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

@Toowoomba **@Springfield** **@Ipswich**

- 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.

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

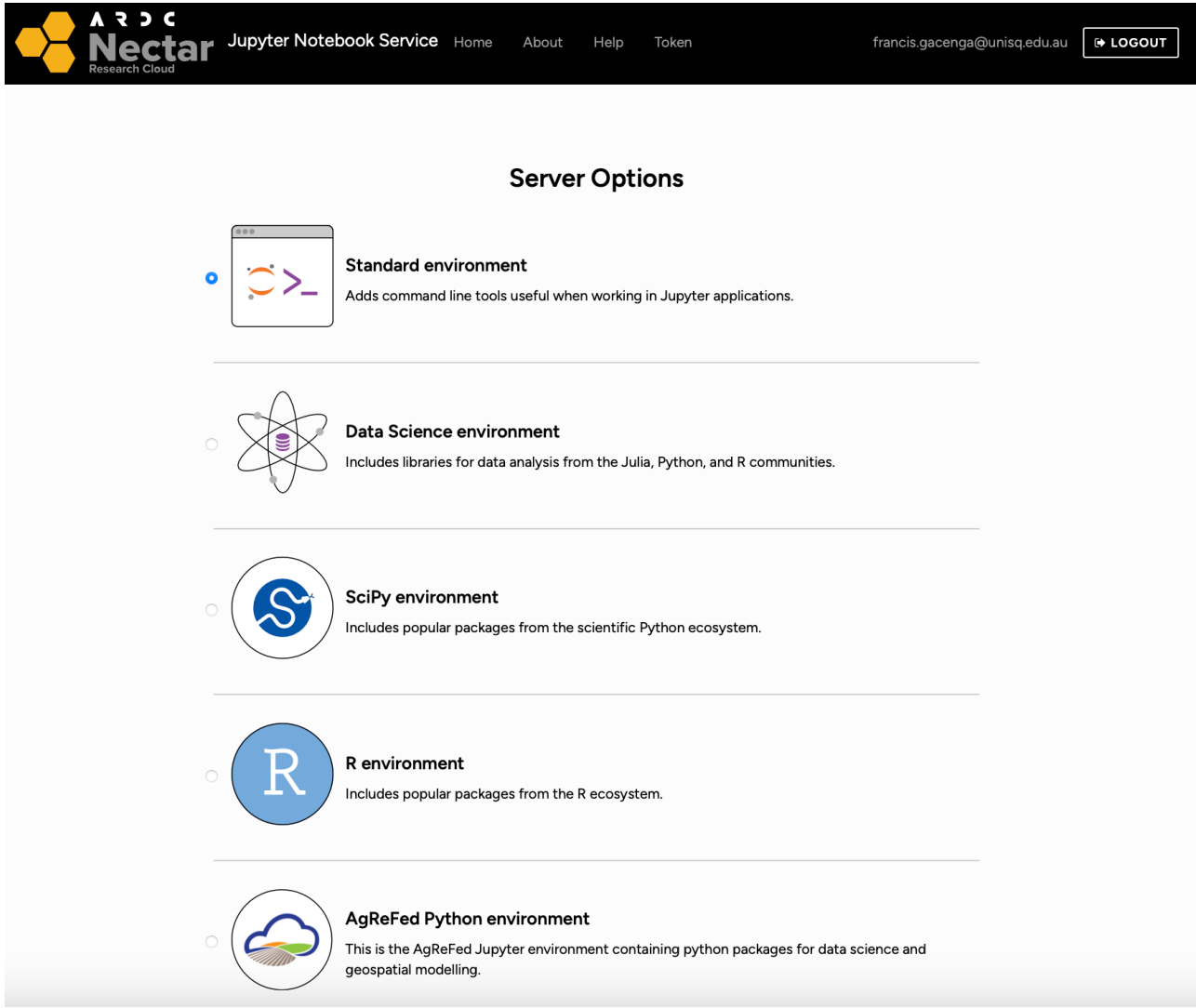
Deliver a FAIR agricultural research dataset for reuse on a widely available cloud-based platform.

- 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.



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

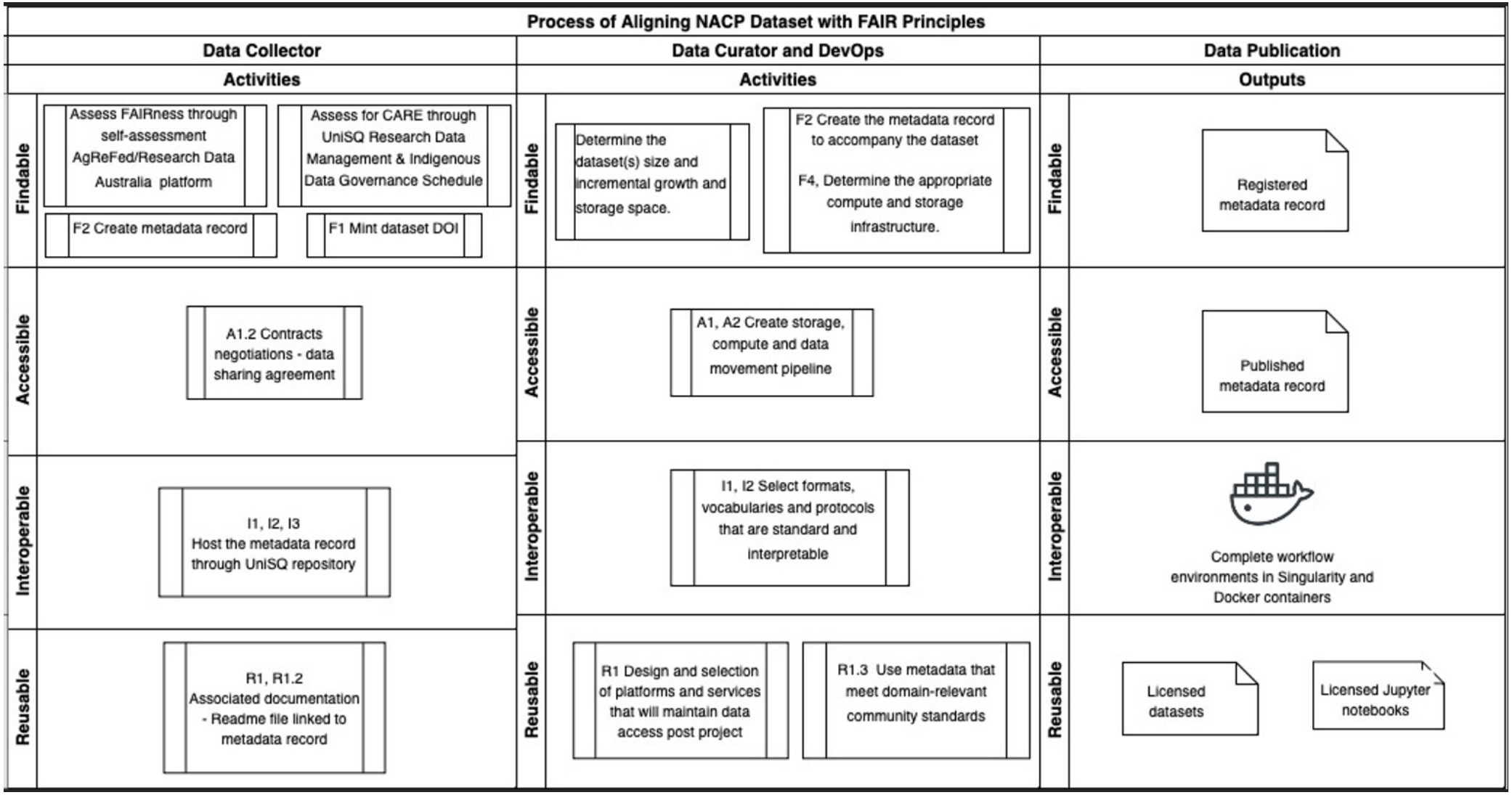
Core Roles

- Data Collection
- Data Curation
- DevOps
- Data Publication

Support Roles

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

Methodology



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

- Dataset
  - ARDC AgReFed Jupyter Platform
- Results
  - Academic & Industry publications
- Training Materials
  - GitHub Repository
  - Seminar Series
- Media
  - Mass and social media release

AgReFed
Agricultural Research Federation

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

AgReFed

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