

SmartFarm Data Management.

iRODS User Group 2020

Agriculture Victoria Research

- Science supporting agriculture

Achieving step change improvements in agriculture through innovation for enduring profitability

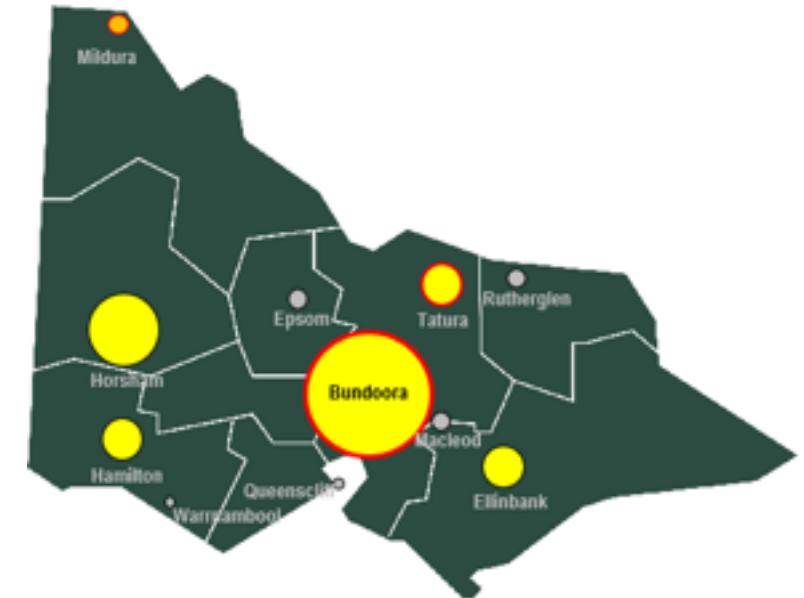
Enhancing response and management of plant and animal pest and disease outbreaks

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

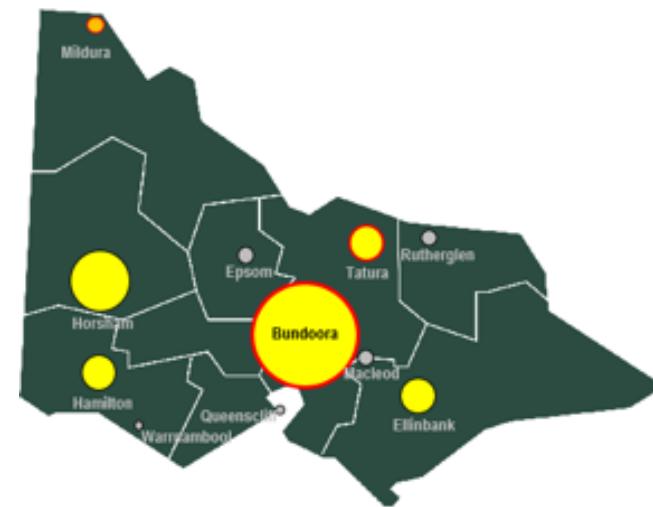
Innovation clusters with ‘hub and spokes’ model and ‘SmartFarms’



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

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.



Virtual SmartFarms Data



OUR RESEARCH CENTRES

Hubs & Spokes

Harvesting Smart Data - Supporting Four Innovation Clusters.

ALL

DAIRY

GRAINS

HORTICULTURE

RED MEAT



AGRIBIO



TATURA



MILDURA



Advanced Air-Based Phenomics Platform



High-Throughput Phenomics - Aerial-Based Platforms



Aerial-based Platforms



3DR Solo



DJI M100



DJI M600



DJI S1000+

High-Throughput Phenomics Ground-Based Platforms



High-Throughput Phenomics - PhenoRover



SICK LMS400
LiDAR



Baumer
ultrasonic sensor



Campbell Scientific
CR3000 datalogger



Navcom RTK GNSS receiver



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

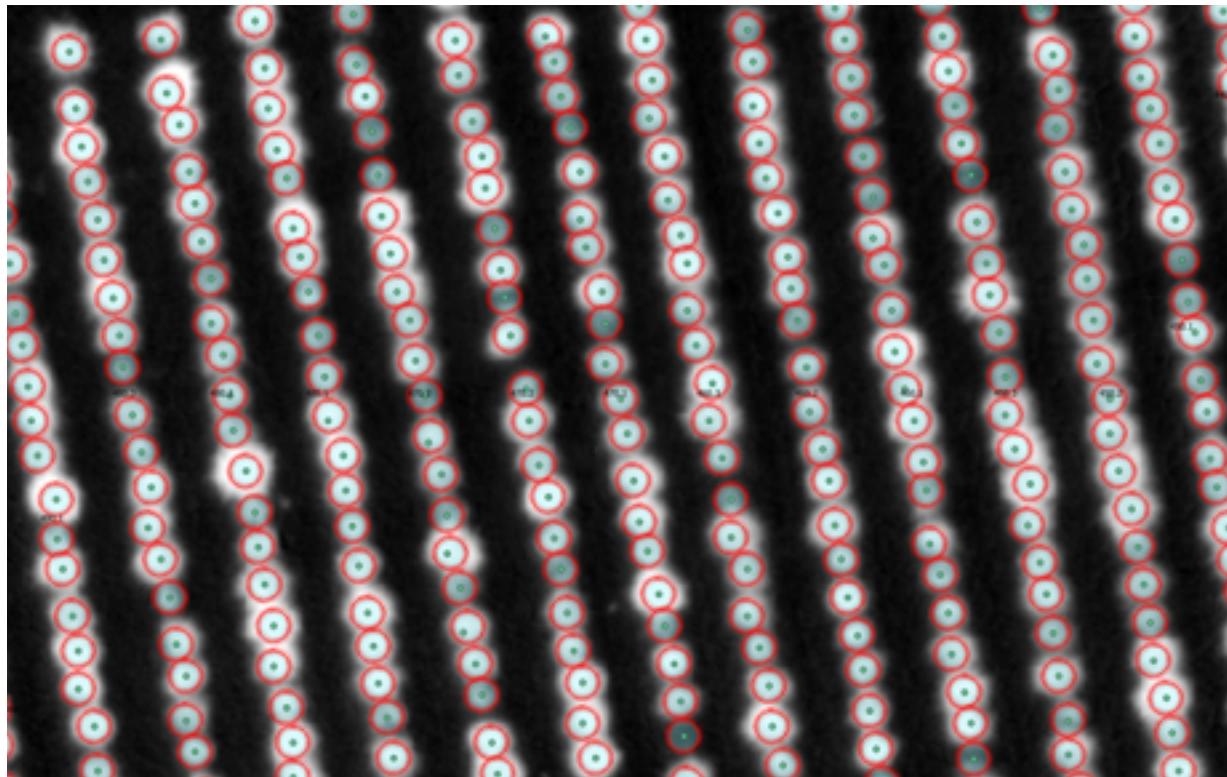
- Geographic Distribution,
- Network Capacity,
- Network Reliability,
- Large Geographic Areas,
- Variety of Sensors to Interface,
- Variety of formats to process
- Variety of required policy.
- Staff capability

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

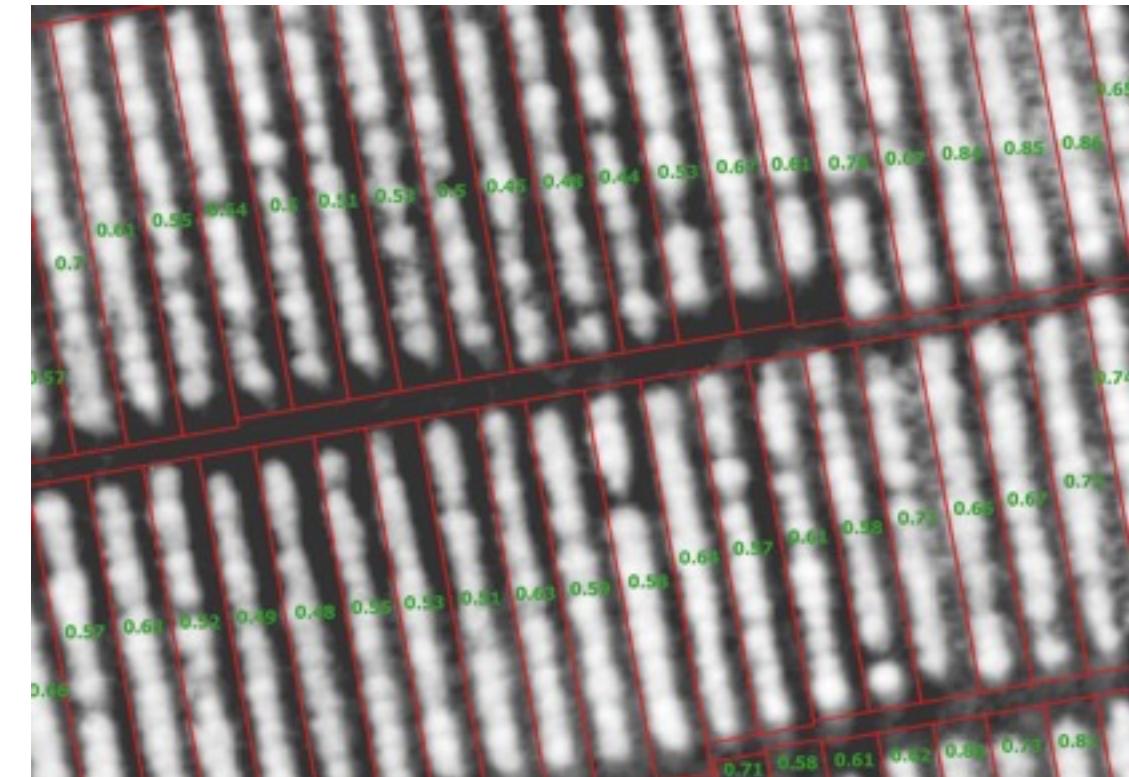


Phenomic Computational Pipeline

Identifying and defining geolocation on single plants

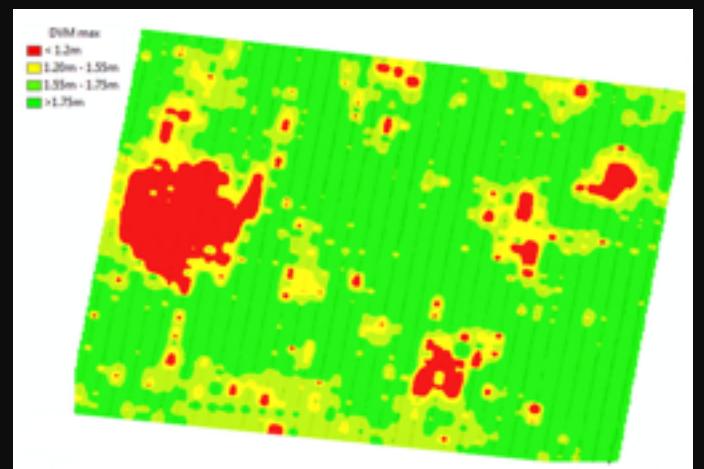
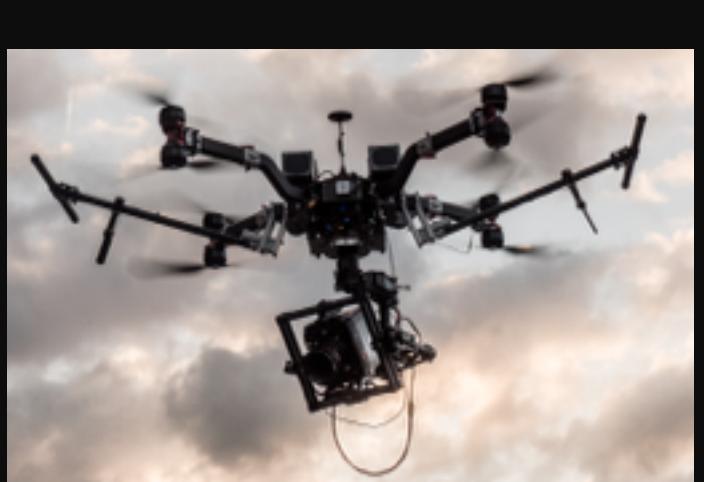


Identifying and defining geolocation on single rows



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.



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

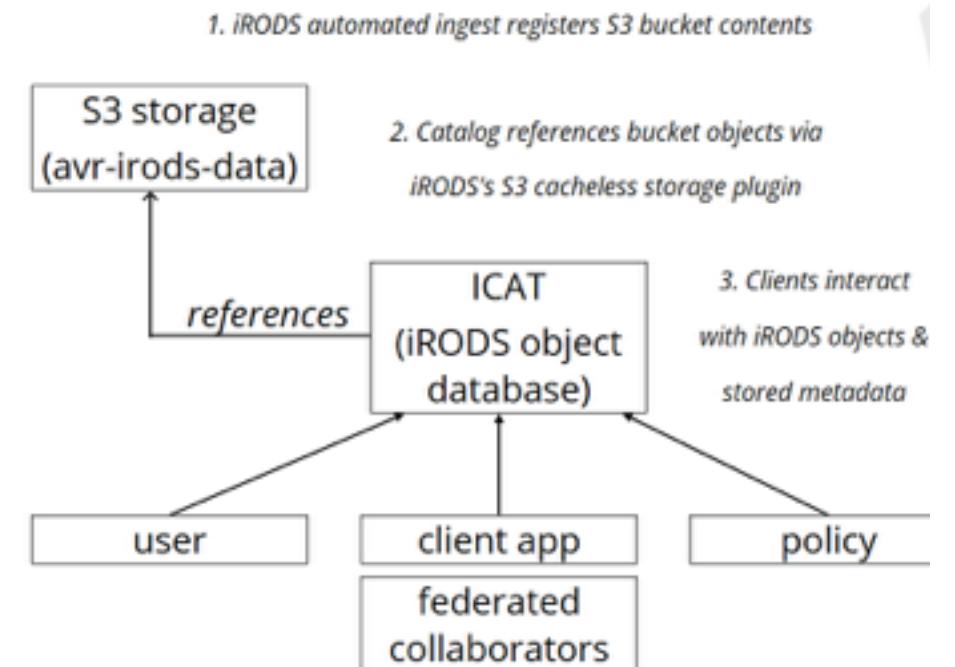
1. Upload existing AVR data as example content into S3 bucket **avr-irods-data**
2. Get S3 files / folders registered to iRODS catalogue
3. Extract salient metadata - e.g. EXIF tags in TIF files
4. Tag Data Objects and Collections to make them Actionable and Discoverable



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.

iRODS and S3 setup



Automated Ingest

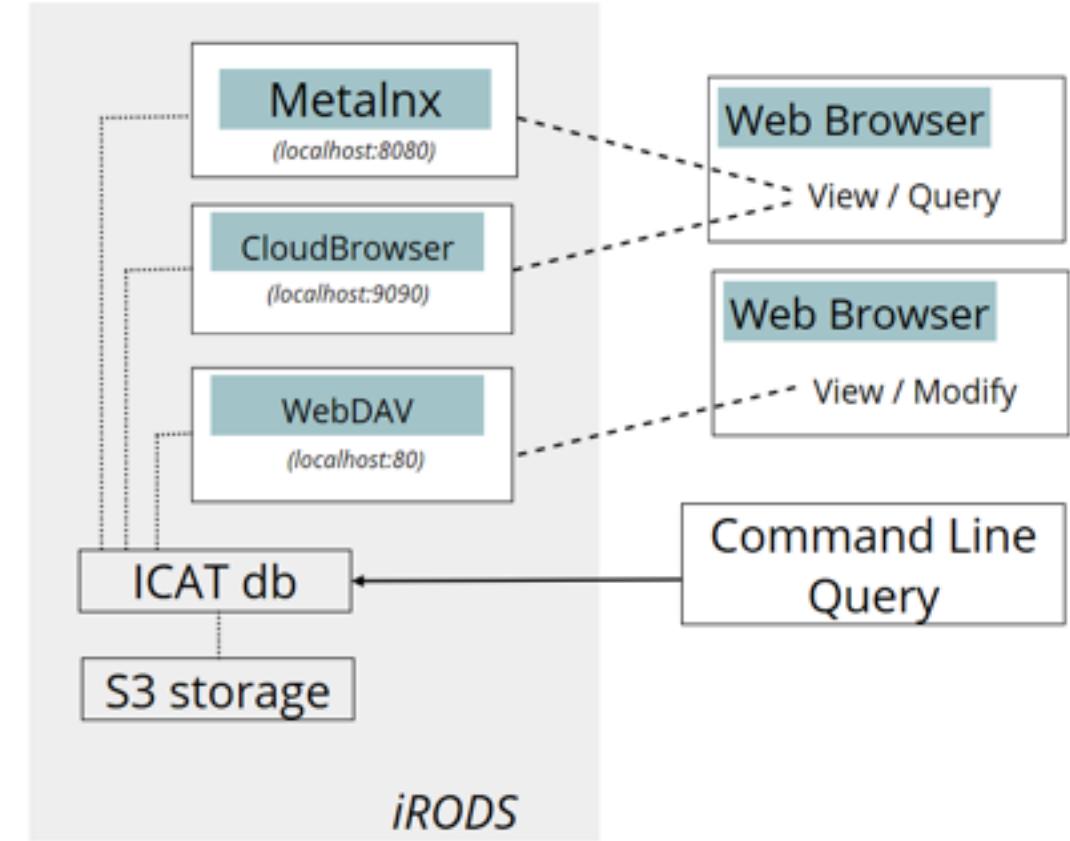
S3 buckets scanned

- 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

iRODS Data/AVU - access and queries



User can view and access data and

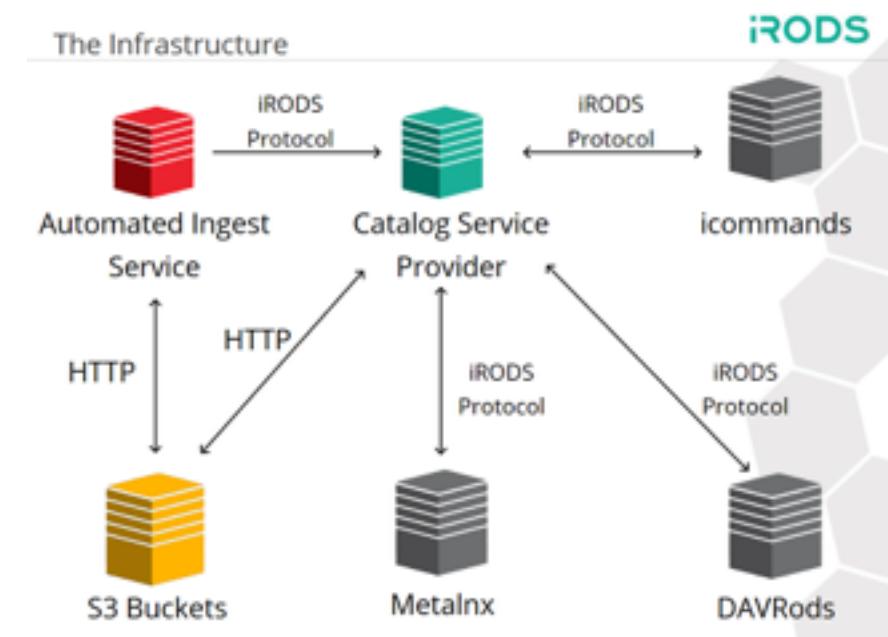
Data Discovery with Metalnx

Automated ingest has provided metadata for data discovery

The metadata can be directly inspected in Metalnx

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.



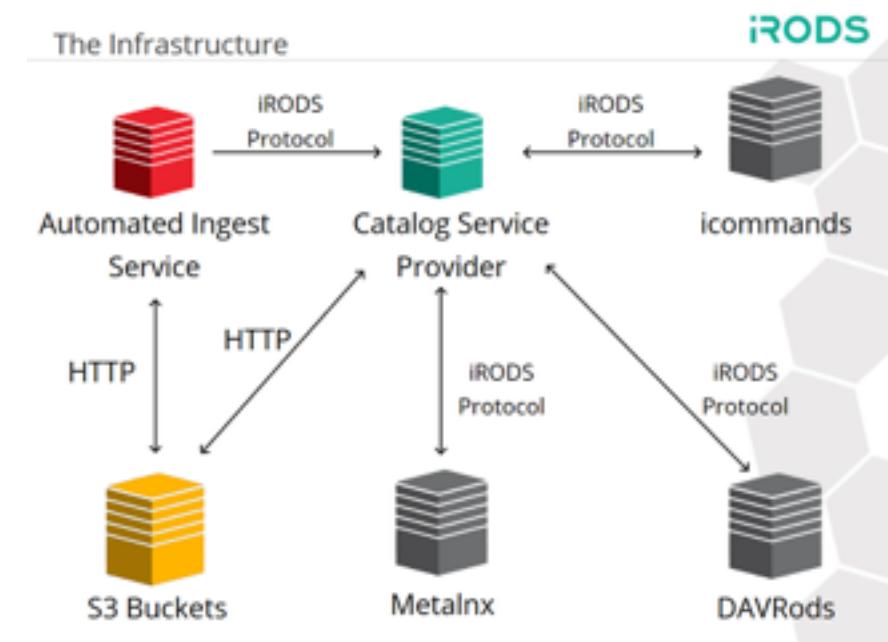
File System Presentations: DAVRods

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



Virtual SmartFarm Data ecosystem - testing new function

The screenshot shows a mobile-optimized website for the Horsham SmartFarm. At the top, there is a navigation bar with four tabs: 'ABOUT', 'PROGRAMS', 'TECHNOLOGY', and 'PEOPLE'. Below the navigation bar is a large, semi-transparent overlay window. The overlay has a close button ('X') in the top right corner and a title 'ABOUT' in bold capital letters. Inside the overlay, there is descriptive text about the Horsham SmartFarm, followed by a map titled 'SMARTFARM HORSHAM' showing the locations of various research facilities across Victoria. The facilities include SmartFarm Mildura Horticulture, SmartFarm Toturo Horticulture, SmartFarm Horsham Grains, SmartFarm Hamilton High-Rainfall Zone, AgriBio, Centre for AgriBioscience, and SmartFarm Ellinbank Dairy. Each facility is represented by a small icon and a label. The background of the overlay features a photograph of a modern agricultural building. The overall design is clean and professional, using a light color palette and clear typography.

ABOUT

The Horsham SmartFarm uses and demonstrates innovative cutting-edge digital technology to increase the productivity, profitability and sustainability of grain growers.

The SmartFarm brings together research staff, infrastructure, plant and equipment and farm facilities to provide a unique research and innovation service involving over 100 science staff.

The SmartFarm incorporates research facilities at the Horsham Grains Innovation Park and the Plant Breeding Centre.

PROGRAMS

- Seed Phenomics and Quality Traits
- Australian Grains Genebank
- Plant Phenomics Victoria
- Gene Technologies (AgriBio)

SMARTFARM HORSHAM

The map illustrates the spatial distribution of SmartFarm facilities across Victoria. It shows a central cluster around Horsham, with branches extending to the west (Mildura), east (Toturo), north (Hamilton), and south (Ellinbank). Each location is marked with a circular icon representing a specific sector or facility.

SmartFarm Mildura Horticulture

SmartFarm Toturo Horticulture

SmartFarm Horsham Grains

SmartFarm Hamilton High-Rainfall Zone

AgriBio, Centre for AgriBioscience

SmartFarm Ellinbank Dairy

AGRICULTURE VICTORIA

Virtual SmartFarm Data ecosystem - example

Neo Bot - Easy Upload to iRODS Server



Last Thursday at 4:37 PM

neo Hello there! 🐄😊 I'm a bot that can help you upload information to our iRODS server.

Today at 10:31 AM

Upload to iRODS

Today at 10:31 AM

neo Please choose from the options the type of file you are uploading:

Today at 10:31 AM

Calibration Target

Today at 10:31 AM

neo Select the project:

Today at 10:31 AM

DairyBio-F1HY

Today at 10:31 AM

neo Select the corresponding sensor used:

Today at 10:31 AM

Sequoia

Today at 10:31 AM

Virtual SmartFarm Data ecosystem example

IRODS UPLOADER

≡

01 Catalog

02

03

04

05

Ready to Upload
Lightspeed
Experience

UPLOAD →



FILE TYPE

05 Spectral Index.

PROJECT

DairyBio-GSS.

SENSOR

REM.

BAND

Green.

SIGNAL

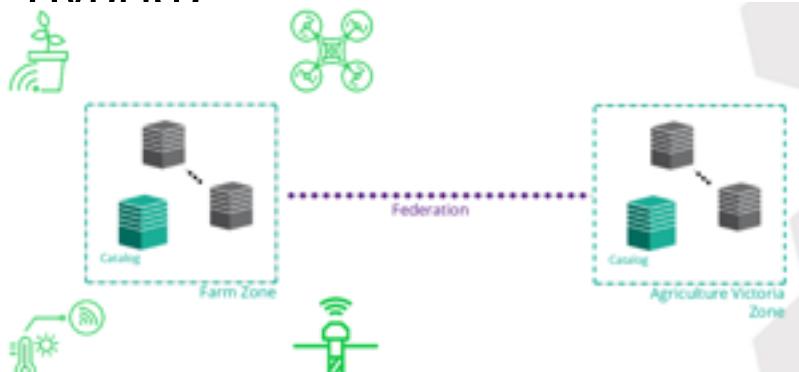
Radiance.

Emerging SmartFarm Data Infrastructure

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

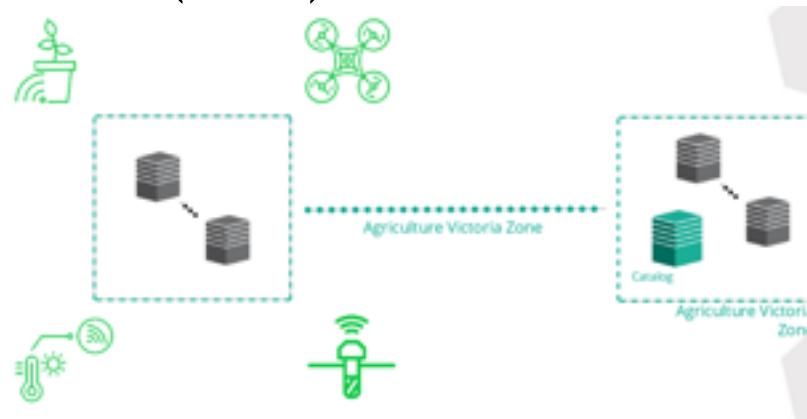
Data is gathered from the UAV over the protocol of choice.

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)

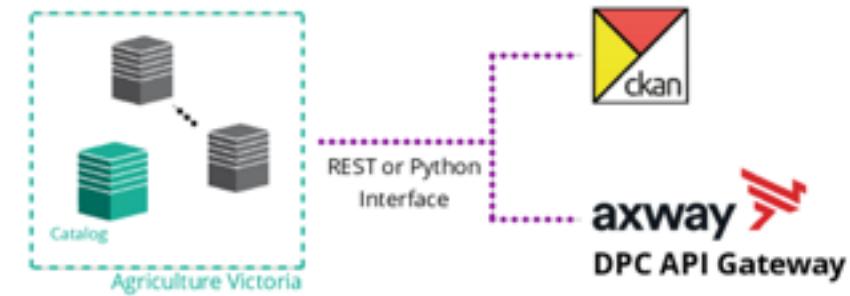


Once data is at rest in the Agriculture Victoria Research namespace i.e.
Horsham_UAV_AVR_Plot1

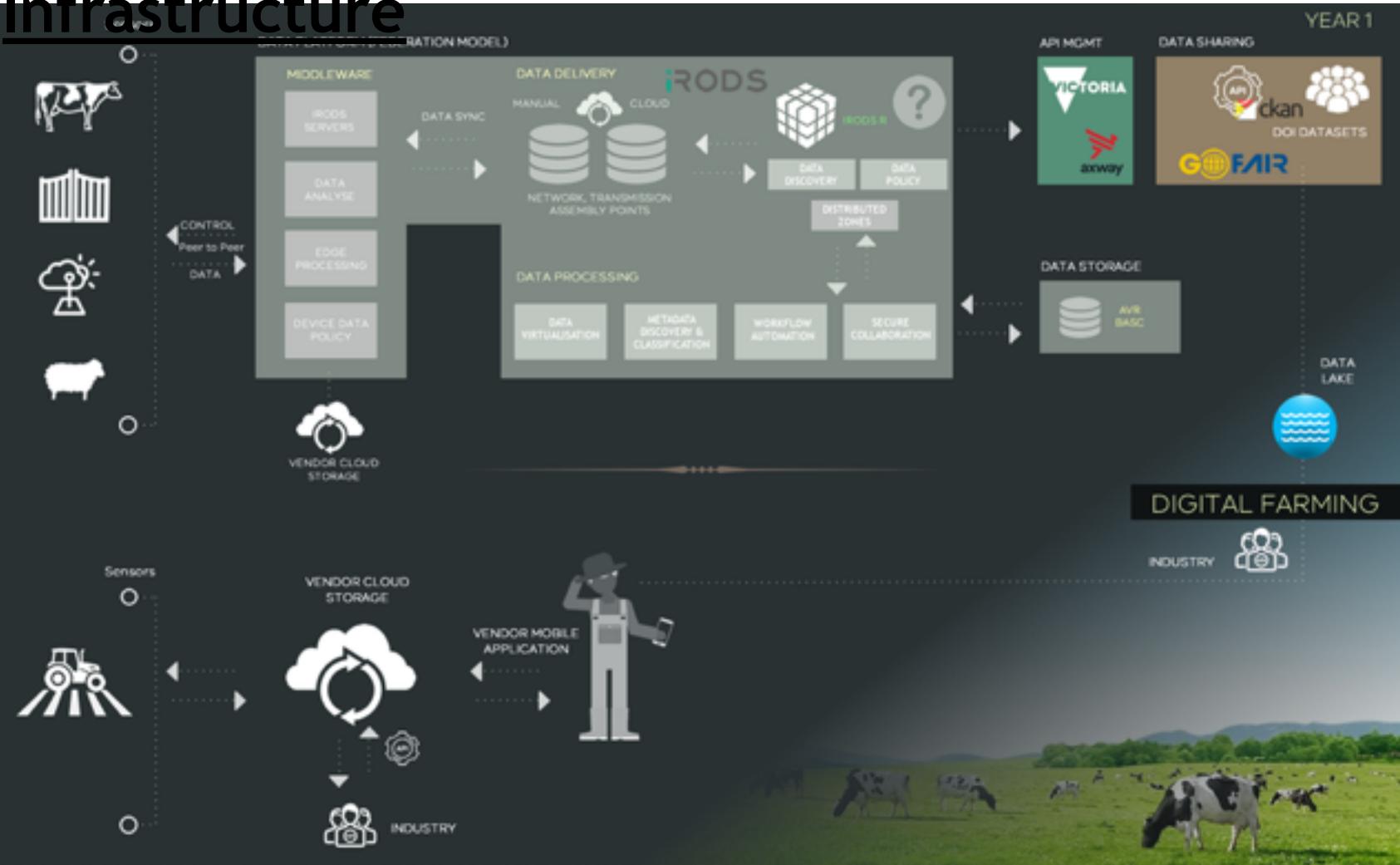
Data may be replicated to HPC storage for analytics.

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

Data may be shared over an iRODS interface : WebDAV, Metalnx, 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 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

The screenshot shows the iRODS graphical user interface. On the left, there's a sidebar with 'Groups', 'Profiles', 'Collections', 'Search', 'Templates', and 'Workflow Links'. The main area has tabs for 'Icons', 'Compact', and 'Details'. Under 'Details', there's a 'Details' section with a 'Name' field containing 'image/tiff'. Below it is a 'Meta' section. To the right, there's a large list of files under 'Name', including '2018-06-05', '2018-06-18 Multispec', etc. At the bottom, there's another list under 'Criteria'.

