SmartFarm Data Management.

iRODS User Group 2020

AGRICULTURE VICTORIA

Agriculture Victoria Research

- Science supporting agriculture

Achieving step change improvements in agriculture through innovation for enduring profitability

Enhancing response and management of plant and animal pest and disease outbreaks

Enhancing the underpinning innovation ecosystem

Six science branches

- Genomics and Cellular Sciences
- Microbial Sciences, Pests & Diseases
- Plant Sciences
- Plant Production Sciences
- Animal Production Sciences
- Agriculture Resources Sciences

Innovation clusters with 'hub and spokes' model and 'SmartFarms'

⇒ An outcome-focused innovation agenda with a clear mission:
science and technology for productivity and biosecurity outcomes

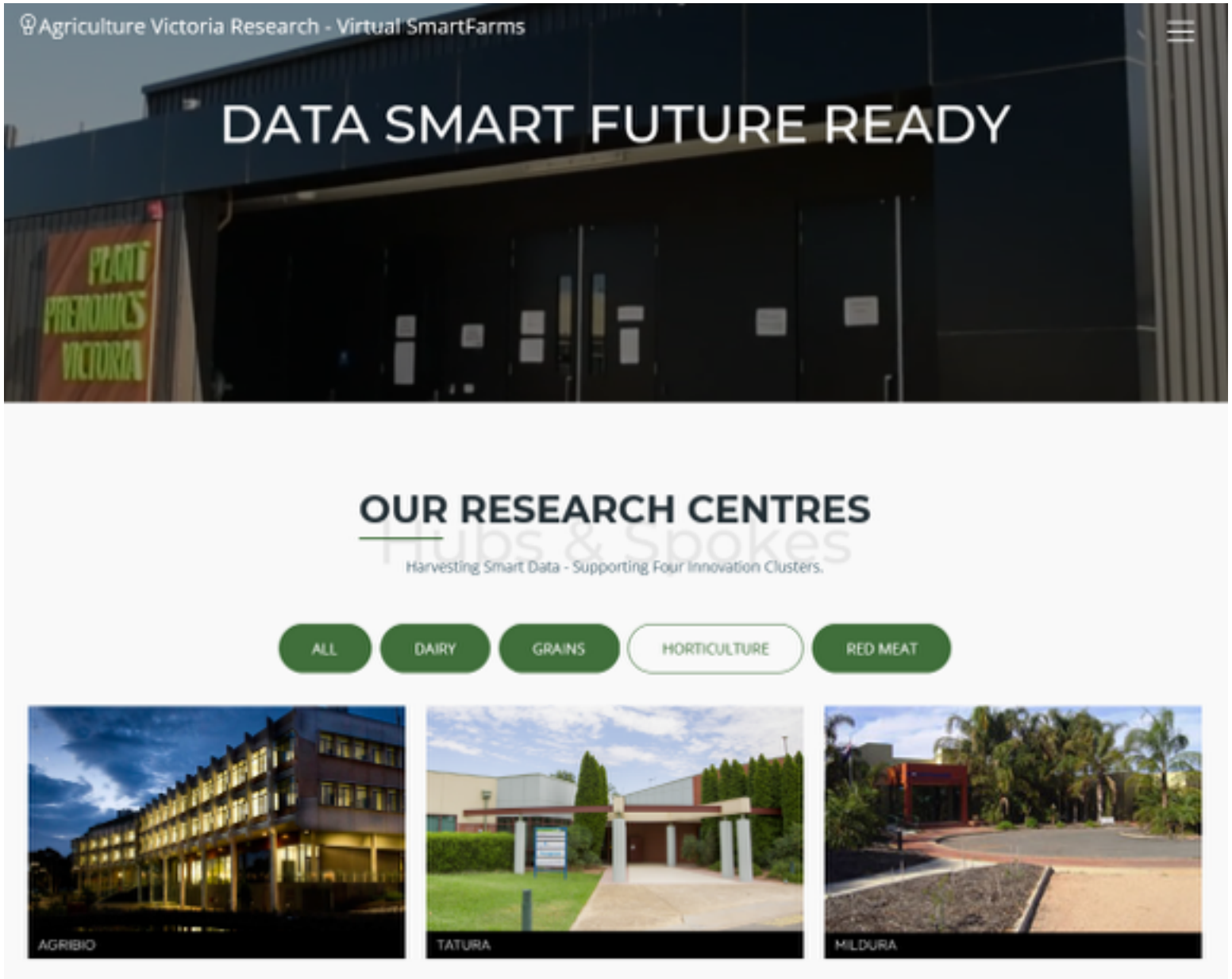
AGRICULTURE VICTORIA

Virtual SmartFarms

The Virtual SmartFarm (VSF) initiative is about connecting AVR's innovation ecosystem using immersive digital technologies that link research SmartFarms with Agribio through an online Hub and Spoke experience.

AGRICULTURE VICTORIA

Virtual SmartFarms Data



AGRICULTURE VICTORIA

Advanced Air-Based Phenomics Platform

High-Throughput Phenomics - Aerial-Based Platforms

Aerial-based Platforms

3DR Solo

DJI M100

DJI M600

DJI S1000+

AGRICULTURE VICTORIA

High-Throughput Phenomics Ground-Based Platforms

High-Throughput Phenomics - PhenoRover

SICK LMS400 LiDAR

Baumer ultrasonic sensor

Campbell Scientific CR3000 datalogger

Navcom RTK GNSS receiver

AGRICULTURE VICTORIA

Ryegrass Reference Population

- Global perennial ryegrass reference population
- Reference population consists of 270,000 plants representing 1,300 experimental varieties
- Weekly measurements on single plants

Challenges of the SmartFarm Data

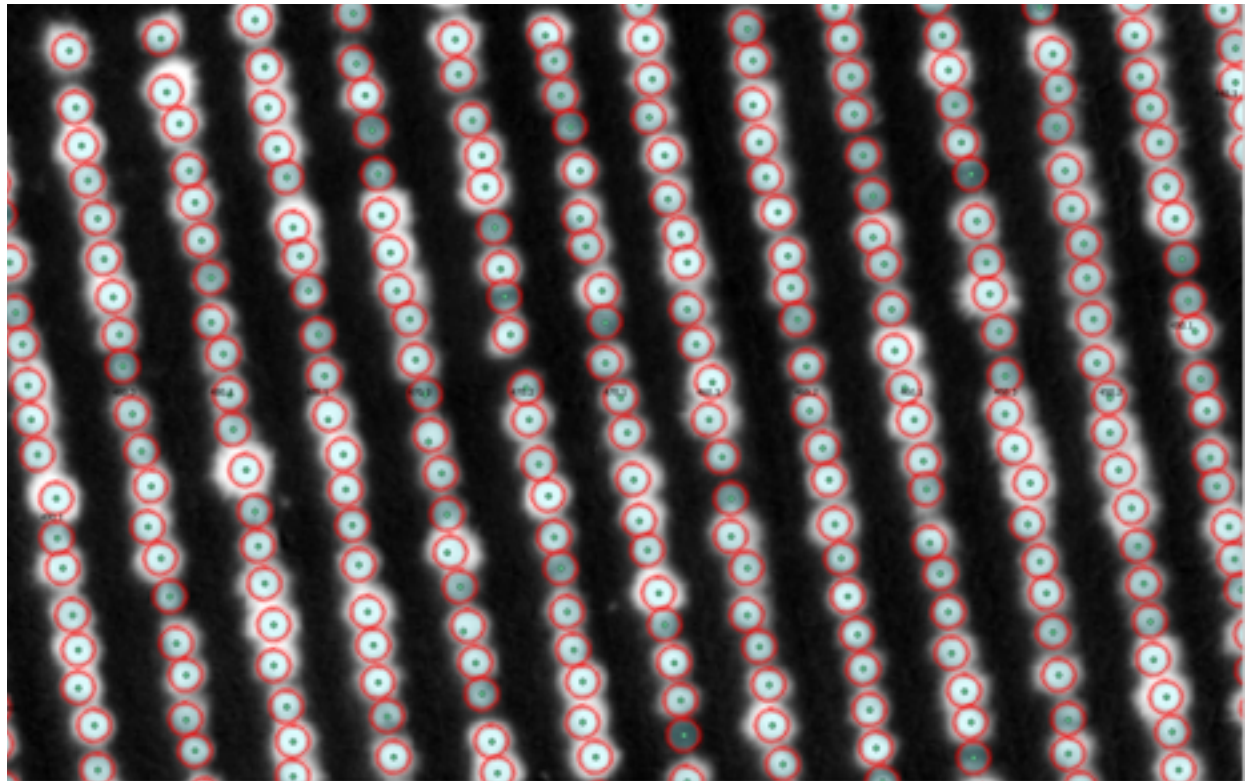
- Geographic Distribution,
- Network Capacity,
- Network Reliability,
- Large Geographic Areas,
- Variety of Sensors to Interface,
- Variety of formats to process
- Variety of required policy.
- Staff capability

Increased reliance on new sensor technology for data collection increasing the challenges of SmartFarm data management.

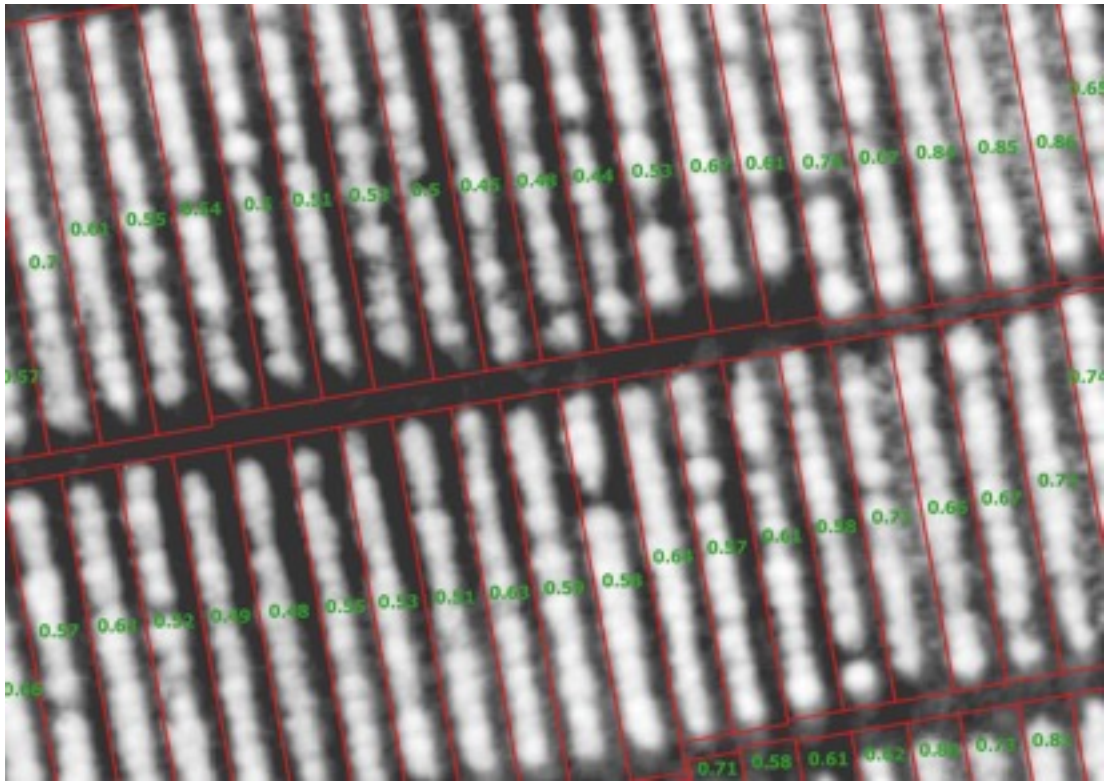
AGRICULTURE VICTORIA

Phenomic Computational Pipeline

Identifying and defining geolocation on single plants



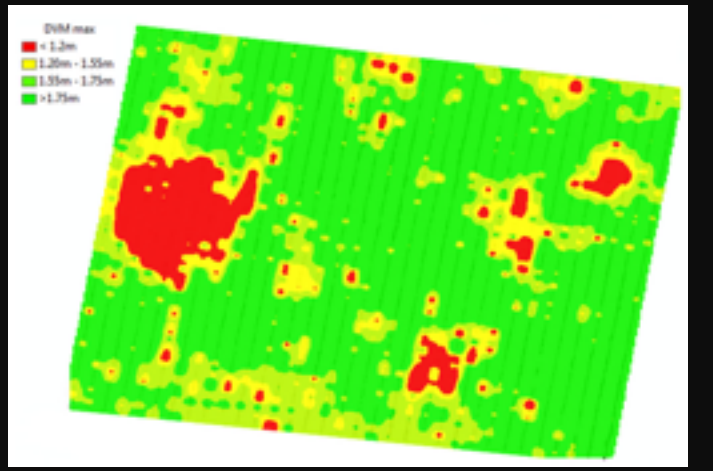
Identifying and defining geolocation on single rows



AGRICULTURE VICTORIA

USE CASE | UAV Data

- **PROBLEM** | Use of sophisticated and data intensive technologies is increasing the complexity of collecting, description, assembly, transport and analysis of data. This use case establishes a forward looking pathway to metadata management, data discovery and use across AVR sites.
- **SOLUTION** | Requires metadata discovery and workflow automation from ingested UAV data, new big data collection and transfer methods that utilise edge computing, coded data policies and simple storage service for access and use.
- **INFRASTRUCTURE** | iRODS (metadata & workflow) and S3 Data Lake (storage and access)
- **CAPABILITIES** | Automated ingest and metadata discovery workflow, metadata policies, algorithms and analytics.



AGRICULTURE VICTORIA

Making data discoverable moves beyond establishing folder schemas to the development of agreed metadata

The appropriate choice of metadata tags, as well as of queries that can be implemented is aided greatly when this body of ingested data needs to be made discoverable.

Examples of the types of metadata tags that might be added to the data:

a. Reflectance data, possibly other multi- or hyper-spectral data from sensors
b. UAV flight parameters, e.g. orientation and GPS position
c. Timestamps

Metadata is critically important in all stages of processing and data discovery, but near the front-end it is particularly good for uses such as logically tying together related datasets, or associating raw data with measured (quantifiable) details of the collection process (precise time and geographic location probably being the most important).

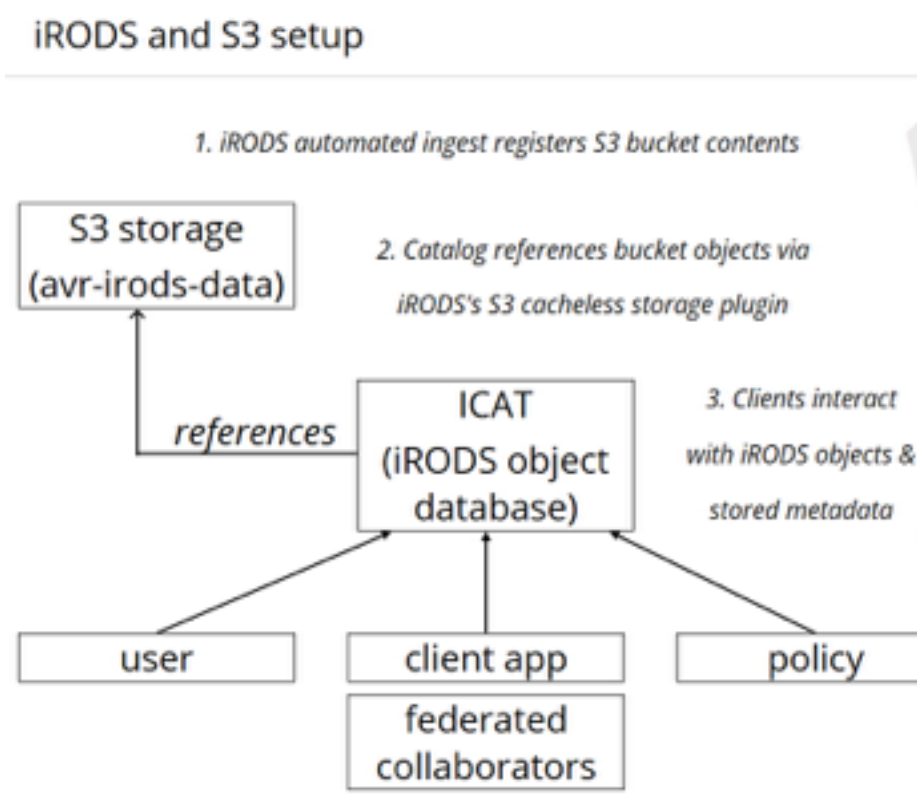
Initial Goals

1. Upload existing AVR data as example content into S3 bucket **avr-irods-data**
2. Get S3 files / folders registered to iRODS catalogue
3. Extract salient metadata – e.g. EXIF tags in TIF files
4. Tag Data Objects and Collections to make them Actionable and Discoverable

AGRICULTURE VICTORIA

The Content

- **Ingest policy registers object in place then extracts metadata**
- **Apply metadata to the object in the catalogue**
  - Metadata headers available in the files
  - Contextual metadata : LZ directory, instrument, etc
- **Demonstrate**
  - Ingest
  - Discovery
  - Data egress
  - Graphical presentation
  - File system presentation : WebDAV & emerging new front ends.



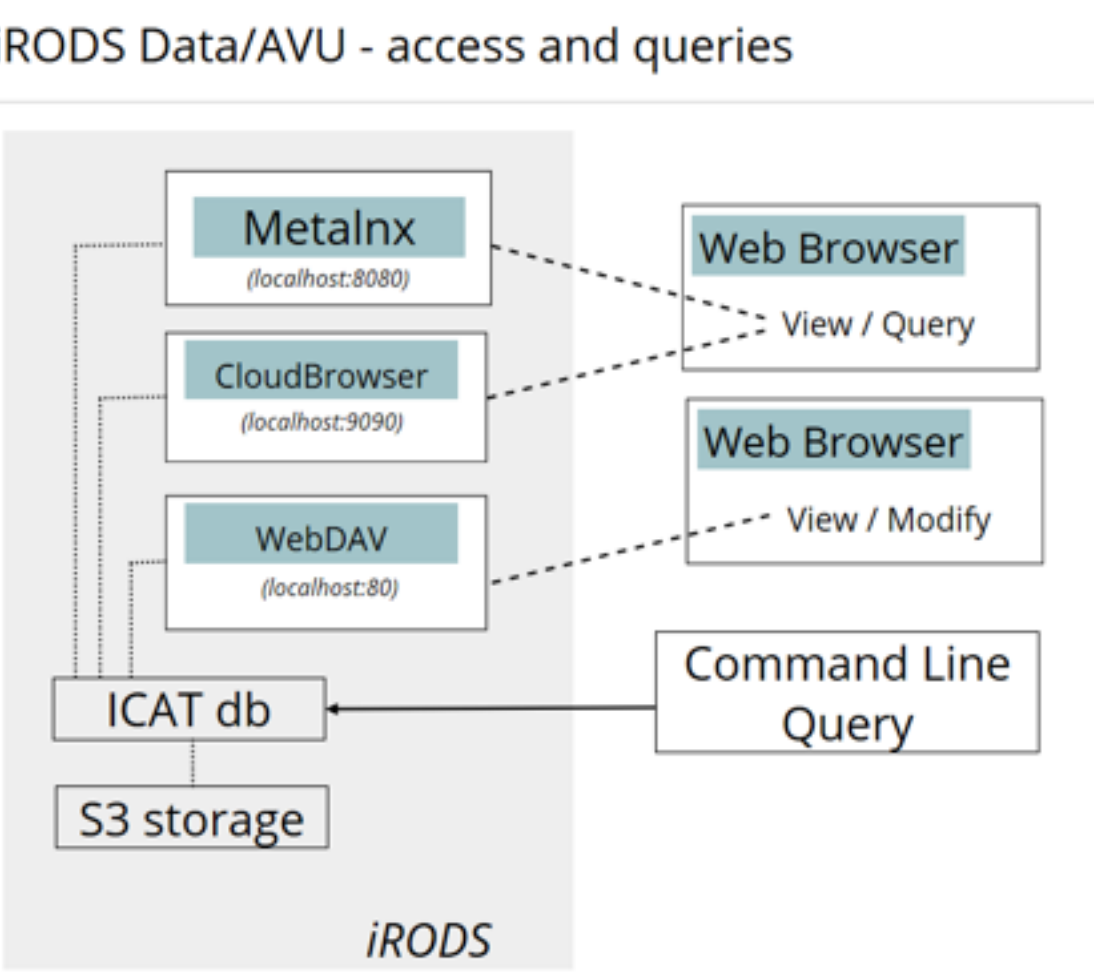
AGRICULTURE VICTORIA

Automated Ingest

S3 buckets scanned
- avr_irods_data
- possibly many others

Any data that is discovered during a scan
- Automatically registered to a storage resource
- Metadata extracted and applied to the object in the catalogue
- Event possible generated for audit trail
- Create opportunities for richer data discovery

User can view and access data and



AGRICULTURE VICTORIA

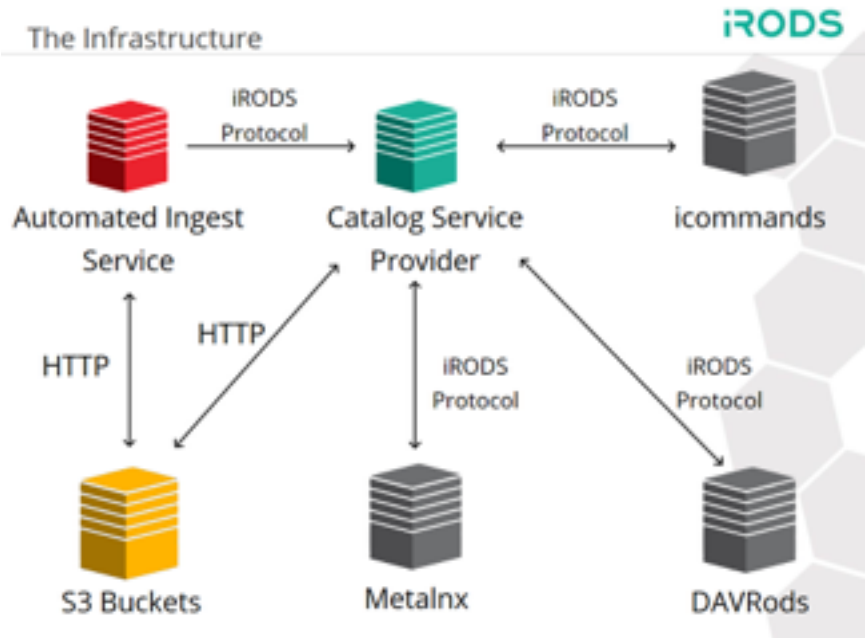
Data Discovery with MetaInx

Automated ingest has provided metadata for data discovery

The metadata can be directly inspected in MetaInx

The query builder can be used to identify data sets of interest via Attribute, Value, Unit matches

Queries to the system metadata may also be performed, searching on values such as file name, collection path, user, etc.



AGRICULTURE VICTORIA

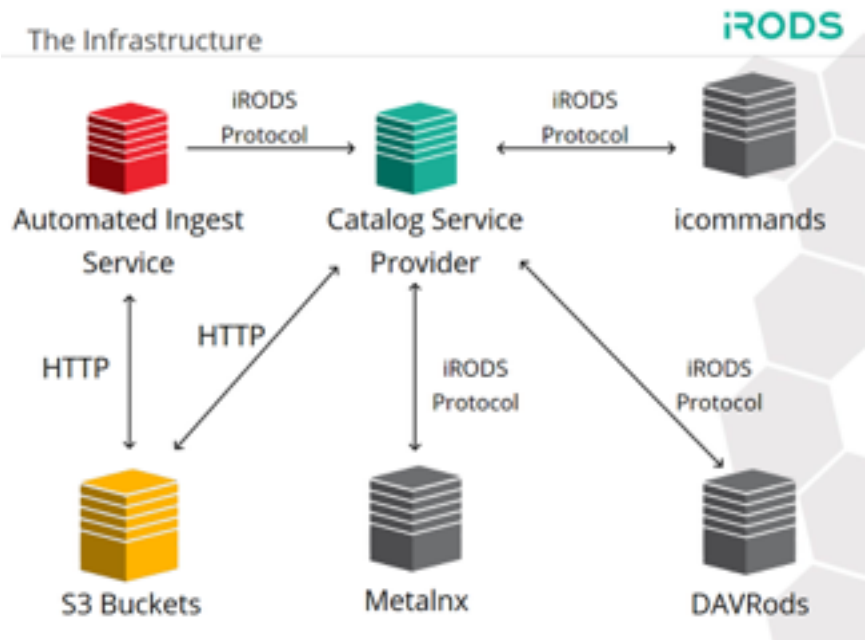
File System Presentations: DAVRods

DAVRods provides both a simple web based interface as well as the ability to mount a folder on the desktop

DAVRods is an Apache Module implemented in C using the native iRODS POSIX API

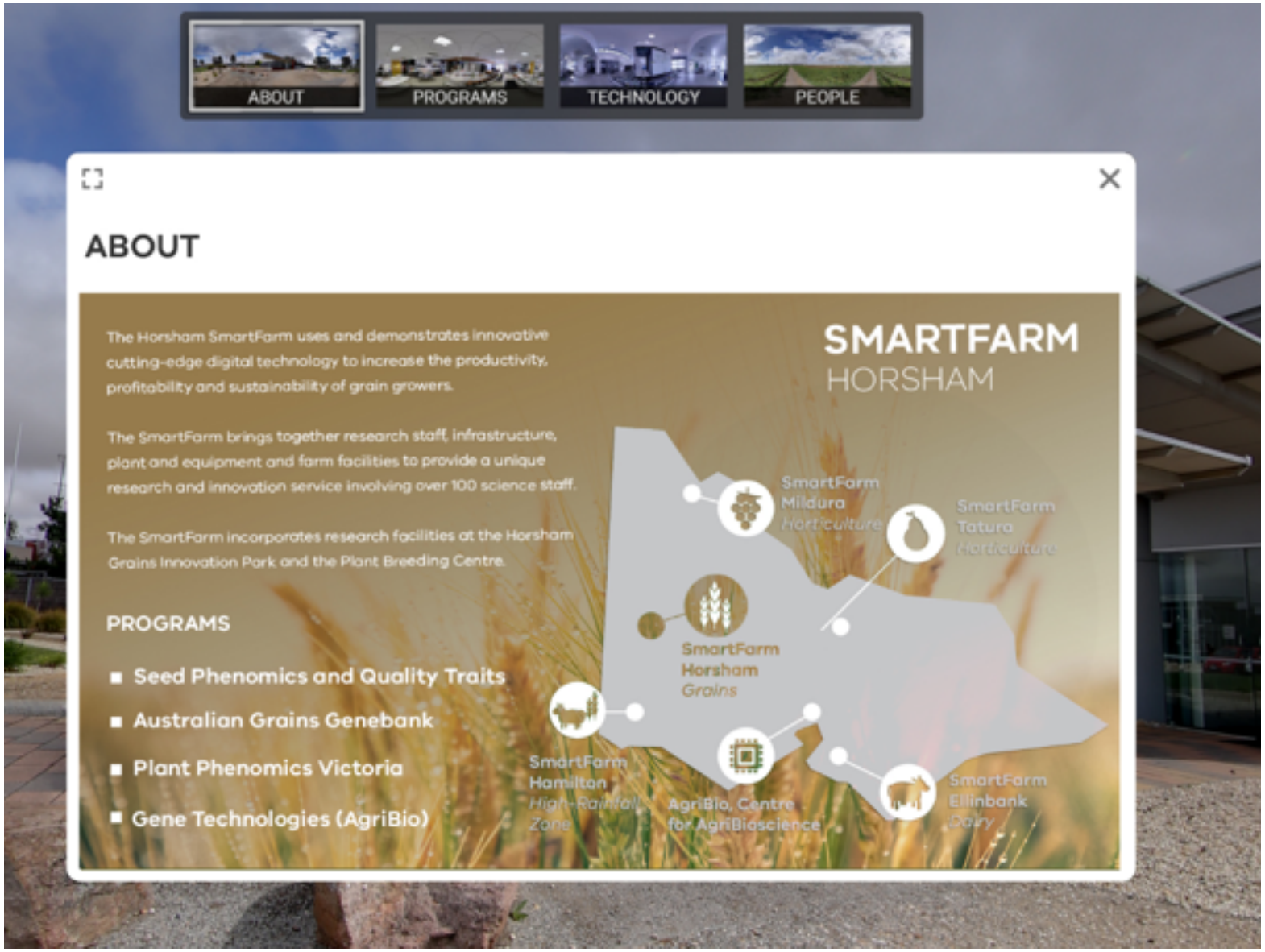
DAVRods can be used to edit data in-place, or to copy data to/ from a users collections.

USE CASE requirements for increased UI and UX specifications



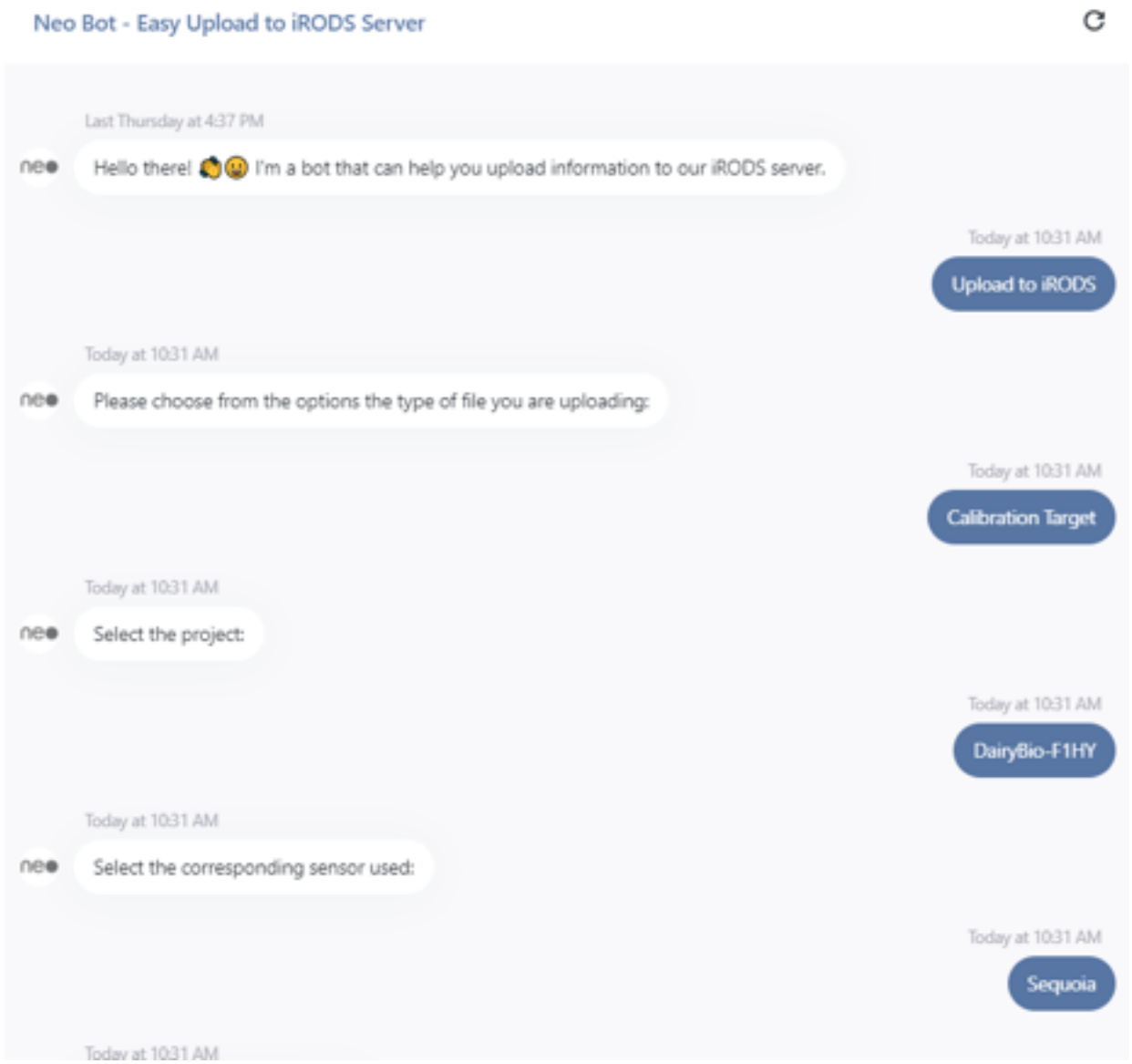
AGRICULTURE VICTORIA

Virtual SmartFarm Data ecosystem - testing new function



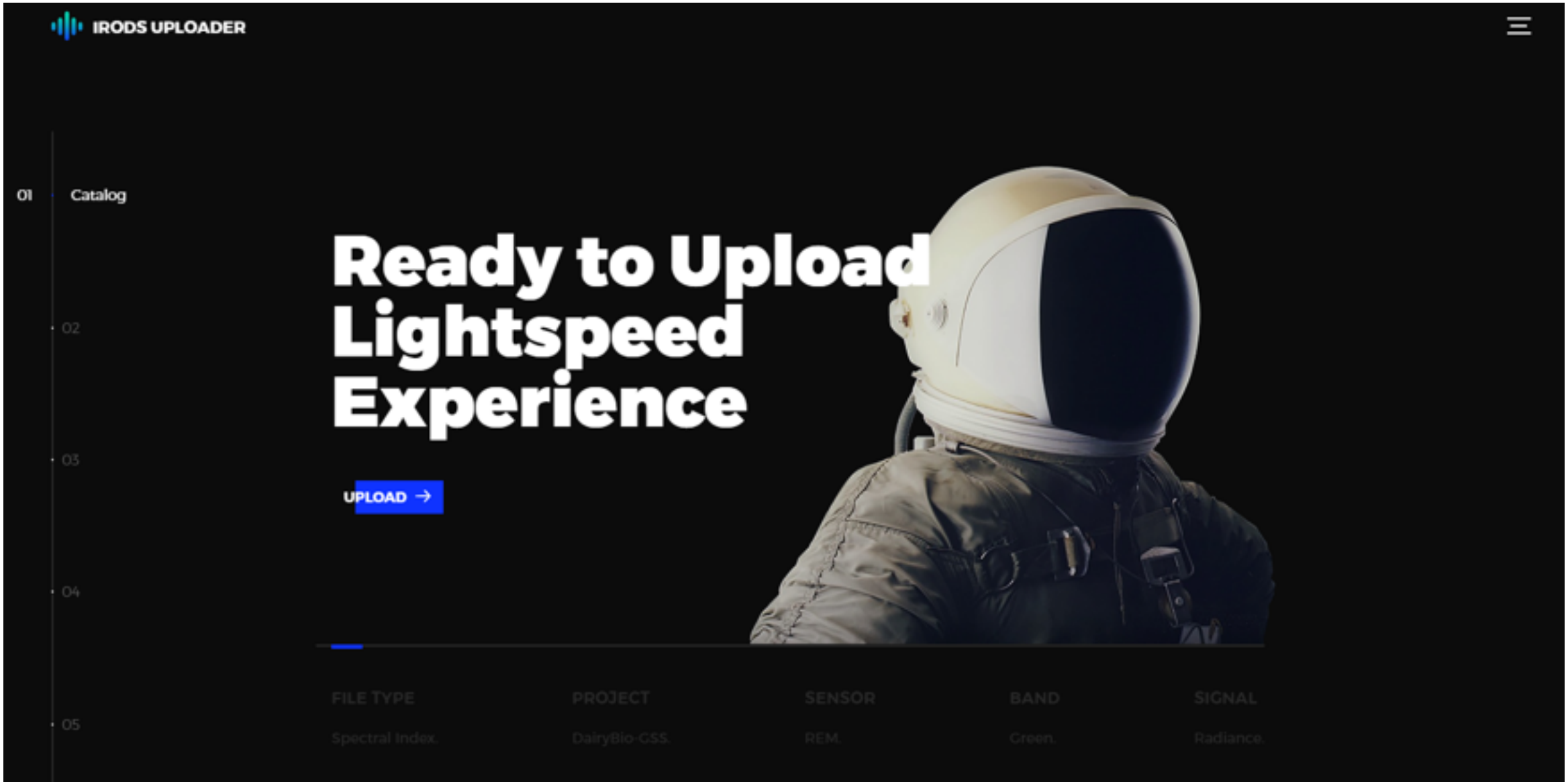
AGRICULTURE VICTORIA

Virtual SmartFarm Data ecosystem - example



AGRICULTURE VICTORIA

Virtual SmartFarm Data ecosystem example

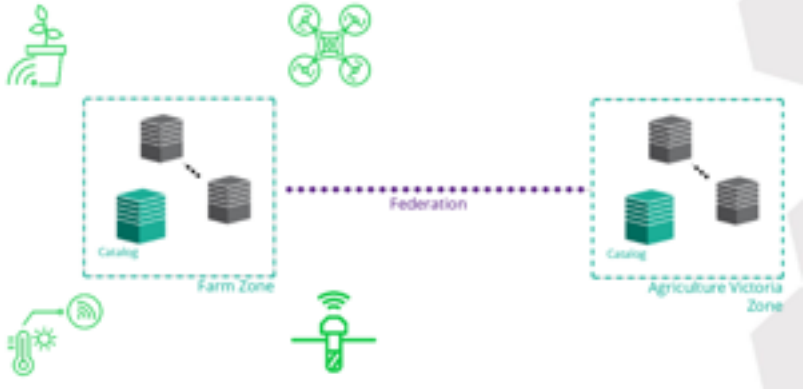


Emerging SmartFarm Data Infrastructure

Each SmartFarm may host their own application (iRODS) to manage metadata description and catalogue for each UAV trial.

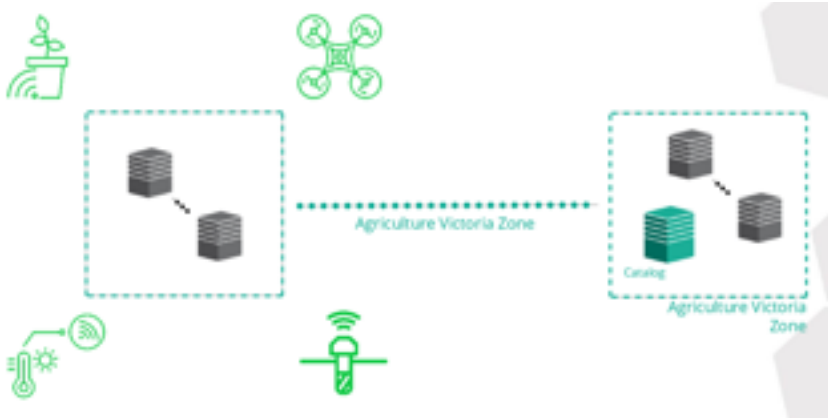
Data is gathered from the UAV over the protocol of choice.

Data is periodically synchronised to Agriculture Victoria Research servers (S3 / Hybrid)



SmartFarm hosts Agriculture Victoria Research servers (S3 / Hybrid)

Data is periodically **replicated** to Agriculture Victoria Research Servers (BASC)

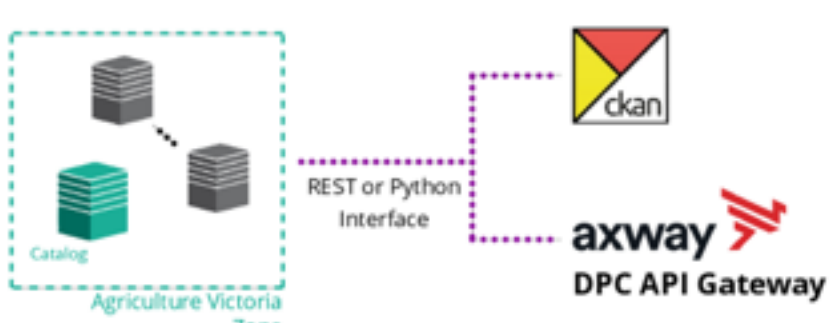


Once data is at rest in the Agriculture Victoria Research namespace i.e. Horsham_UAV_AVR_Plot1

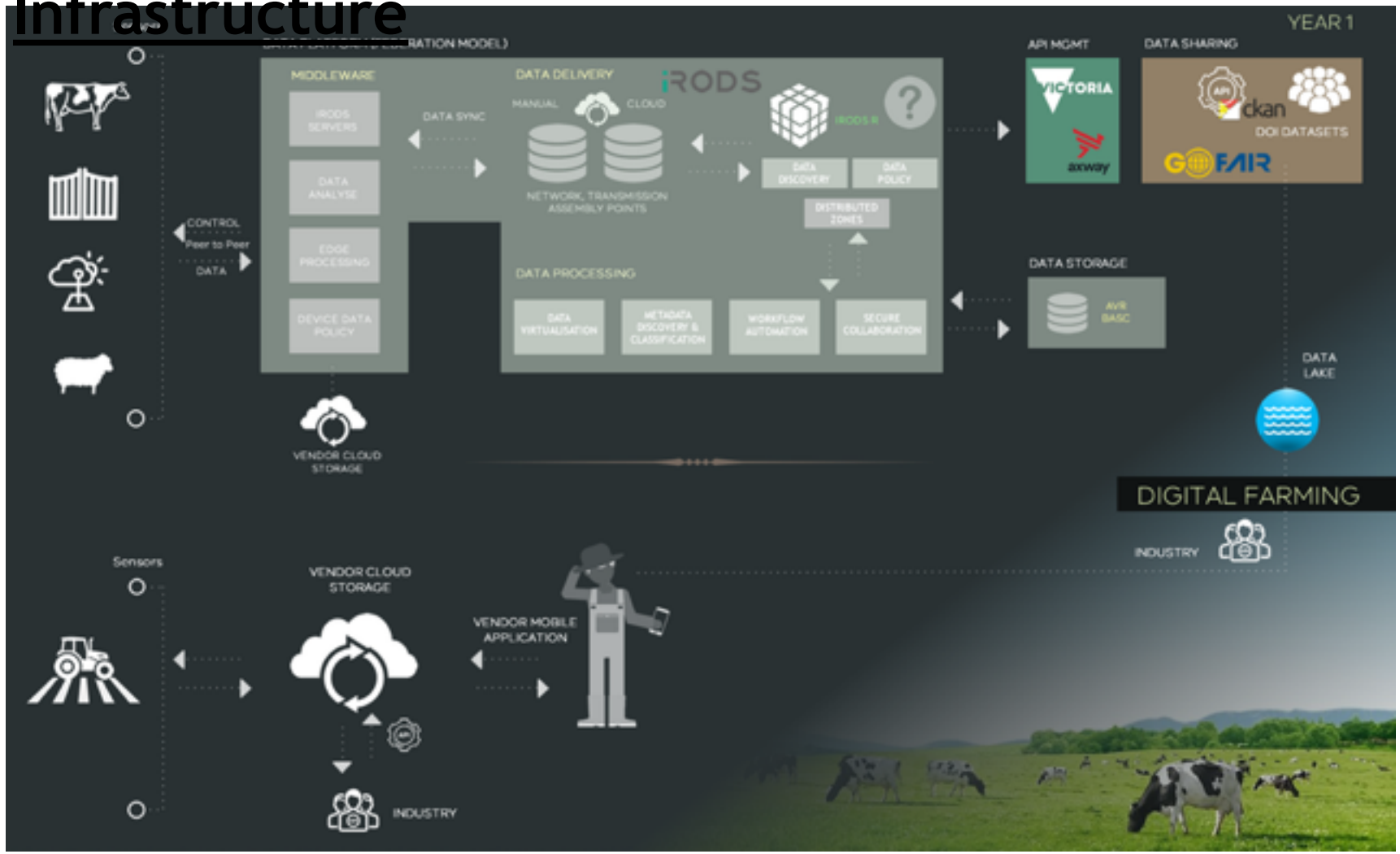
Data may be replicated to HPC storage for analytics.

Data may be published to CKAN or made accessible via the API gateway

Data may be shared over an IRODS interface : WebDAV, MetaInx, NFS, Command Line.



SmartFarm Data Infrastructure



iRODS is facilitating the data transfer and movement of data from remote geographic SmartFarms.

Deploying iRODS at the edge on these SmartFarm minimises the impact of network traffic and development of data policies.

By virtualising this data and correctly cataloguing this into specific iRODS zones, we effectively maintain our data is “optimised” to our SmartFarm data architecture.

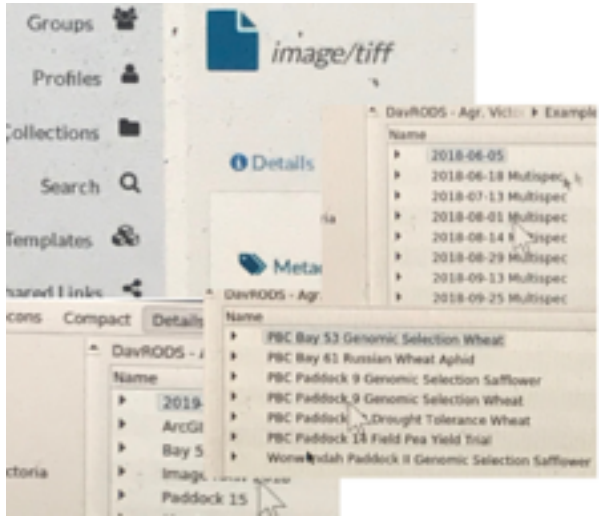
This is supports our API strategy and makes it easier for our researchers data to be consumed in formats that their clients expect.

Completed Use Case. Next iteration endorsed with iRODS

- Testing data ingest to S3 bucket and open source metadata management application (iRODS).
- A new capability in data discovery and workflow automation – new AI and enhanced UX
- Enable data classification and reporting to support rapid assessment of data assets and use.
- Fast track data processing and transfer to defined repositories for management and use.
- Better manage data sovereignty, preservation and reproducibility for researchers.

```
# POLICY CONFIGURATION VARIABLES
SCANNED_RESOURCE = 'example_scanned_resc'
DESTINATION_RESOURCE_ROOT = 'example_dest_resc_root'
LIST_OF_DESTINATION_RESOURCE_LEAVES = ['a','b','c']
```

iRODS



AGRICULTURE VICTORIA

The Integrated Rule-Oriented Data System (iRODS) is open source data management software used by research organizations and government agencies worldwide