Agricultural Research Federation (AgReFed)

A federated agricultural research data community

Bruce Simons, Andrew MacLeod
Federation University Australia

Project Partners

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.

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).

Stewardship and Governance Framework

- Aims to improve the sharing and reuse of agricultural data.
- Describes a socio-technical system that:
  - Brings independent organisations together
  - Will guide agricultural data providers; and
  - Be implemented by data provider using common technical infrastructure.
- The ‘social architecture’ key concepts:
  - Independent and autonomous **Data Provider Communities**
  - The collective **AgReFed Community** within which they participate;
- Guiding principles
  - FAIR Data
  - Trusted Repository (Core Data Seal)

Findable Interoperable
Accessible Reusable

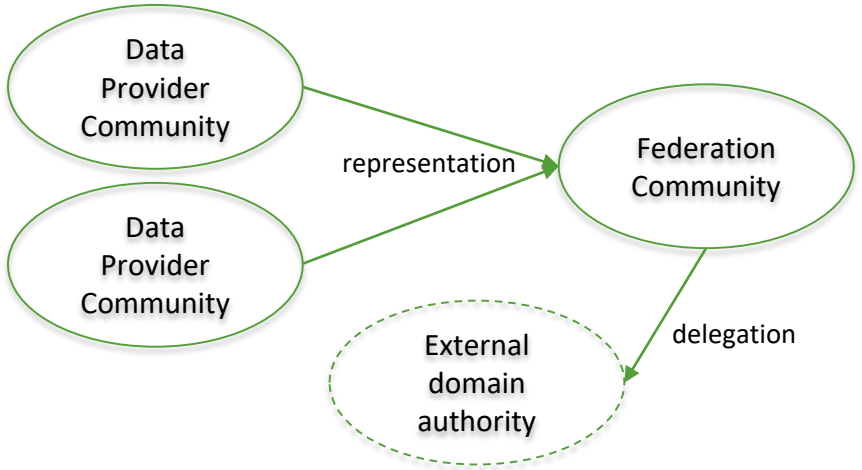
CORE TRUST SEAL

AgReFed

AgReFed communities

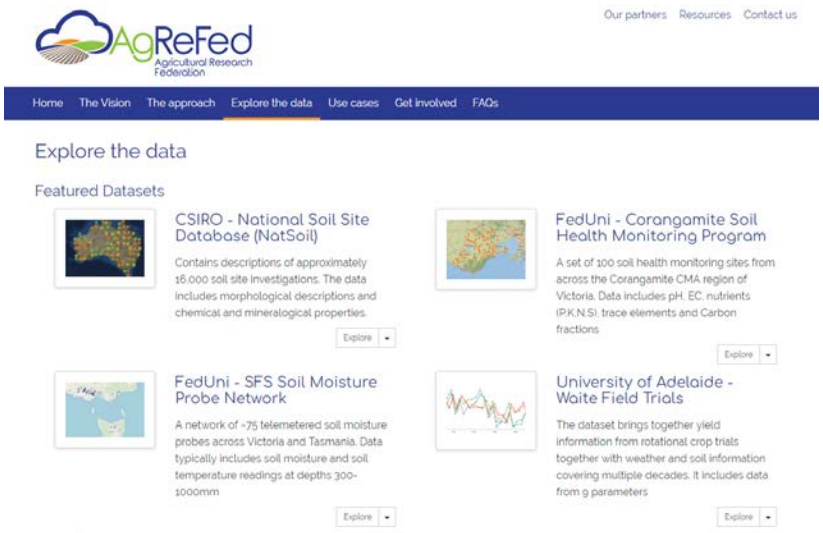
Communities

- **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.

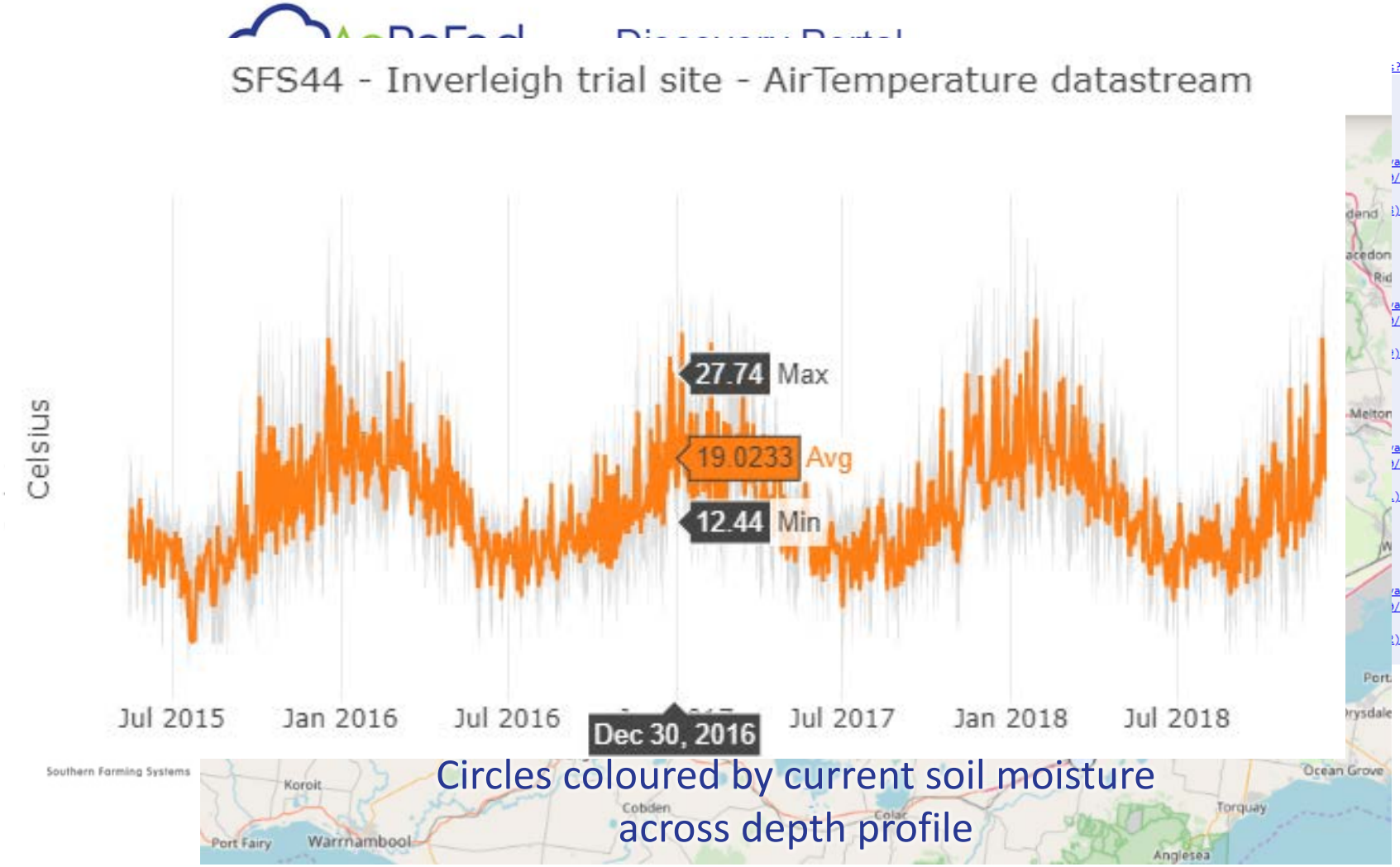


Exemplar datasets

- CSIRO - National Soil Site Database (NatSoil)
- CSIRO – Soil and Landscape Grid of Australia
- FedUni - SFS Soil Moisture Probe Network
- FedUni - Corangamite Soil Health Monitoring Program
- University of Adelaide - Waite Field Trials
- UWA/DPIRD – Frost Nursery Trials
- UNE – SMART Farm SensorNETS



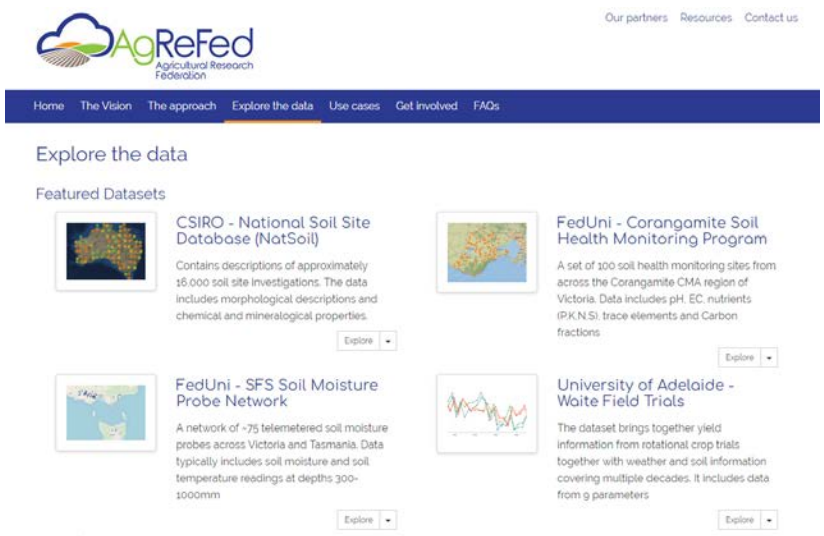
Sensor network - Soil Moisture and Weather



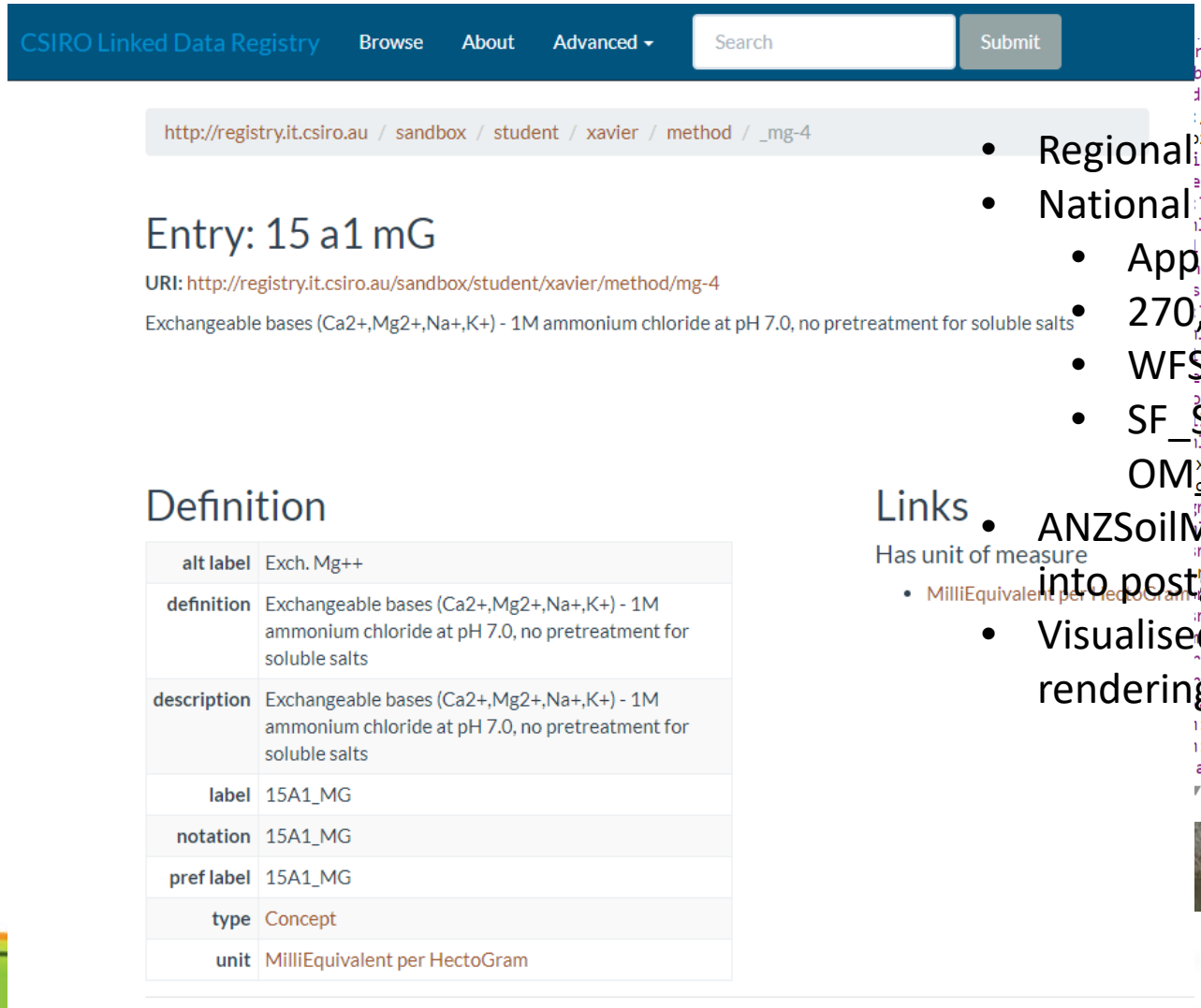
Circles coloured by current soil moisture across depth profile

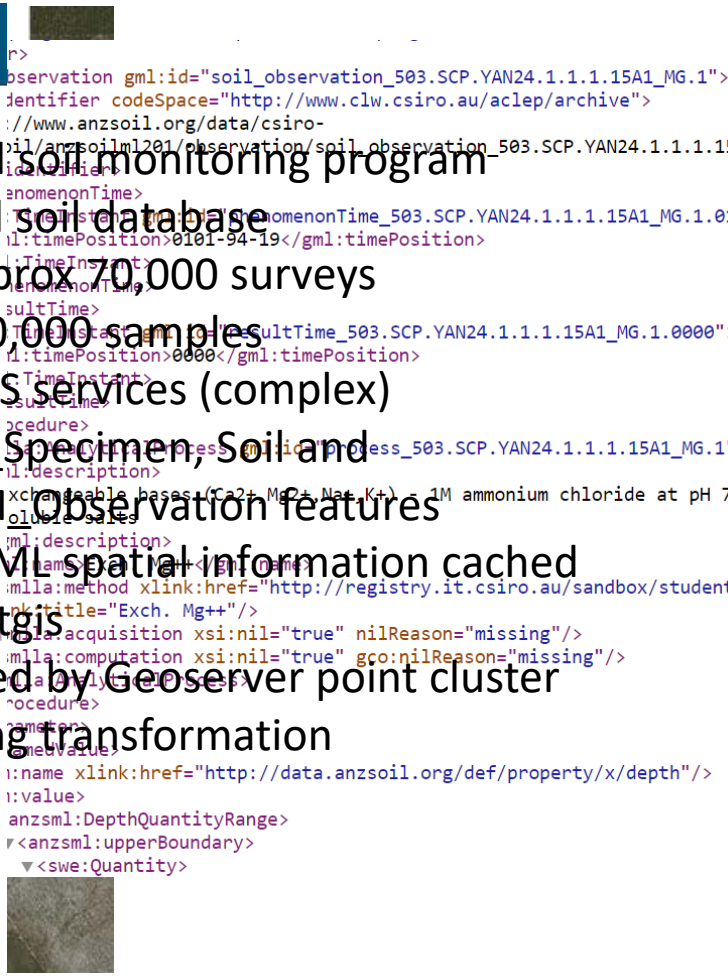
Exemplar datasets

- CSIRO - National Soil Site Database (NatSoil)
- CSIRO – Soil and Landscape Grid of Australia
- FedUni - SFS Soil Moisture Probe Network
- FedUni - Corangamite Soil Health Monitoring Program
- University of Adelaide - Waite Field Trials
- UWA/DPIRD – Frost Nursery Trials
- UNE – SMART Farm SensorNETS



Soil sampling and observations

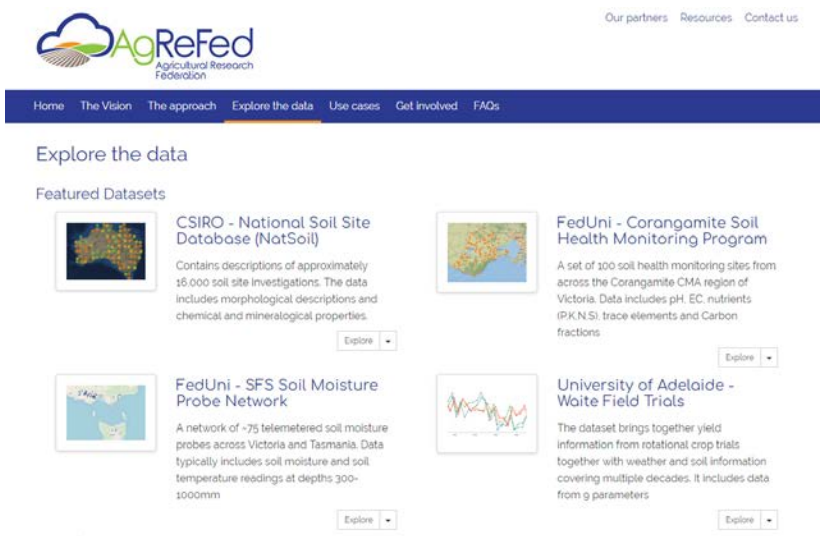




- Regional soil monitoring program
- National soil database
  - Approx 70,000 surveys
  - 270,000 samples
  - WFS services (complex)
  - SF_Specimen, Soil and OM_Observation features
- ANZSoilML spatial information cached into postgis
- Visualised by Geoserver point cluster rendering transformation

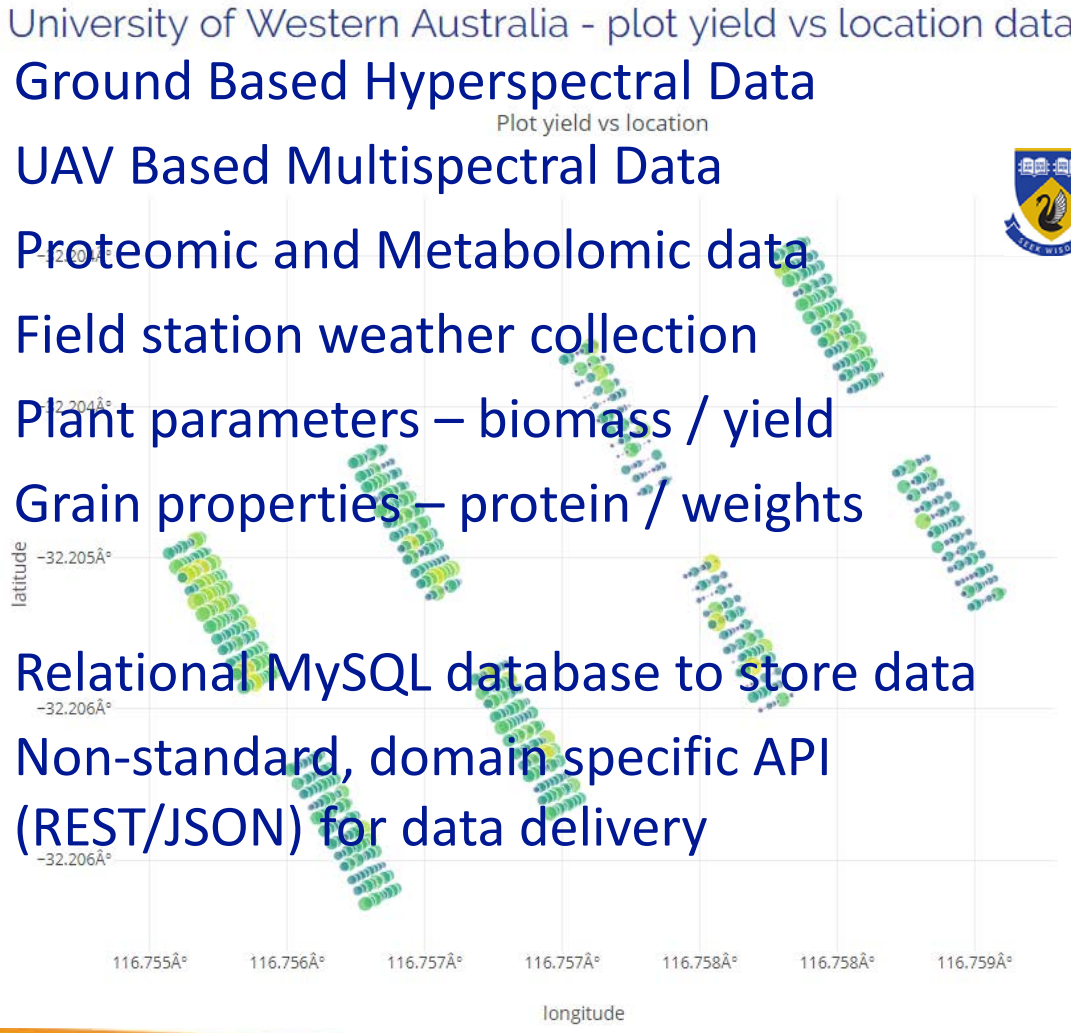
Exemplar datasets

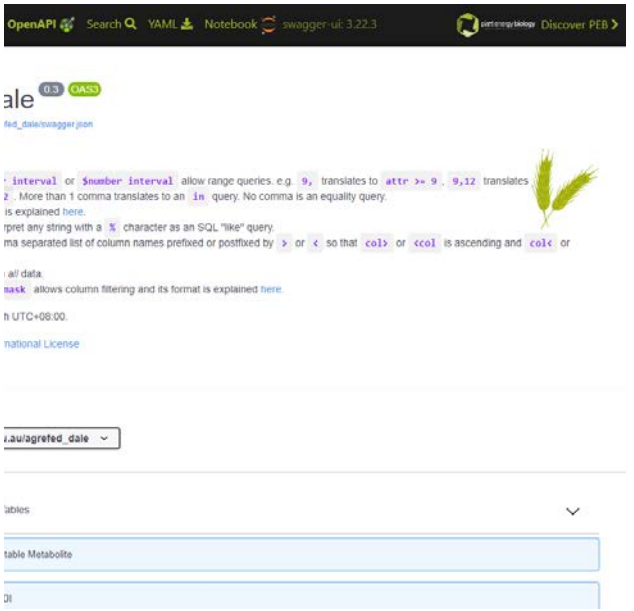
- CSIRO - National Soil Site Database (NatSoil)
- CSIRO – Soil and Landscape Grid of Australia
- FedUni - SFS Soil Moisture Probe Network
- FedUni - Corangamite Soil Health Monitoring Program
- University of Adelaide - Waite Field Trials
- UWA/DPIRD – Frost Nursery Trials
- UNE – SMART Farm SensorNETS



Field and greenhouse trial research

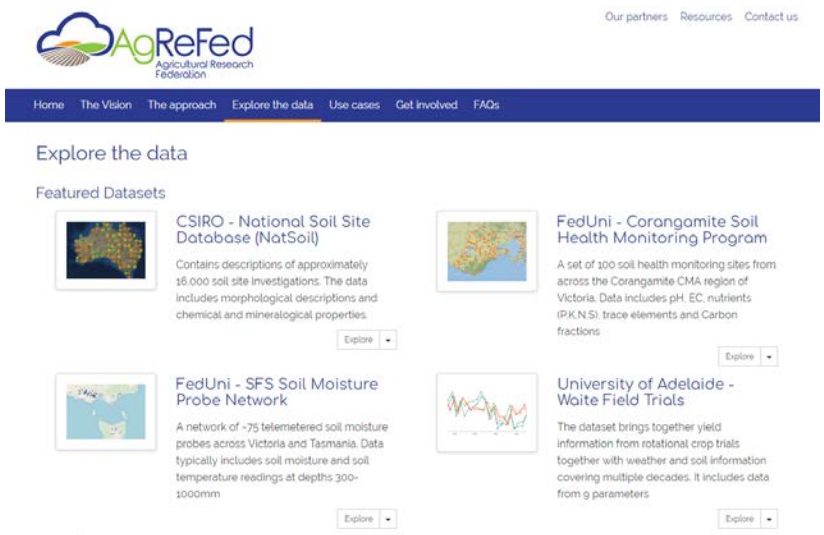
- Ground Based Hyperspectral Data
- UAV Based Multispectral Data
- Proteomic and Metabolomic data
- Field station weather collection
- Plant parameters – biomass / yield
- Grain properties – protein / weights
- Relational MySQL database to store data
- Non-standard, domain specific API (REST/JSON) for data delivery





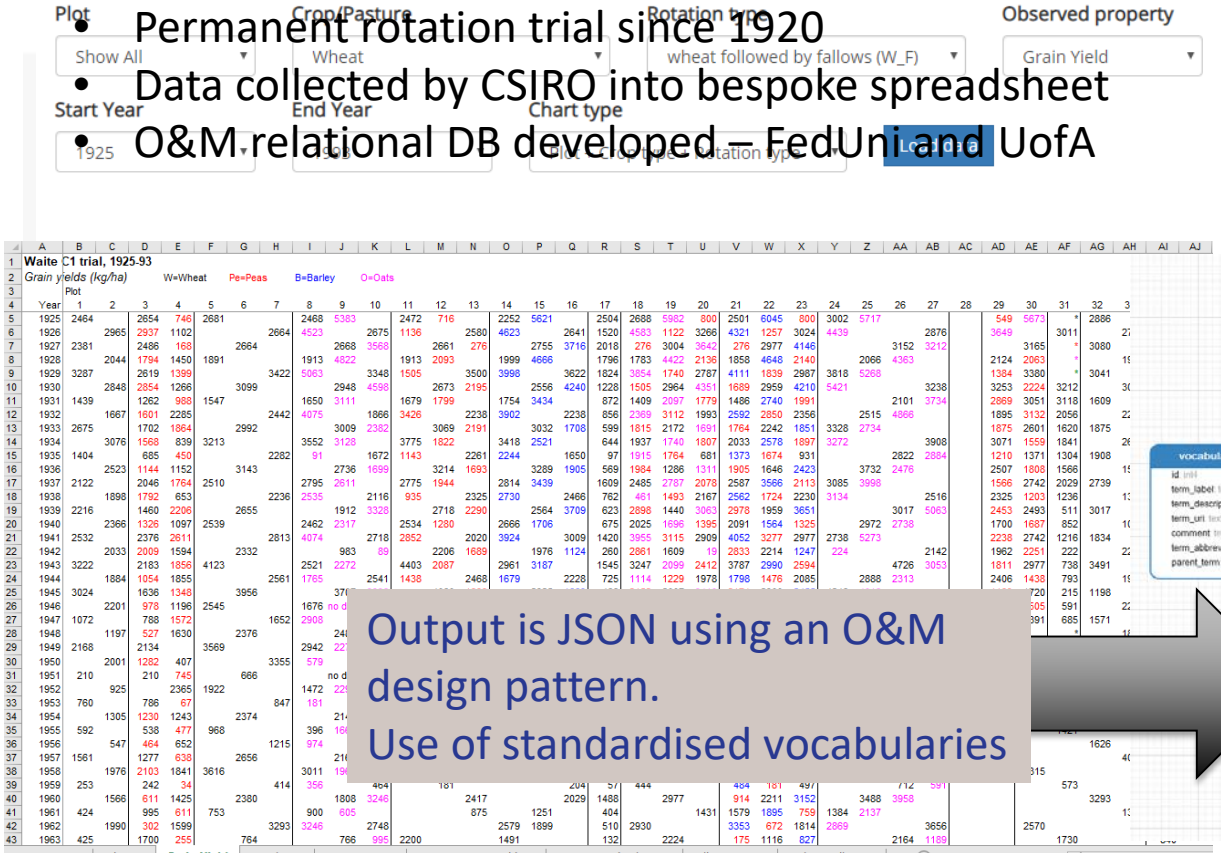
Exemplar datasets

- CSIRO - National Soil Site Database (NatSoil)
- CSIRO – Soil and Landscape Grid of Australia
- FedUni - SFS Soil Moisture Probe Network
- FedUni - Corangamite Soil Health Monitoring Program
- University of Adelaide - Waite Field Trials
- UWA/DPIRD – Frost Nursery Trials
- UNE – SMART Farm SensorNETS

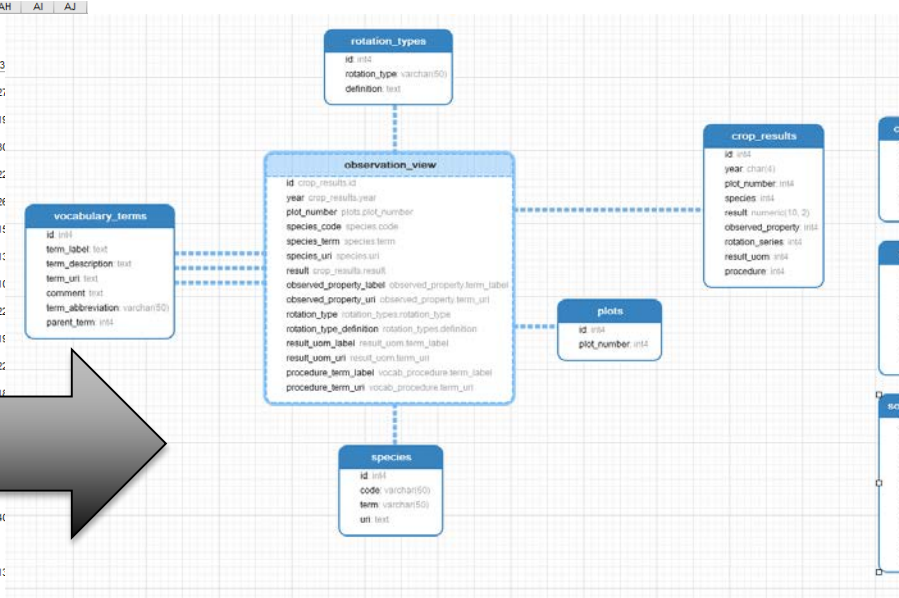


Historical trial research

- Permanent rotation trial since 1920
- Data collected by CSIRO into bespoke spreadsheet
- O&M relational DB developed – FedUni and UofA



Output is JSON using an O&M design pattern.
Use of standardised vocabularies




```
[
  - {
    id: 9999,
    year: "1925",
    plot_number: 1,
    "Grain Yield",
    http://id.agrisemantics.org/gacs/C2241",
    : "wheat followed by fallows",
    ram/hectare (grain yield)",
    purl.obolibrary.org/obo/UO_0000283 ",
    nknown",
    p://www.opengis.net/def/nil/OGC/0/unknown"
  },
  - {
    ram/hectare (grain yield)",
    purl.obolibrary.org/obo/UO_0000283 ",
    nknown",
    p://www.opengis.net/def/nil/OGC/0/unknown"
  },
  - {
```


10003	1927	1	W_F	Wheat	Grain Yield	unknown	2381	kilogram/hectare (grain yield)

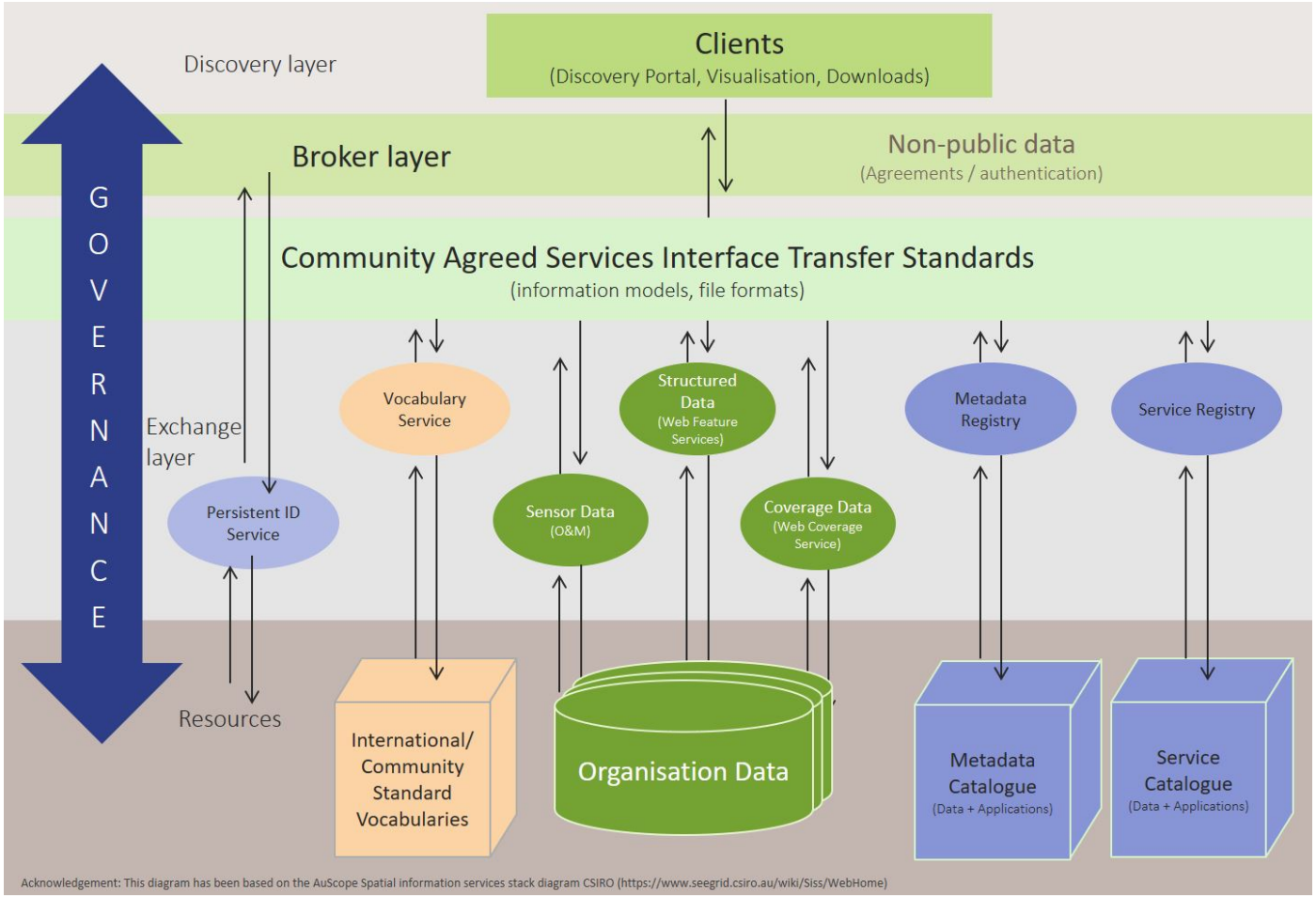
From ‘my data’ to ‘our FAIR data’

- Individual providers’ heterogeneous data
- Data providers assessing their data FAIR-ness and Trusted-ness
- Maturity of provider starting point determined level of standardisation reached
  - Structure - full OGC schema vs OGC pattern vs local structure
  - Content – external vocabulary services vs local terminology
  - Technology – OGC compliant stack vs custom API’s


AgReFed
Agricultural Research
Federation

Extra Slides

AgReFed Computational Viewpoint Stack



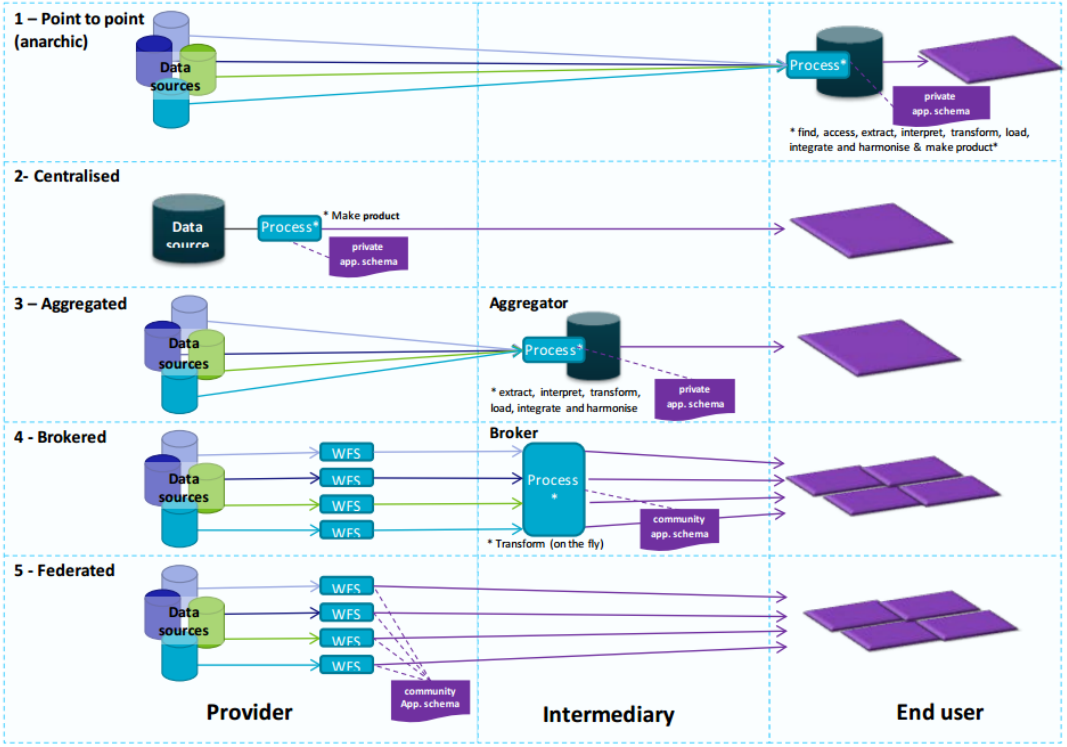
This diagram provides a high level view of the interactions between provider and users via provider and common computational components.

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.

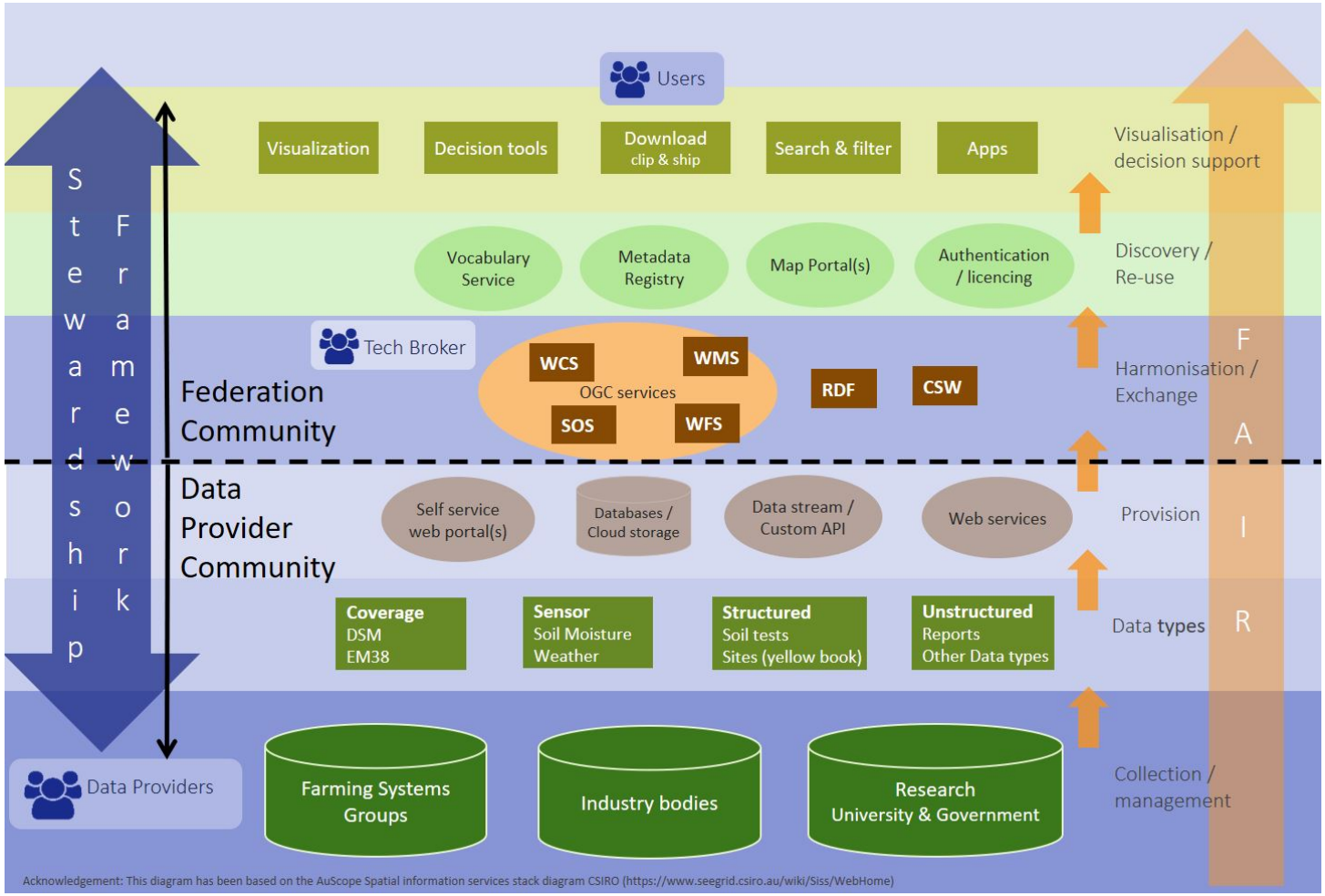
- For federated and brokered data, source data resides with the data provider system. Both solutions encourage currency and validity of data.
- A standards based Service Oriented Architecture (https://en.wikipedia.org/wiki/Service-oriented_architecture) is utilised, including metadata cataloguing and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

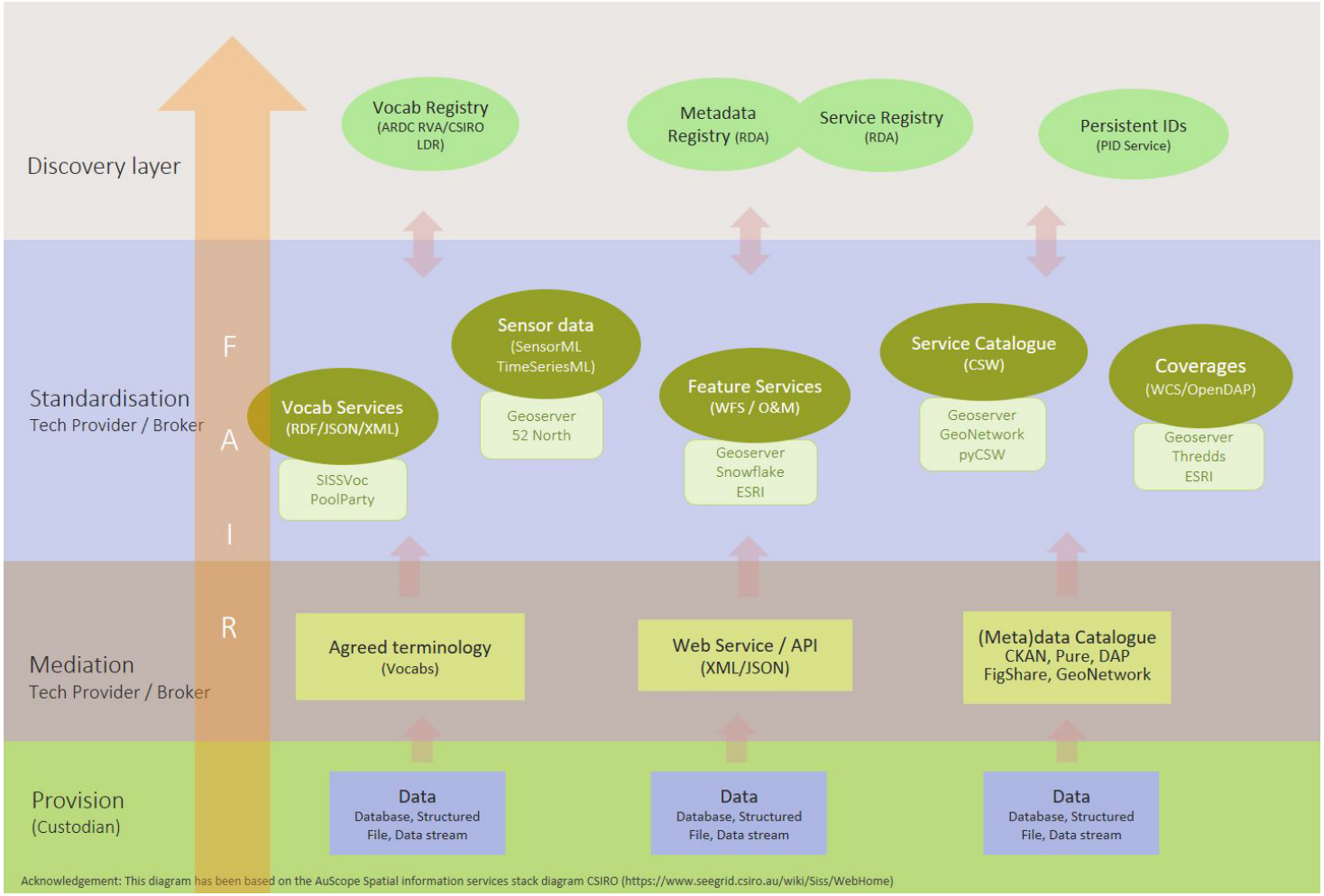
1 https://publications.csiro.au/rpr/pub?pid=csiro:EP155525

AgReFed Engineering viewpoint



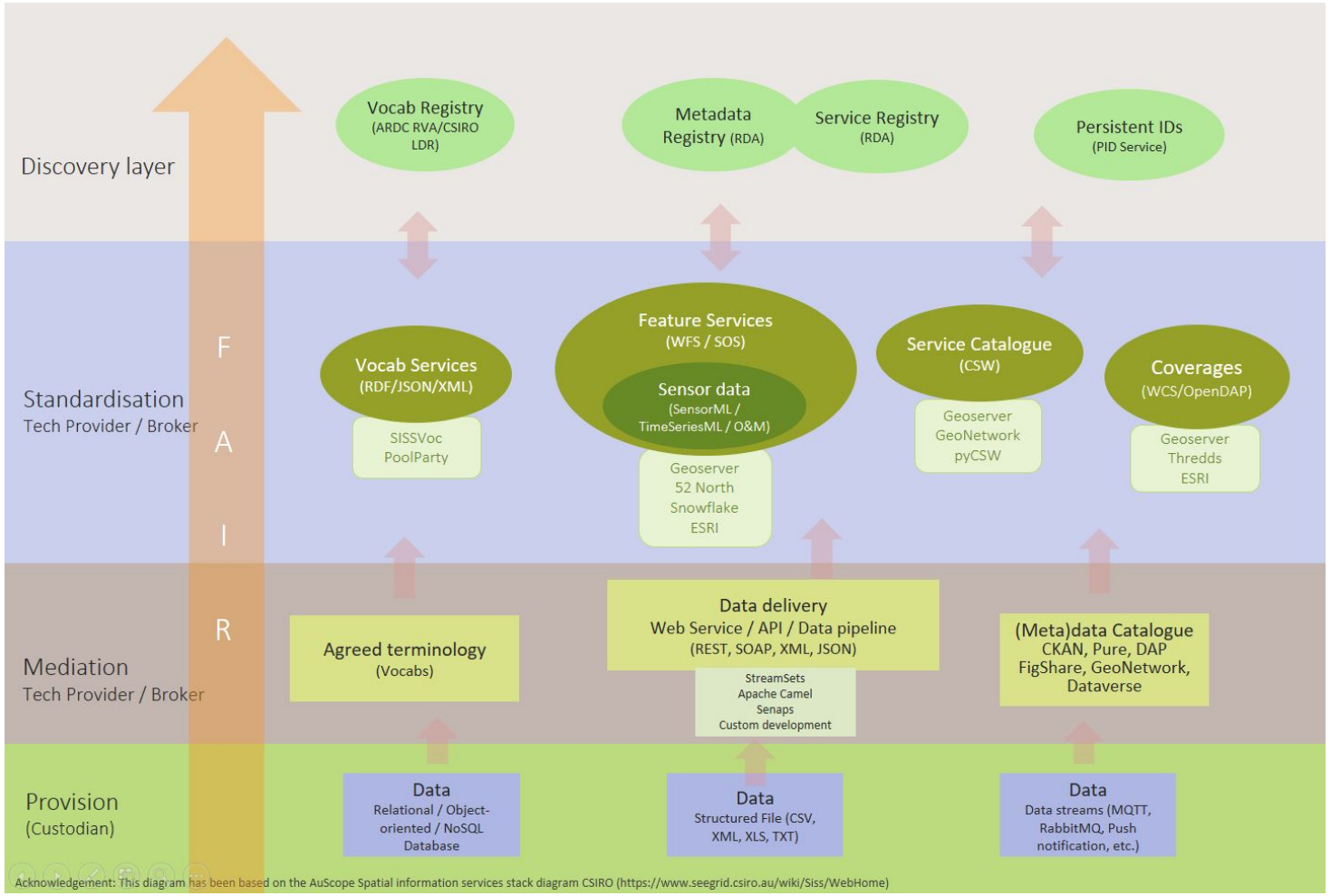
This diagram provides a view of the components, mechanisms and functions required to support distributed interactions among objects in the system, including whose responsibility they are. When this diagram is fleshed out in more detail it will show where components are deployed.

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



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