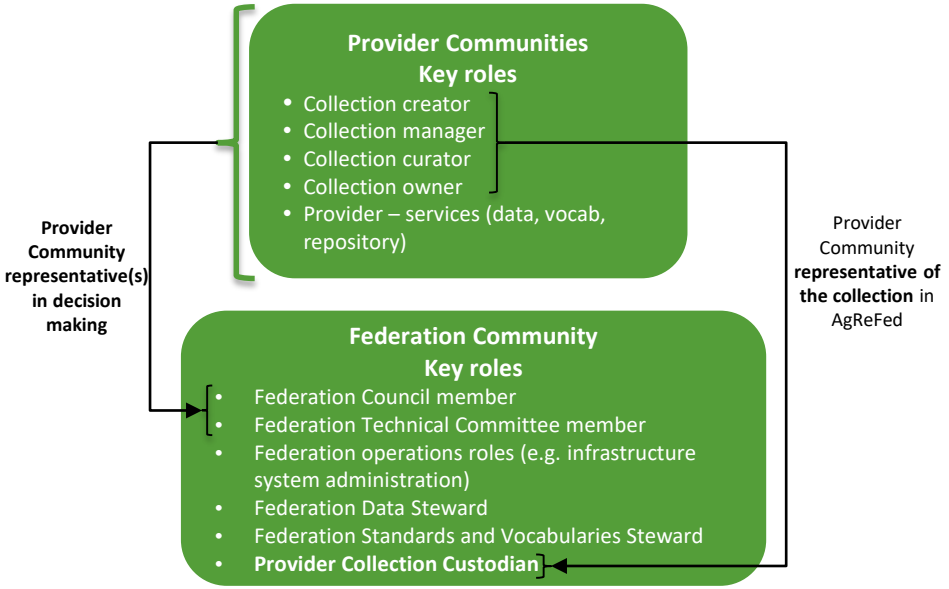


Social architecture comprises:

- Descriptions of the **communities** comprising the AgReFed;
- The **roles** performed by community members to govern and contribute to AgReFed efforts;
- The **policies** that are produced through governance mechanism; and
- The two main **alignment processes** (p20) for data and repositories adapted from FAIR principles and CoreTrustSeal requirements (p that are used to align individual data providers with agreed collective direction

Federated Community–Provider Community role relations

An outline of stakeholder communities and their proposed participatory roles and interactions in the AgReFed community



Stakeholder participation and relationships in AgReFed are defined using Communities and Roles

Data Provider Communities

- Each provider of a dataset or data collection is represented as its own autonomous community (comprising actors and systems involved in providing the repository, data services and curation, management and provisioning of data)
- Each community is **autonomous**, making decisions about how it organises itself.

The Federation Community and 'Contract'

- The community composed of all the various roles, that enable the AgReFed to operate
- The 'Contract' between Provider Communities and the Federation Community is expressed through AgReFed policies, roles and alignment process that determine the behaviours.

Participation in governance

- Participants are given **real voice** and **decision rights** in decisions that affect them and the community to which they are contributing
- Provider involvement in decision making is the key to building and sustaining AgReFed.

What you need to know and do: Role filling

As a data provider, to participate in AgReFed you will be expected to:

1. *Nominate one or more people to act as representative(s) of your Data Provider Community in the AgReFed Federation Council and Federation Technical Committee.*
2. *Identify a Provider Collection Custodian – a person to represent the interests of and act as point for contact for the data collection(s) that you contribute to AgReFed*

Federation Data Steward

- Reviews and validates AgReFed FAIR Data Self-assessments
- Reviews and validates AgReFed Trusted Repository Self-assessments

Federation Standards and Vocabularies Steward

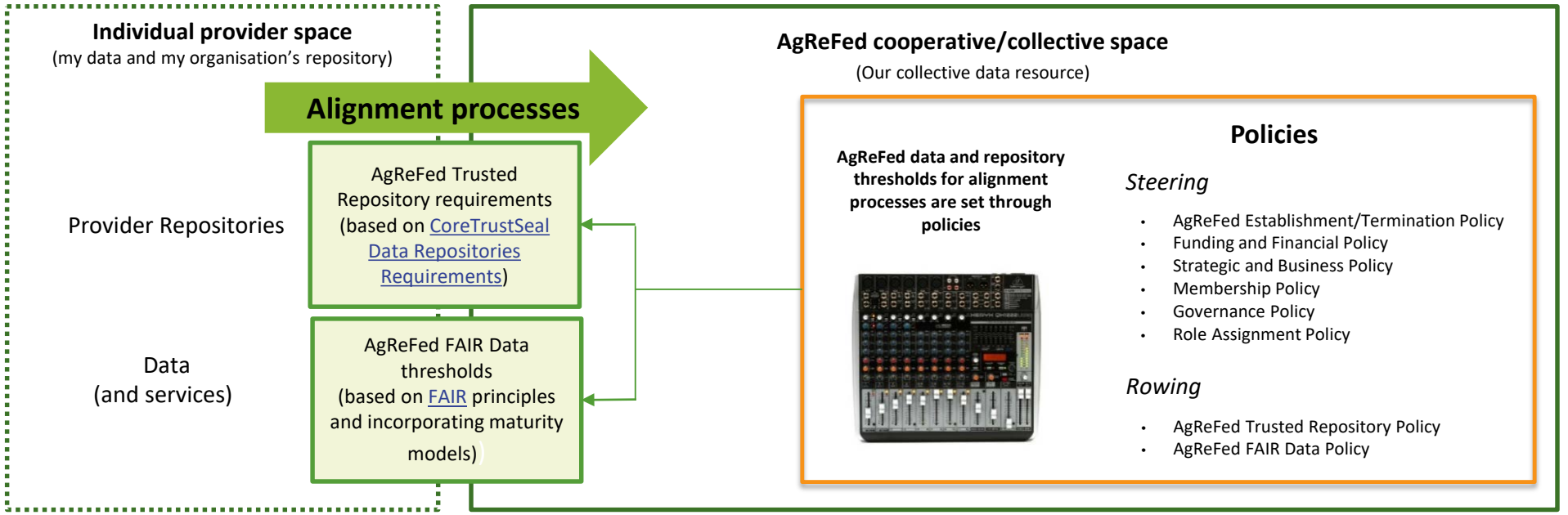
- Coordinates and administers standards development and governance

These roles could be filled by a member of a Provider Community or third parties tasked with the role. A single person may perform multiple roles.

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 AgReFed levels of FAIRness and repository trustedness
- AgReFed FAIR Data Policy (p21-22) and AgReFed Trusted Repository Policy (p23) are used to define the qualifying levels required for data to be provided as AgReFed Data



Key policy recommendations and priorities

There is a need to determine what kind of entity AgReFed participants will be interacting with e.g. partnering with. Is it Federation University Australia as the lead agency of AgReFed or a different (to be established) organisational form, for example a new legal entity such as cooperative or club? Clarifying the organisational form for AgReFed is a critical requirement that must be addressed before the other policies can be enacted. The following provides a brief list of the policies that will be required in order to establish and operate AgReFed and recommended priority actions.

Steering policies

AgReFed Establishment Policy

- Define the organisational form for AgReFed and policies for its creation and disbandment
- Co-develop a mission statement and strategic plan for the AgReFed, with AgReFed inaugural members to guide next steps

Membership Policy

- Define and formalise the responsibility of Provider Communities in AgReFed
- Define and endorse the protocols for joining and exiting AgReFed

Strategic and Business Policies

- High level business decisions such as areas of focus, strategy and engagement that guide future direction of AgReFed.
- Determine what if any business decisions are part of the Federation Council's decision domain. If not, who decides?

Funding and Finance Policies

- Are Provider Community members involved in financial decision making or does the lead agency and or Federation Council make these decisions?
- Establish policies for how financial matters are dealt with, such as securing and distributing funding.

Role Assignment Policies

- Refine, develop Terms of Reference (ToR) for and formalise (through the Federation Council) the initial recommended roles that have been defined in this document
- Define and formalise a process by which roles can be established and may be changed (and by whom) over time, for example the addition of new roles or the change in responsibilities or accountabilities for an existing role

Rowing policies

AgReFed FAIR Data Policy (p21-22)

- Adoption and promulgation of the AgReFed FAIR Data Policy, together with proposed initial AgReFed FAIR Data qualifying thresholds
- Determine permissible licensing and access arrangements for data e.g. the ability to advertise commercial data
- Agree on initial data standards

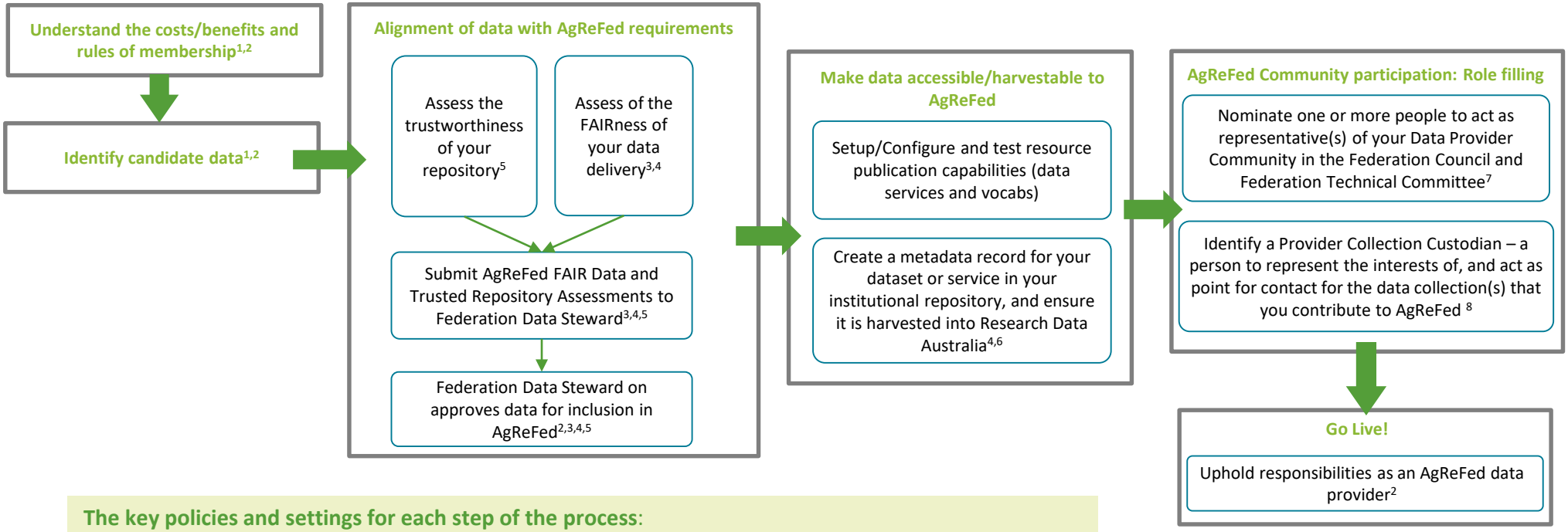
AgReFed Trusted Repository Policy (p23)

- Further define and adopt (via the TC) the AgReFed Trusted Repository Policy described in this document, together with proposed initial qualifying levels

Priority governance actions to establish AgReFed

1. ToRs for the Federation Council and Federation Technical Committee developed by the by AgRDC project Stewardship Framework team, and reviewed by Steering Committee
2. Establish the Federation Council and Federation Technical Committee by Data Providers nominating representative to sit on these bodies
3. At the inaugural meeting of the Council: formally establish the Council, Technical Committee, and key AgReFed roles; and sign off on ToRs and an initial set of Establishment and Membership policies (over time, the Funding and Financial policies and the Strategic and Business Policies would be developed and promulgated by the Council)
4. At a subsequent inaugural Technical Committee meeting, an initial core set of rowing policies would be promulgated

Process model for data provider participation



The key policies and settings for each step of the process:

1. AgReFed mission statement (See Establishment Policy for development, p10)
2. AgReFed Membership Policy (for development, see p10)
3. AgReFed FAIR Data Policy and initial qualifying thresholds (as proposed see p21-22, and see p38)
4. Minimum metadata requirements (as proposed see p36-37)
5. AgReFed Trusted Repository Policy and initial qualifying thresholds (as proposed see p23, and see p39)
6. These processes will be outlined within the Rowing (Technical) Policies (p10).
7. Outlined in 'The Federation Community – roles and accountabilities' p17: AgReFed Federation Council and Technical Committee membership
8. Outlined in 'The Federation Community – roles and accountabilities' p17: Provider Collection Custodian role

The Stewardship Framework ‘Viewpoints’

The Guidelines for the Stewardship Framework is described using the Reference Model for Open Distributed Processing (RM-ODP), which provides five architectural ‘Viewpoints’ for specifying distributed information systems; and is mainly focussed on the first:

- **Enterprise Viewpoint - purpose, scope and policies of a system, and socio-economic and institutional environment within which it operates (the social architecture)**
- Information Viewpoint - scope and nature of the data content within AgReFed (the information architecture)
- Technical Viewpoints - the Computational, Engineering and Technology Viewpoint configuration of the system (the technical architecture)

These Viewpoints are described in the next section of this document. The Enterprise Viewpoint is the focus of the document and has been articulated in detail, while the information and technical viewpoints are presented in high level summary form.

Enterprise Viewpoint

focuses on the purpose, scope and policies of a system and the socio-economic and institutional environment within which it operates

The Enterprise viewpoint describes the social architecture of the Agricultural Research Federation.

A federation is defined as 'an encompassing political or societal entity formed by uniting smaller or more localized entities: such as a federal government or a union of organizations.'[1] The term is also used as a key concept in enterprise architecture to capture the notion of independent autonomous components (actors, systems, policies, etc.) coming together in a collaborative effort.

The Enterprise Viewpoint presents

- A community-based approach to defining the Stewardship Framework
- The decision making mechanisms through which community decisions can be made
- The polices, roles and processes through which the community collaborates to achieve FAIR agricultural data

1 https://www.merriam-webster.com/dictionary/federation

AgReFed Agricultural Research Federation

Key framework concepts

The Enterprise Viewpoint defines relevant aspects of the organisational, business and social context within which the AgReFed exists. It aims to define the purpose of the system, its stakeholders and how they interact with the system, and their requirements. This document focusses on the Provider Community (rather than the end user community) and how data and other providers participate in and make decision about AgReFed.

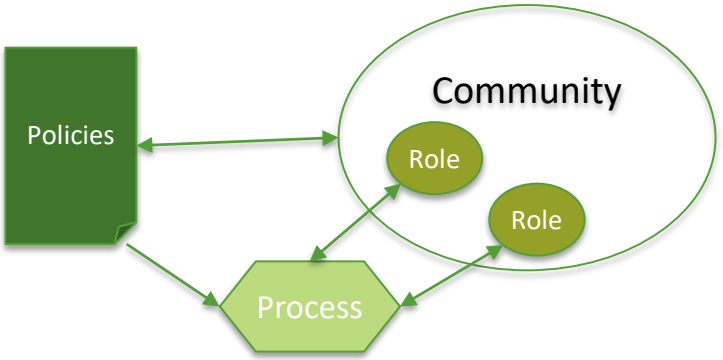
A **community** defines a set of participants and how they **behave** in order to achieve shared objectives.

The behaviour of the community is shaped and guided by the definition of **policies** (e.g. related to data standards or decision-making) that define how different **roles** interact within and between communities.

The relevant behaviours of the community can be described as a series of **processes** that conform to agreed policies (or rules).

For AgReFed, two key **alignment processes** are used as a means to align existing practices and approaches of individual data providers to agreed AgReFed approaches. The two principle alignment processes are AgReFed FAIR Data Policy including qualifying thresholds and AgReFed Trusted Repository requirements, based on FAIR principles and CoreTrustSeal requirements, respectively (see appendices p35. for more).

Policies set by the community determine acceptable minimum levels of FAIRness of data (p21-22) and trustworthiness of repositories (p23) that providers need to meet in order to share data through AgReFed.



Community - is a collection of participants (human and other entities) that interact to achieve shared objectives;

Policy- agreements that are produced by, and guide and shape the behaviour of the community;

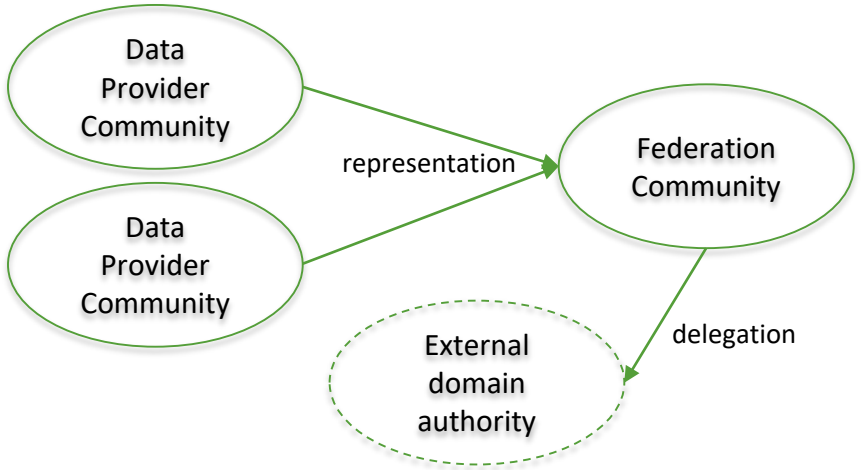
Roles - the roles that enable the AgReFed to function. They comprise governance and management roles, and the substantive stewardship roles that enable delivery of FAIR agricultural data;

Processes - are the interactions between roles, shaped by policies that enable the achievement of community objectives.

AgReFed communities

A number of discrete organisations are involved in the establishment and operation of the AgReFed. These are defined as communities:

- **The Federation Community** – this is the federated community composed of all the various roles that enable the AgReFed to operate;
- **Data Provider Communities** – a community involved in the provision of a data collection or dataset to the AgReFed, represented by multiple roles filled by actors from one or more organisation(s).
- **Domain Authority Communities** – the (virtual) organisations responsible for developing, publishing and governing vocabularies, information models and other kinds of standards. These may be pre-existing external communities that govern standards that are relevant for AgReFed.



It is anticipated that engagement with and delegation to external domain authorities will occur as the AgReFed matures and data standards and vocabularies are identified, adopted by AgReFed and incrementally applied. They are not dealt with in any more detail in this document.

The Federation Community – objectives, roles, and accountabilities

The Federation Community is responsible for two objectives:

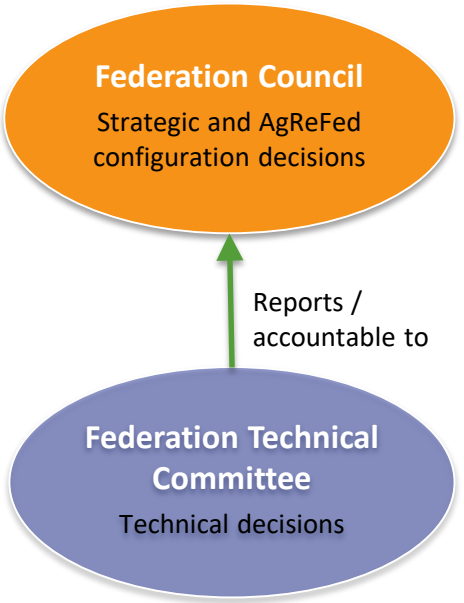
Objective 1: Steering – establishing, maintaining (and terminating) the AgReFed. This encompasses governance, coordination, management, and funding of the AgReFed to enable it to realise objective 2.

It is proposed that two levels of decision making are established through two related bodies, with representatives of each Data Provider Community nominated to act as members of each body:

- **Federation Council - Strategic and business decision making** – responsible for the overall direction, management and (business) operation of the AgReFed
- **Federation Technical Committee** – decisions about common technology choices, data standards and acceptable qualifying levels. The technical body would report and be accountable to the Federation Council.

These decision making bodies would be responsible for setting and maintaining the policies and determining the alignment processes for the AgReFed. Each body is defined in terms of its authority, functions, the domains over which it exercises decision authority, decision rights and representation (the proposed design of each decision making body is provided on the next page).

Objective 2: Rowing - achieving community objectives – defining the interactions between parties that are necessary to achieve its objective; that is the policies, roles, responsibilities and processes of the 'moving parts' of the system that ensure FAIR agricultural data can be discovered, accessed (from AgReFed Trusted Repositories), and used by end users of the AgReFed.



The Federation Community – roles and accountabilities

Objective 1: Steering - Function and characteristics of authority structures

	Federation Council	Federation Technical Committee
Source of authority / mandate	AgReFed Community with endorsement of AgRDC project Steering Committee will give itself the authority to make agreed decisions as in its ToR	Will be authorised and provided with delegation of decision rights by the Federation Council
Main functions	1. Establishment, and decisions about the overall direction and operation or the AgReFed. 2. Establishment and oversight of technical level authority and advisory structures (e.g. the Technical Committee and potentially a user reference group) 3. Oversight of the operations of the AgReFed 4. Representing AgReFed and communicating with those outside AgReFed 5. Engage with a authorise external domain authorities.	1. Making decisions about technical decision domains 2. Providing recommendations to the Federation Council in relation to agreed decision domains 3. Oversight of the alignment processes and Federations
Decision domains	**Enterprise Viewpoint – agreements and policies about:** 1. The purpose and objectives of the community 2. The strategic direction and (business) operation of the AgReFed 3. Collective business processes that the community will use 4. Deployment and maintenance procedures for Federation - testing deployment, implementation plans as well as the governance apparatus 5. Establishment of Federation roles 6. Exit strategy articulating approaches to exiting and terminating the AgReFed	**Information Viewpoint - agreements about:** 1. Determining what constitutes acceptable data for sharing through the AgReFed 2. Authorising the publication of data through the AgReFed 3. Determining acceptable levels of FAIRness for publication 4. Determining acceptable levels of AgReFed Trusted Repository requirements 5. Identifying common information models (data structures) that will be supported by the community 6. Determining agreed semantics - provider specific and agreed community vocabularies and ontologies **Computational Viewpoint – agreements about:** 1. The design and deployment of common infrastructure elements such as resource discovery mechanisms 2. The computational interfaces to be supported by each Provider Community 3. The supported end-user experience (portal) **Engineering and Technology Viewpoints** – where the components (services) are deployed using what technology
Decision rights	Decide on all decision domains specified above, as well as endorse Computational Viewpoint recommendations from Technical Committee	To be determined but broadly ... decide on information viewpoint, and identify Computational Viewpoint issues (for endorsement by the Federation Council)
Representation	Representatives of each Provider Community. Each community nominates one representative to the Federation Council and the Technical Committee. Observers can be included (with decision rights TBD)	Each Provider Community nominates one representative to the Technical Committee.
Decision making processes	One organisation one seat/voice Consensus based decision making (with vote or chair decides if no consensus is reached)	Representatives of all capability data and technology providers. TBC . Maybe a 'pigs and chickens'[1] model of commitment could be used to inform representation and decision rights.

1 – a fable about differing levels of commitment (skin in the game. See https://en.wikipedia.org/wiki/The_Chicken_and_the_Pig

The Federation Community – roles and accountabilities

Objective 2: rowing

It is proposed that two key substantive roles (and one optional role) be established:

Provider Collection Custodian

The Provider Collection Custodian role is filled by a nominated Provider Community representative to represent the interests of the collection in AgReFed and has responsibility for ensuring the collection's sustained FAIRness according to agreed policies.

Specific accountabilities include:

1. Ensuring that the AgReFed 'Contract' is honoured i.e. data remains FAIR according to agreed AgReFed FAIR Data Policy
2. The creation and maintenance of accurate collection metadata
3. Ensure that the metadata record is harvestable by Research Data Australia (RDA)
4. Collection vocabulary publication
5. Liaise with the Provider Community roles to ensure data collections' updates are consistent with Federation policies (e.g. around metadata quality)
6. Liaise with the repository manager to ensure sustained access

This Federation Community role maps most closely to the Provider Community collection manager role(s).

A Provider Collection Custodian may be responsible for more than one collection and a collection must have at least one Custodian.

Federation Data Steward

A Federation Data Steward is responsible for administering the policies and alignment processes that are set by the Federation Council and Technical Committee by working with Data Provider Communities to assist them in complying with agreed policies.

Responsibilities include:

1. Approving data sets for publication
2. Liaising with and providing support to Data Provider Communities to complete their AgReFed FAIR Data and Trusted Repository Self-assessments
3. Providing advice, guidance and technical support to Data Provider Communities
4. Validating collection metadata
5. Validating AgReFed FAIR Data Self-assessments
6. Validating AgReFed Trusted Repository Self-assessments

Federation Standards and Vocabularies Steward (optional)

- Coordinating standards and vocabulary development
- Validating published vocabularies
- Coordinating standards and vocabulary governance

Data Provider Community – function and roles

Each AgReFed data provider is represented as its own community. This community comprises the actors engaged in operating the repository, establishing services and curation, management and provisioning of data (including metadata). Each community is **autonomous** and makes its own decisions about how it organises itself.

Data Provider Communities are represented in the AgReFed through nominated representatives filling Federation Community decision making roles on the Federation Council and Technical Committee (page 17), and the Provider Collection Custodian role (page 18).

Defined, scoped, configured by the data provider independently of the AgReFed.

A Provider Community could be scoped to:

- A single data set or collection
- An organisation with multiple collections
- Multiple collections from multiple organisations e.g. a data cooperative
- Providers of common infrastructure and services as part of the AgReFed

Possible Data Provider Community roles

A number of Provider Community roles necessary for the provision of FAIR data are *suggested*, and are mapped to Collection-Party relations from the RIF-CS schema (as used by Research Data Australia), to assist the representation of these roles in collection metadata records.

Provider Community roles	Collection-Party relation (RIF-CS)	Description
Provider Collection Creator	hasCollector	has been collected, generated, created or aggregated by the related party
Provider Collection Manager/Custodian	isManagedBy	is maintained and made accessible by the related party (includes custodian role)
Provider Collection Owner	isOwnedBy	legally belongs to the related party
Provider Collection Enhancer	isEnrichedBy	additional value provided to a collection by a party (i.e. formatting or describing to enable sharing and reuse) who is not already represented by another role, e.g. manager
Provider Service Provider	isManagedBy	provider of vocab service, data service, repository as a service

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 AgReFed levels of FAIRness and repository trustedness
- **AgReFed Trusted Repository Policy** and **AgReFed FAIR Data Policy** are used to define the qualifying levels required for data to be provided as AgReFed Data

AgReFed FAIR Data policy (p21-22) is based on FAIR principles.

The AgReFed FAIR Data Self-assessment is based on the ARDC FAIR Data Self-assessment Tool but incorporates maturity models to measure and track increases in interoperability delivered through services.

The development of the Assessment is explained further in the Appendix, p38.

AgReFed Trusted Repository Policy (p23) is based on CoreTrustSeal Data Repositories Requirements.

The development of the Assessment is explained further in the Appendix, p39.

Rowing: AgReFed FAIR Data Policy

The AgReFed FAIR Data Policy

1. AgReFed FAIR Data Self-assessment (p22) will be used as the data alignment process for AgReFed
2. Current thresholds required for qualification as AgReFed FAIR data are shown in the table on page 22
3. Standards for data provision including vocabulary standards will be defined as part of the AgReFed FAIR Data Policy settings
4. **Additional data structure semantics and syntax standards may be specified as part of the AgReFed FAIR Data Policy**
5. The Federation Data Steward will review and approve (or reject) data providers' AgReFed FAIR Data Self-assessments
6. Any disputes in relation to the validation of assessment will be escalated to the Federation Technical Committee for review and decision
7. The FAIR assessment process and settings (including qualifying threshold levels) may be modified by the Federation Technical Committee

See page 22 for detail

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 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)	
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	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			
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	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 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			

AgReFed FAIR Data Policy 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

* Question 3 specifies minimum metadata requirements for collections and services (see appendix – pages 36-37)

AgReFed FAIR Data assessment – Initial settings

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 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)
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 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
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		
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	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 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		

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

* Question 3 specifies minimum metadata requirements for collections and services (see appendix – pages 36-37)

Rowing: AgReFed FAIR Data Policy and initial settings

AgReFed Trusted Repository Policy

1. The AgReFed Trusted Repository Self-assessment process will be used as the AgReFed alignment process
2. Assessment (scope and requirements) may be modified by the AgReFed TC
3. Current thresholds for qualification are proposed as shown in the table on this page. If R2, R3, R4, R11 and R13 to R16 requirements are met, the repository qualifies as an AgReFed Trusted Repository
4. An additional requirement that repository metadata must be harvestable by Research Data Australia (RDA)
5. The Federation Data Steward will review and qualify (or reject) Data Provider Communities' AgReFed Trusted Repository Self-assessment.
6. Any disputes in relation to the validation of assessment will be escalated to the AgReFed TC for review and decision
7. Qualifying threshold levels can be reset by AgReFed TC

	Requirement	AgReFed policy
R1	**Mission/Scope**: The repository has an explicit mission to provide access to and preserve data in its domain	
R2	**Licenses**: The repository maintains all applicable licenses covering data access and use and monitors compliance	Yes
R3	**Continuity of Access**: The repository has a continuity plan to ensure ongoing access to and preservation of its holdings	Yes
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	Yes
R5	**Organisational Infrastructure**: The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission	
R6	**Expert Guidance**: The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant)	
R7	**Digital Object Management**: The repository guarantees the integrity and authenticity of the data	
R8	**Appraisal**: The repository accepts data and metadata based on defined criteria to ensure relevance and understand ability for data users	
R9	**Documented Storage Procedures**: The repository applies documented processes and procedures in managing archival storage of the data	
R10	**Preservation Plan**: The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way	
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	Yes
R12	**Workflows**: Archiving takes place according to defined workflows from ingest to dissemination	
R13	**Data Discovery and Identification**: The repository enables users to discover the data and refer to them in a persistent way through proper citation	Yes
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	Yes
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.	Yes
R16	**Security**: The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users	Yes

Information Viewpoint

defines the scope of data, and specifies standards used to achieve interoperability

Principles

The AgReFed data integration and interoperability principles are:

- The data provider's repository should be trustworthy (page 40)
- The data provider follows the FAIR principles (i.e. the data should be as "FAIR" as possible) (page 40)
- The dataset interoperability should be mature (page 26)

Data scope

The initial focus was on the AgRDC project exemplar datasets, which include crop yield, rotation information, weather and climate, hyper- and multi-spectral imagery, molecular analysis, and soil measurements (from sensors and other sources).

The full scope of potential data to be made available through the AgReFed is much broader, and indeed could be **any data from Agriculture production, or Agriculturally-relevant research**, as defined by the Federation Council.

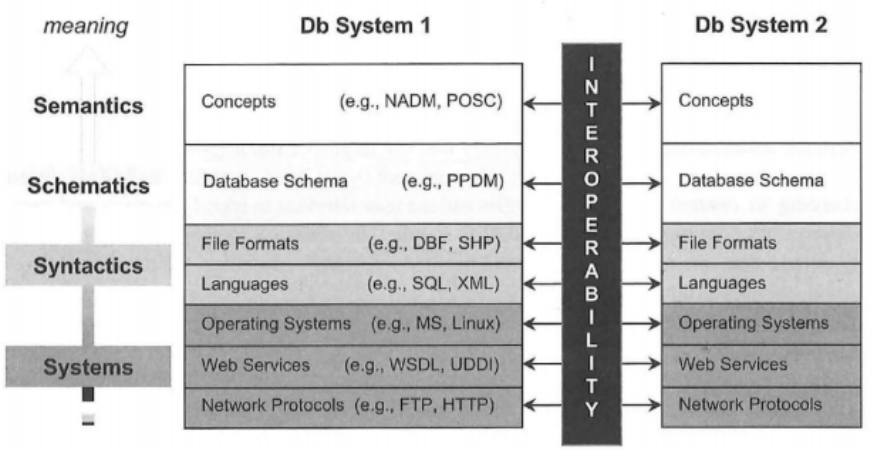
AgReFed
Agricultural Research Federation

Data interoperability maturity

Where possible, data should be made interoperable through the adoption of agreed ontologies and vocabulary services (semantic interoperability), agreed data structures (schematic interoperability), and standard communication protocols (syntactic interoperability), along with the adoption of appropriate social architecture for maintaining cross-domain interoperability.

From a technical perspective, levels of interoperability[1] are:

- **semantic interoperability**: harmonised concepts and content (for example through the use of common controlled vocabularies in standard formats such as SKOS);
- **schematic interoperability:** agreed data structures achieved through the use of common information exchange models that define appropriate data types and structures (such as O&M, ANZSoilML, SKOS);
- **syntactic interoperability** achieved through the use of common file formats (such as NetCDF, CSV) and exchange languages (such as XML, JSON, RDF), and spatial representations or patterns for using the standard formats for geographic data (such as GML, KML, GeoJSON, TopoJSON, SHP); and
- **system or technical interoperability** achieved through the use of standard communication protocols (such as HTTP, WSDL) and standard delivery mechanisms (such as OGC-WFS, OGC-WCS, OGC-WMS, Net-CDF).



Interoperability levels increasing vertically for greater exchange of meaning between database systems[1]

1 Brodaric, B. and M. Gahegan (2006). "Representing geoscientific knowledge in cyberinfrastructure: Some challenges, approaches, and implementations." Geological Society of America Special Papers 397: 1-20.

AgReFed data interoperability implementation

Vocabularies (semantics)

AgReFed data must use published **vocabularies** to describe concepts within the data.

These may vary from flat vocabulary term lists provided and managed by an AgReFed data provider and hosted in Research Vocabularies Australia (RVA), to domain specific vocabularies such as soil vocabularies hosted by CSIRO's Linked Data Registry (LDR), to externally governed and managed vocabularies, such as the Quantities, Units, Dimensions and Data Types (QUDT) ontologies and the AGROVOC controlled vocabulary.

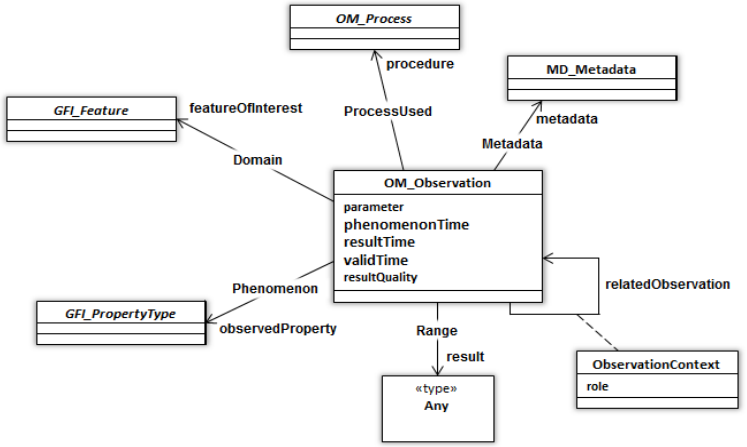
Each vocabulary used by providers will need to be assessed as to its appropriateness and governance arrangements, as part of the AgReFed FAIR data assessment (p22).

In future, AgReFed may declare and publish AgReFed vocabularies, i.e. vocabularies that are to be adopted by the community as common terms used in the data they publish.

Information models (schematics)

Observational data: The preferred AgReFed implementation for observational data (e.g. crop yield or soil moisture) is the ISO/OGC Observations & Measurements (O&M) information model. Specifically, the OGC O&M XML specification.

Physical sample data: Likewise, OGC O&M XML specification can be used for for physical sample and specimen data, or a domain specific variation, such as ANZSoilML.

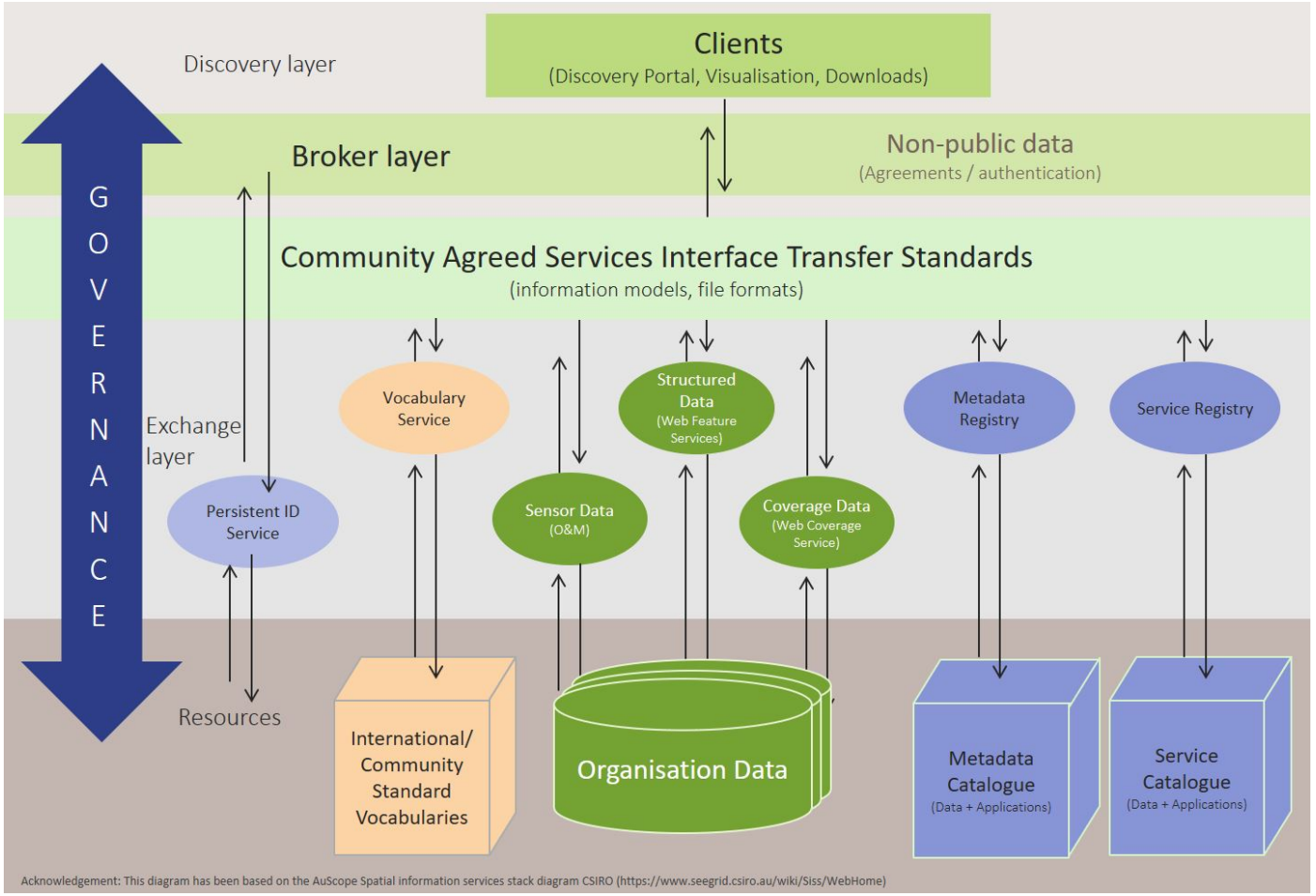


A UML representation of the ISO/OGC Observations & Measurements information model (Source: Cox, Simon. Geographic Information: Observations and Measurements. Online: Open Geospatial Consortium Inc.; 2010. http://www.opengis.net/doc/om/2.0)

Computational Viewpoint

focuses on the distribution of a system through functional decomposition into objects which interact at interfaces

AgReFed Computational Viewpoint Stack



This diagram provides a high level view of the interactions between provider and users via provider and common computational components.

Engineering Viewpoint

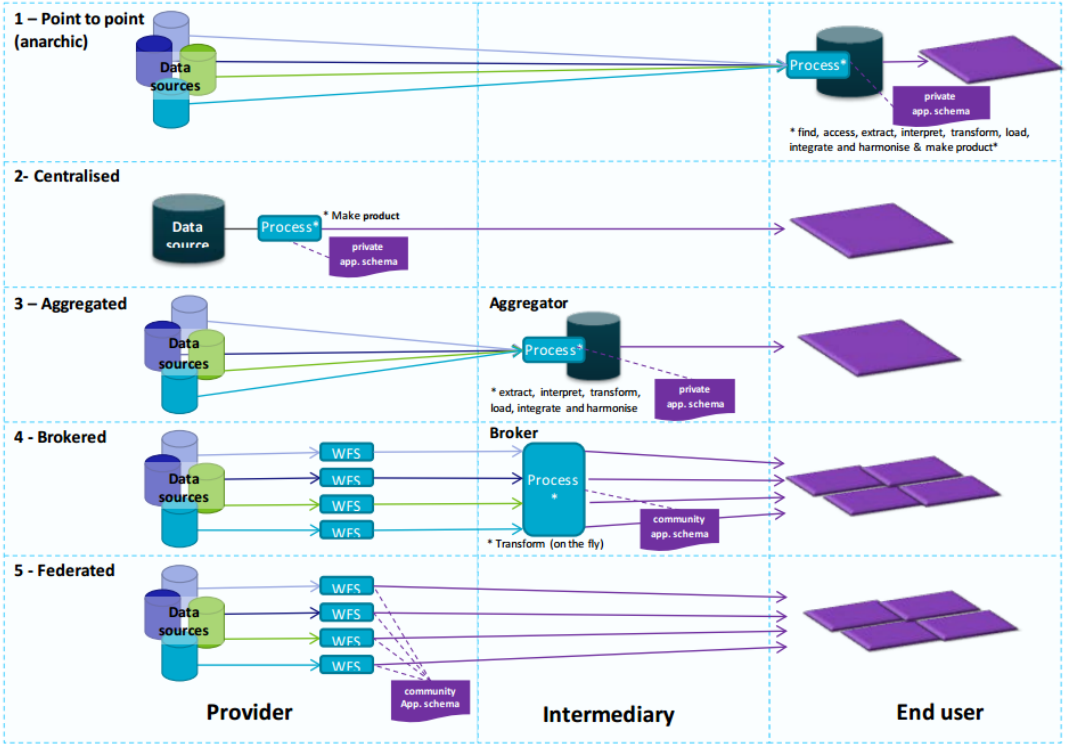
focuses on the mechanisms and functions required to support distributed interaction among objects in the system

Distributed data supply chain patterns

The preference for the AgReFed is to use a **federated** approach to data supply.

However, where it makes integration easier to achieve, elements of the brokering and aggregation patterns will also be used.

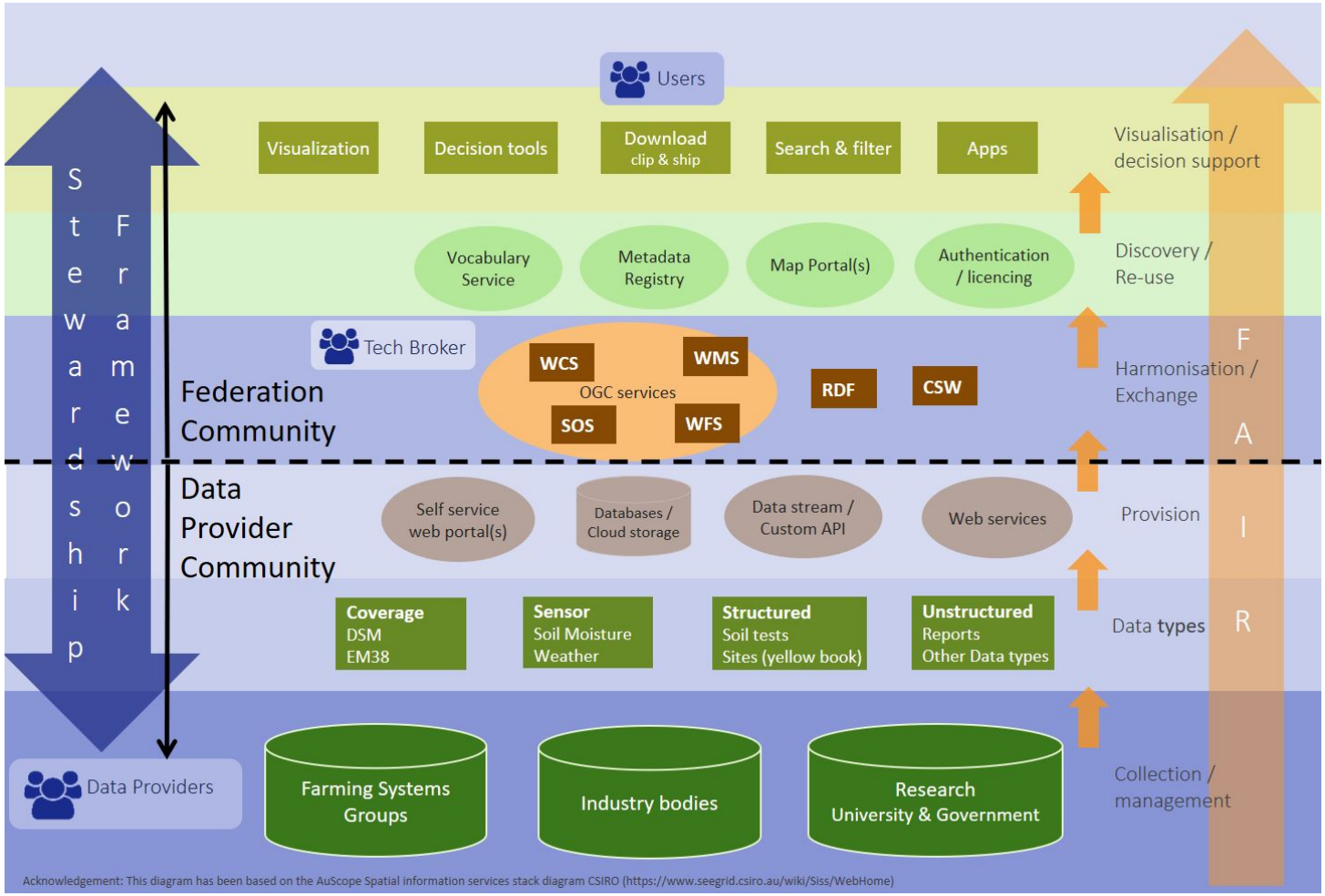
- For federated and brokered data, source data resides with the data provider system. Both solutions encourage currency and validity of data.
- A standards based Service Oriented Architecture (https://en.wikipedia.org/wiki/Service-oriented_architecture) is utilised, including metadata cataloguing and vocabulary linking. These 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.



Distributed data supply chain patterns (Box et al., 2015)

1 https://publications.csiro.au/rpr/pub?pid=csiro:EP155525

AgReFed Engineering viewpoint



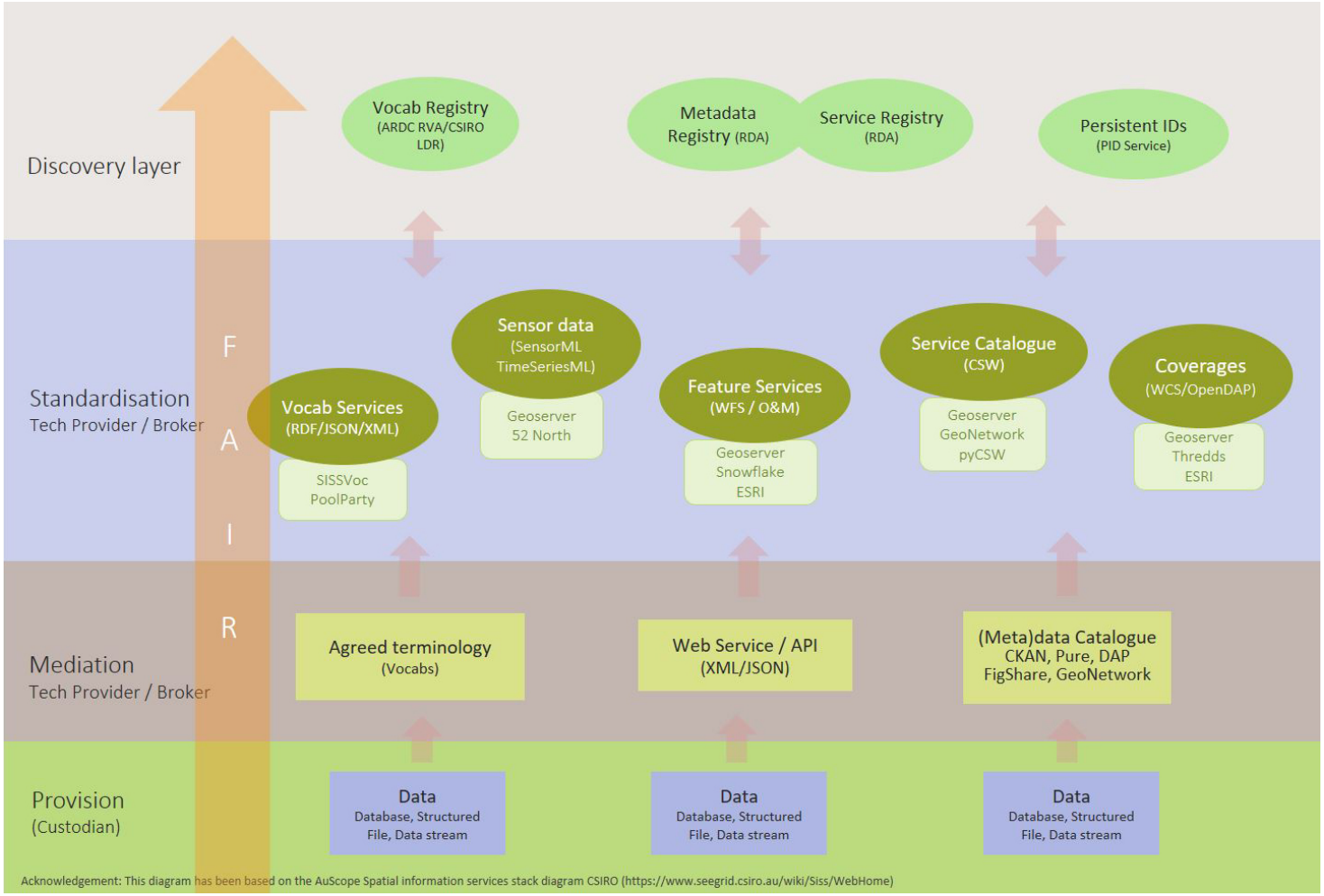
This diagram provides a view of the components, mechanisms and functions required to support distributed interactions among objects in the system, including whose responsibility they are. When this diagram is fleshed out in more detail it will show where components are deployed.

Technology Viewpoint

focuses on the choice of technology for a system

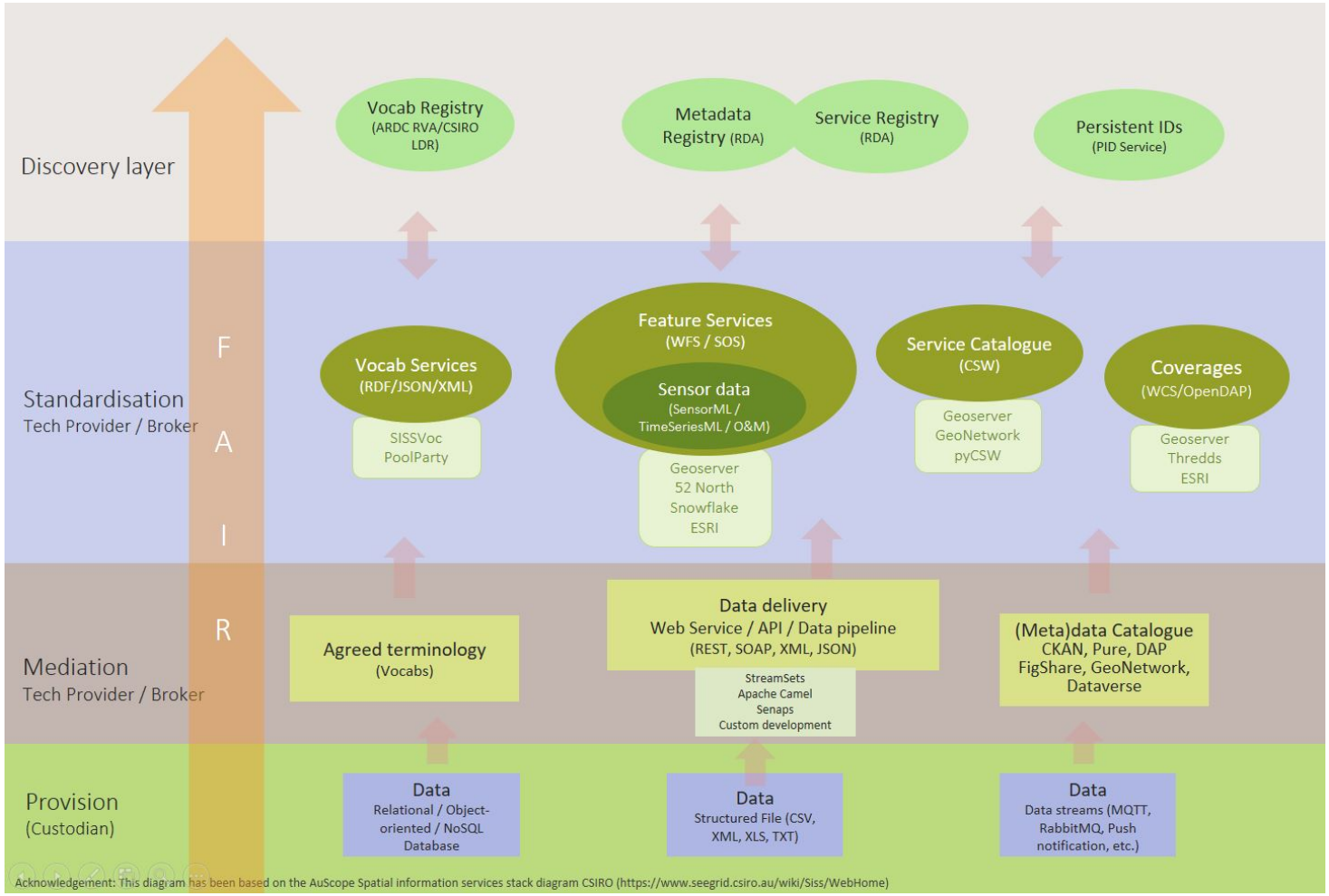
"*Technology standards allow different systems and services to work together through standard interfaces. Ideally, when the standards are implemented in products or online services independently, the resulting components 'plug-and-play', that is, they work together seamlessly.*"
Reference Model for Open Distributed Processing: RM-ODP

AgReFed Technology viewpoint – Feature instance



This diagram provides a **(spatial) 'feature oriented' view** of the technology stack which emphasises the delivery of geospatial feature data

AgReFed Technology viewpoint – Sensors instance



Acknowledgement: This diagram has been based on the AuScope Spatial information services stack diagram CSIRO (https://www.seegrid.csiro.au/wiki/Siss/WebHome)

This diagram provides a sensor oriented view of the technology stack which emphasises the delivery of sensor data.

Appendices

Minimum metadata elements for collection records

Minimum metadata requirements for AgReFed Collection Descriptions in Research Data Australia, to meet AgReFed FAIR Data Policy (Q3).
Fields based on RIF-CS schema (as used by Research Data Australia).
See https://documentation.ands.org.au/display/DOC/Collection for more information on each field.

Information type (field)	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, AGRIVOC terms should be used.
Spatial coverage (if relevant)	The geometry for the location the resource relates to.
Temporal coverage (if relevant)	The time period the resource relates to, in W3C Date/Time Format.

Minimum metadata elements for service records

Minimum metadata requirements for AgReFed Collection Descriptions in Research Data Australia, to meet AgReFed FAIR Data Policy (Q3). Fields based on RIF-CS schema (as used by Research Data Australia).

See https://documentation.ands.org.au/display/DOC/Collection for more information on each field.

Information type (field)	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.
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, AGRIVOC terms should be used.
Spatial coverage*	The geometry for the location the resource relates to (a point or a polygon).
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

AgReFed FAIR Data – alignment process and policy settings

Overview

The FAIR principles[1,2] 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[3]. They provide for a continuum of increasing Reusability or 'FAIRness', rather than a prescriptive standard which can be failed.

The principles reference community standards and best-practices, rather than defining a specific implementation. It should also be noted that FAIR does not necessarily equal 'Open' nor 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.

Data providers can use the AgReFed FAIR Data Self-Assessment (see page 21) to assess data products, and inform how each component of findability, accessibility, interoperability and reusability can be improved incrementally.

How is FAIRness assessed in AgReFed?

Inaugural AgReFed data providers used the ARDC FAIR Data Self-assessment Tool to assess the FAIRness of their data products. In response to AgReFed data provider feedback relating to the complexity of data and data-service relationships that couldn't be accommodated by the original assessment, the questions from the ARDC FAIR Data Self-assessment Tool were modified. This formed the AgReFed FAIR Data Self-assessment (see page 21).

AgReFed FAIR Data Policy

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 be need to be met for sharing of that data through the AgReFed. These levels may vary between datatypes (e.g. genomics, yield, etc) and may change over time. A description of the current AgReFed FAIR Data Policy settings are described on page 21.

The completed AgReFed FAIR Data Self-Assessments will be reviewed and validated by Federation Data Stewards against the agreed AgReFed FAIR Data Policy.

1 Wilkinson, Mark 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
2 https://www.go-fair.org/fair-principles/
3. Mons, Barend, 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

AgReFed Trusted Repositories - alignment processes and policies

Overview

In a distributed information system such as the AgReFed, the data providers' repositories need to be reliable and trustworthy. The CoreTrustSeal certification provides a process whereby custodians can measure repository compliance levels against sixteen identified characteristics of trustworthy repositories, the Core Trustworthy Data Repositories Requirements (hereafter referred to as the CoreTrustSeal requirements):

R1. Mission/Scope
R2. Licenses
R3. Continuity of Access
R4. Confidentiality/Ethics
R5. Organisational Infrastructure
R6. Expert Guidance
R7. Digital Object Management
R8. Appraisal
R9. Documented Storage Procedures
R10. Preservation Plan
R11. Data Quality
R12. Workflows
R13. Data Discovery and Identification
R14. Data Reuse
R15. Technical infrastructure
R16. Security

How are the CoreTrustSeal requirements being applied in AgReFed?

As with the AgReFed FAIR Data Self-assessment, providers wishing to share data via the AgReFed will conduct an assessment of the repository through which the data will be made persistently available (AgReFed Trusted Repository Self-assessment).

The AgReFed uses a simplified process based on the CoreTrustSeal requirements, but instead of five levels of compliance for each of the requirements, AgReFed assesses whether each requirement has been implemented or not.

The Data Provider Communities will undertake AgReFed Trusted Repository Self-assessments of their repositories. The Federation Data Steward (or other agreed community role) will review and validate the assessment against the agreed AgReFed Trusted Repository Policy.

AgReFed Trusted Repository Policy

The AgReFed Trusted Repository Policy set by the AgReFed (through the Federation Technical Committee) will be used to determine the acceptable level of trustworthiness that will be need to be met for sharing of data through the AgReFed.

Initially it is proposed that for AgReFed the repository meet requirements relating to Licences (R2), Continuity of Access (R3), Confidentiality/Ethics (R4), Data Quality (R11), Data Discovery and Identification (R13), Data Reuse (R14), Technical Infrastructure (R15), and Security (R16). These requirements may change over time (see page 22).

Glossary

Term	Definition
Agricultural or Agriculture data	Data from Agriculture production, or Agriculturally-relevant research, as defined by the Federation Council.
AgReFed Data	Data presented by data providers that is compliant with the 'contract' between Provider Communities and the Federation Community (as expressed through AgReFed policies, roles and alignment processes)
AgReFed FAIR Data	Data that has been assessed by the data provider and reviewed (by the Federation Data Steward) as being compliant with the AgReFed FAIR Data Policy. The collective data resource that is a culmination of the Federation Community (composed of all the various roles that enable the AgReFed to operate) and AgReFed policies and alignment processes.
AgReFed Trusted Repository	Is a repository that has been assessed and certified as compliant with the AgReFed Trusted Repository Policy by the Federation Collections Steward
Alignment processes	
Contract	The 'Contract' between Provider Communities and the Federation Community is expressed through AgReFed policies, roles and alignment process
Data governance	The exercise of authority and control (planning, monitoring, and enforcement) over the **management** of **data** assets (Source: DAMA (2009) DAMA Guide to the Data Management Body of Knowledge)
Federation Community	The community composed of all the various roles, that enable the AgReFed to operate.
Machine readable	Format in a standard computer language (not English text) that can be read automatically by a web browser or computer system. (e.g. xml). Traditional word processing documents and portable document format (PDF) files are easily read by humans but typically are difficult for machines to interpret. Other formats such as extensible markup language (XML), JSON, or spreadsheets with header columns that can be exported as comma separated values (CSV) are machine readable formats. As HTML is a structural markup language, discreetly labeling parts of the document, computers are able to gather document components to assemble tables of contents, outlines, literature search bibliographies, etc. https://en.wikipedia.org/wiki/Machine-readable_data
Observational data	Measurement of an observation of a phenomenon, e.g. air temperature, crop yield, or soil moisture.
Physical sample / specimen	A sample collected of a physical object, e.g. soil sample.
Stewardship	Data stewardship is the management and oversight of data assets, that do not (necessarily) belong to the stewards themselves, in order to improve their findability, accessibility, reusability, and quality.


AgReFed
Agricultural Research
Federation