



University of
Southern
Queensland

Making an Agricultural Research Dataset FAIR: A case study of the Australian Drought Monitor Dataset

Francis Gacenga, Richard Young – Research Infrastructure

Duc-Anh An-Vo, David Cobon, Jillian Jackson – Centre for Applied Climate
Sciences

Acknowledgement of Country

I would like to acknowledge the traditional owners on the land on which we gather.
I would also like to pay my respects to Elders – past, present and emerging.

In this presentation:

- how we applied the FAIR & CARE principles to the Australian Drought Monitor dataset
- How we enabled its re-use and re-purposing.





University of
**Southern
Queensland**



- The University of Southern Queensland (UniSQ) is a young, dynamic university dedicated to providing quality programs and degrees in a flexible and supportive environment.



- In just over 50 years, UniSQ has become a prominent teaching and research institution providing education worldwide.

@Toowoomba

@Springfield

@Ipswich

AgReFed Platform Project Background



Funding – ARDC + Partners
50:50 co-investment



Partners – Federation
University (Lead), University
of Sydney, University of
Southern Queensland, QCIF



Collaborators – UWA, UoA,
DPIRD, GRDC, CSIRO



Co-operating – Universities,
Industry, Government

Problem Being Solved

Deliver a FAIR agricultural research dataset for reuse on a widely available cloud-based platform.

- Technical aspects – Information Systems, Technologies and Ag Science
- Social – contractual and organisational aspects

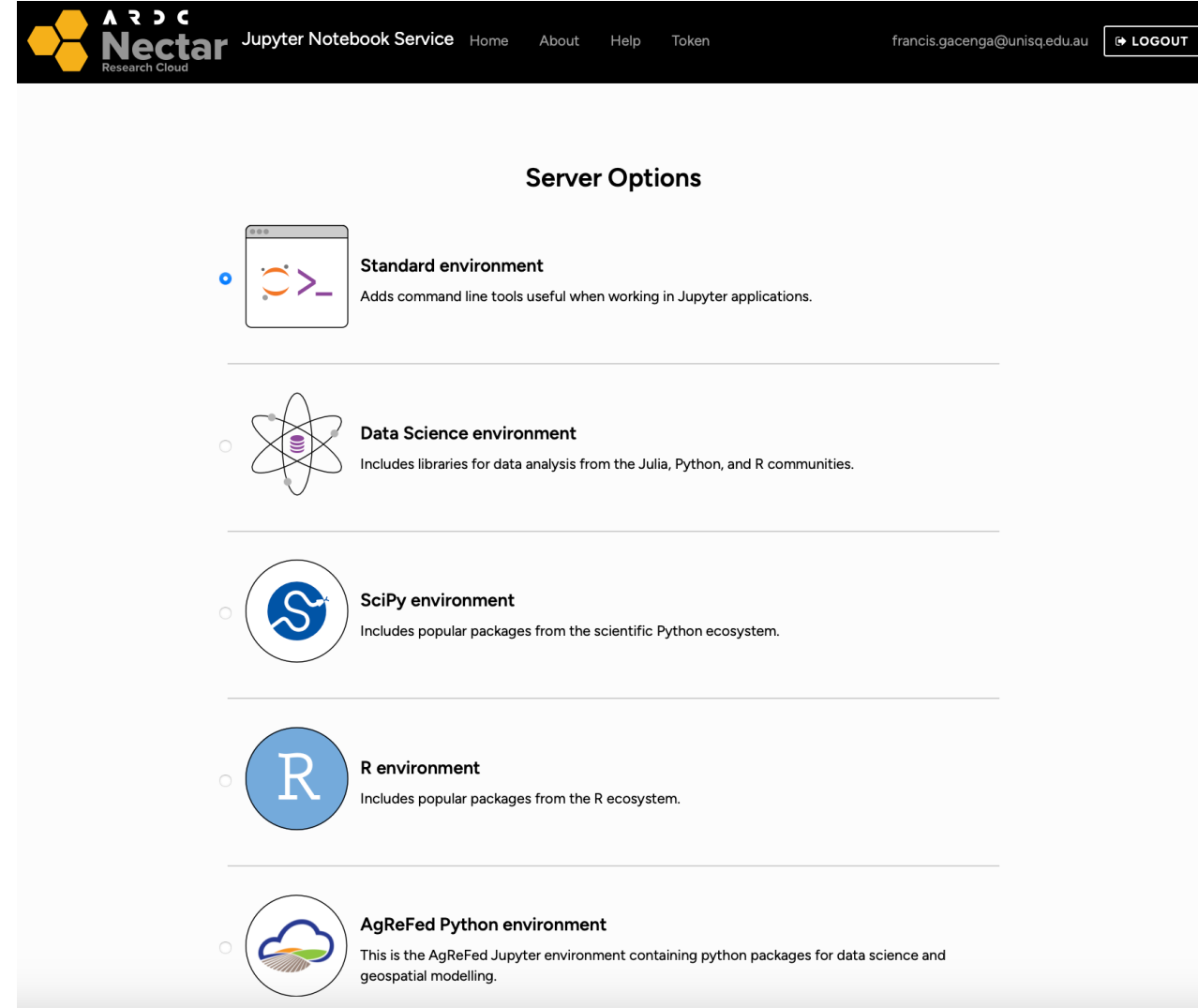


What do you see?



Plan

- Identify a dataset that would be made FAIR and reusable via the AgReFed platform running on ARDC's Jupyter Hub service,
- Utilise federated cloud architecture interconnecting cloud-based systems for data storage and retention,
- Use an approach that can be replicated and that is founded on reviewed literature.



The screenshot shows the Nectar Jupyter Notebook Service interface. At the top, there is a navigation bar with the Nectar Research Cloud logo, the text "Jupyter Notebook Service", and links for "Home", "About", "Help", and "Token". On the right side of the navigation bar, the email address "francis.gacenga@unisq.edu.au" and a "LOGOUT" button are visible. The main content area is titled "Server Options" and lists five environment choices, each with a radio button and a description:

- Standard environment**: Adds command line tools useful when working in Jupyter applications.
- Data Science environment**: Includes libraries for data analysis from the Julia, Python, and R communities.
- SciPy environment**: Includes popular packages from the scientific Python ecosystem.
- R environment**: Includes popular packages from the R ecosystem.
- AgReFed Python environment**: This is the AgReFed Jupyter environment containing python packages for data science and geospatial modelling.

<https://jupyterhub.rc.nectar.org.au/hub/login>

The Original Dataset

The Australian Drought Monitor (ADM) was developed by UniSQ's Centre for Applied Climate Sciences under the Northern Australia Climate Program (NACP) specialised for Australian conditions.

The ADM comprises a software package developed in Python that calculates Combined Drought Indicators (CDI) at a range of time scales e.g., last month, last three months, last six months and last twelve months, based on the United States Drought Monitor (USDM, Svoboda *et al.* 2022).

Originally the only NACP project published outputs based on the dataset were the maps hosted on the NACP website (<https://www.nacp.org.au>) and journal publications on associated research methods and results. The dataset and source code were not published and lacked accompanying metadata.

Data Owners: The NACP-ADM project and subsequent development of the ADM dataset was jointly funded through Meat and Livestock Australia (MLA), the Queensland Department of Agriculture and Fisheries' Drought and Climate Adaptation Program and UniSQ.

Team Roles

Core Roles


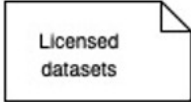
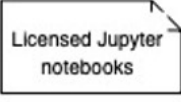
- Data Collection
- Data Curation
- DevOps
- Data Publication

Support Roles

- Contract negotiation
- Metadata and data registration
- Copyright and licensing advise
- Central IT systems



Methodology

Process of Aligning NACP Dataset with FAIR Principles					
Data Collector		Data Curator and DevOps		Data Publication	
Activities		Activities		Outputs	
Findable	Assess FAIRness through self-assessment AgReFed/Research Data Australia platform	Assess for CARE through UniSQ Research Data Management & Indigenous Data Governance Schedule	Determine the dataset(s) size and incremental growth and storage space.	F2 Create the metadata record to accompany the dataset	Registered metadata record
	F2 Create metadata record	F1 Mint dataset DOI		F4, Determine the appropriate compute and storage infrastructure.	
Accessible	A1.2 Contracts negotiations - data sharing agreement		A1, A2 Create storage, compute and data movement pipeline		Published metadata record
Interoperable	I1, I2, I3 Host the metadata record through UniSQ repository		I1, I2 Select formats, vocabularies and protocols that are standard and interpretable		 Complete workflow environments in Singularity and Docker containers
Reusable	R1, R1.2 Associated documentation - Readme file linked to metadata record		R1 Design and selection of platforms and services that will maintain data access post project	R1.3 Use metadata that meet domain-relevant community standards	Licensed datasets  Licensed Jupyter notebooks 

FAIR aligned outputs



Project Outputs: Datasets, Workflows and Training Materials



Findable: Persistent Identifiers (PIDS e.g. DOIs, ORCID) for metadata and data



Accessible: to humans and machines via Servers, VMs and Containers



Interoperable: using open source (Jupyter Notebooks and GitHub) & NCRIS funded infrastructure



Reusable: Use appropriate licensing to protect IP ownership and attributed use.

CARE aligned outputs



Project Outputs: Datasets, Workflows and Training Materials



Collective benefit: ensured that sharing dataset benefits broader community – AgReFed co-op



Authority to Control: clearly defined data ownership and established ethical and legal access



Responsibility: assigned roles for data stewardship, curation and sharing, documented practice



Ethics: adhere to ethical standards throughout the data management cycle.

Technical Solution Development - Iterative



Iteration 0 – Concept phase Thematic Real-time Environmental Distributed Data Services (THREDDS) data server



Iteration 1 – CERN VM-File System on Microsoft Azure Pilot



Iteration 2 – Hybrid Solution - UniSQ On Premise VM server and Nectar VMs still using CERN VM-FS



Iteration 3 – Integration with AgReFed platform using Linux FS, NetCDF, and Github

Delivered Outputs



Data server with management and access capability to deliver monthly update drought monitor datasets for AgReFed and NACP projects



Licensed CDI (Combined Drought Indicator) FAIR dataset published on AgReFed platform



Drought Monitor metadata published on the RDA (Research Data Australia) registry



Documentation of development and Findable, Accessible, Interoperable, Reusable (FAIR) process

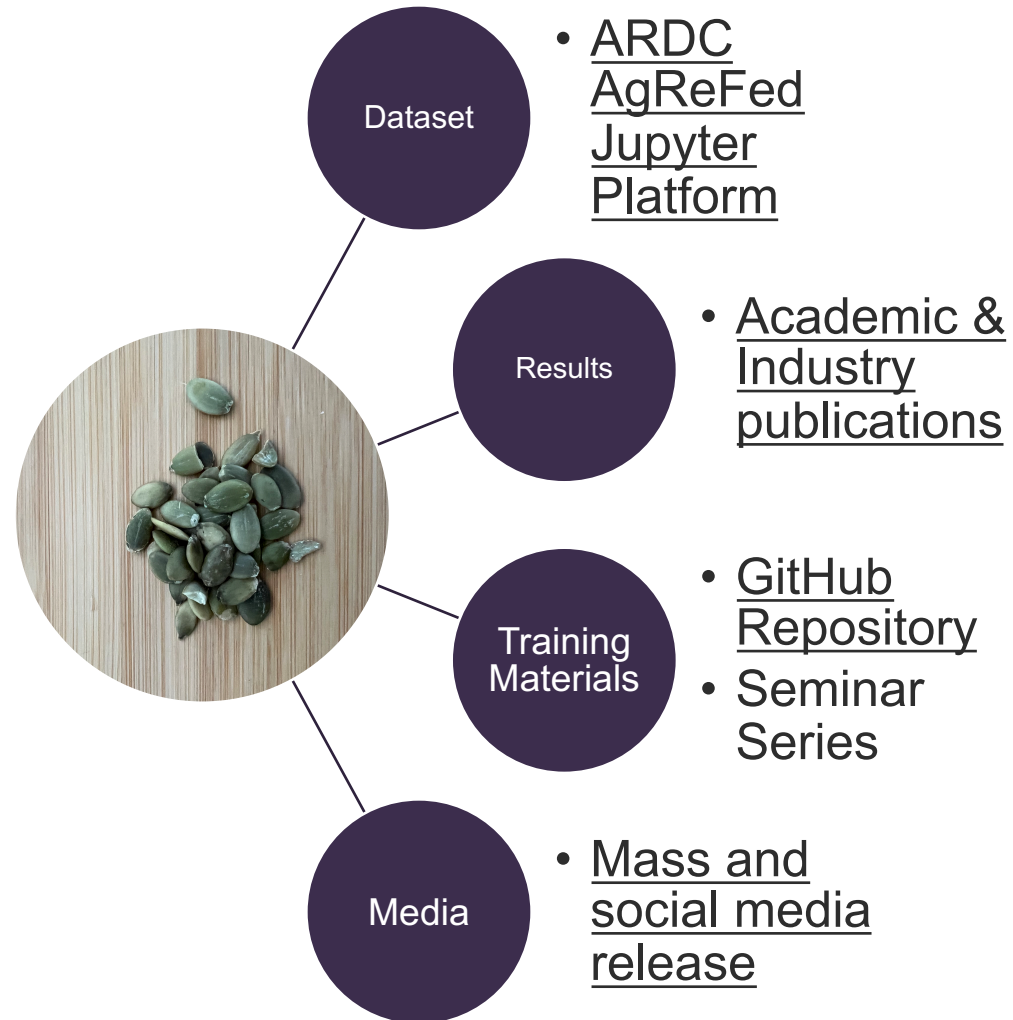


Drought Monitor Use Cases – for application and training via the AgReFed Platform



User Training Material

Propagation – published outputs



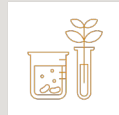
FEDERATED COMMUNITY

Universities

Industry

Government

AUSTRALIAN AGRICULTURAL RESEARCH DATA



PLATFORM

Tools

Tutorials

Case studies

Examples



FAIRness of Australian agricultural research data

IMPACT

Promotes
collaboration

Informs
decision-making

Accelerates
innovation

Builds
capacity

Acknowledgements

Partners

[Australian Research Data Commons](#)

[Federation University](#)

[University of Sydney](#)

•[QCIF](#)

[MLA and QDAF](#)

Internal Project Contributors

Centre for Applied Climate Sciences

Duc-Anh An-Vo, Shahbaz Mushtaq, David Cobon, Jillian McCulloch, Laura Guillory

Research Infrastructure Team

Francis Gacenga, Richard Young, Chris Davey, Sam Rose

UniSQ Enterprise Services

ICT, Library, Contracts and Legal teams





University of
**Southern
Queensland**

Contact:

francis.gacenga@unisq.edu.au

unisq.edu.au