

# **Open Access Week**

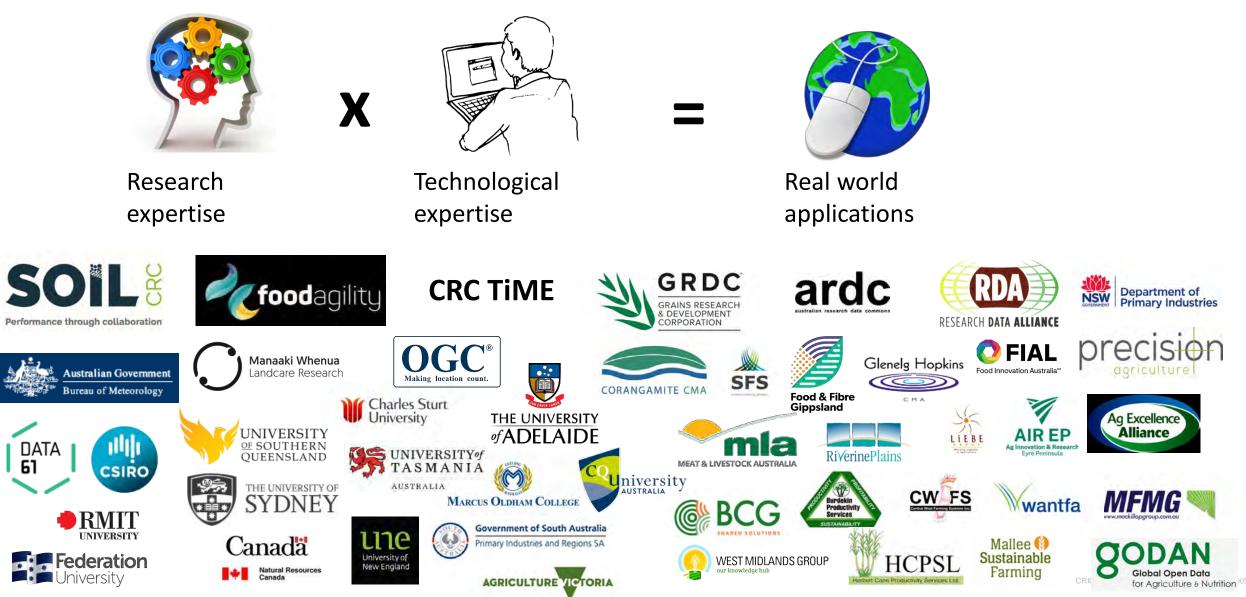
## Open access in research practice: the CeRDI experience

A/Prof Peter Dahlhaus, Principal Research Fellow
Dr Megan Wong, Research Associate
Dr Nathan Robinson, Senior Research Fellow
Centre for eResearch and Digital innovation (CeRDI)

Wednesday 21 October 2020 – Online Research Seminar



Centre for eResearch and Digital Innovation

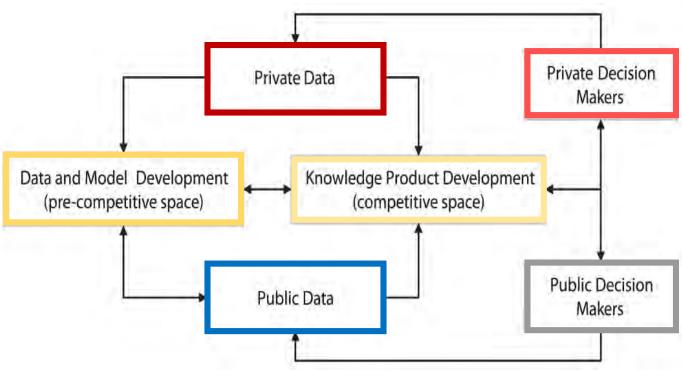


## Spatial data portals - Data federation - Interoperability

30<sup>+</sup> applications across a wide range of domains

Open source, open standards architecture

Data contributed from public and private sectors



#### www.cerdi.edu.au/FedUniSpatial

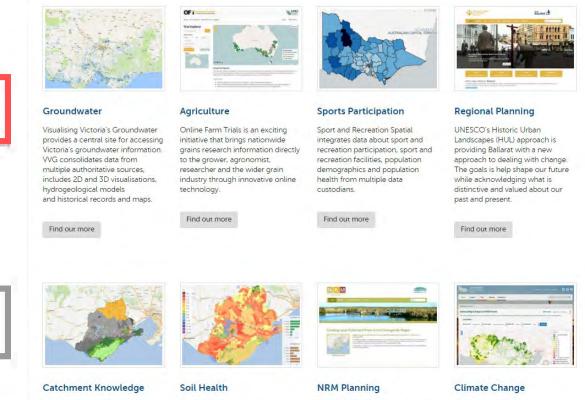
The NRM Planning Portal enables

The South West Climate

#### 🛞 FedUni Spatial

The Corangamite Knowledge

Our spatial research showcases the diverse range of projects that FedUni is supporting through the team at CeRDI. Projects are at various stages of their evolution but share common goals to inform 'big picture' understanding and enhance decision making, create greater efficiencies in communication, increase the quality of information and support policy formulation and evaluation.

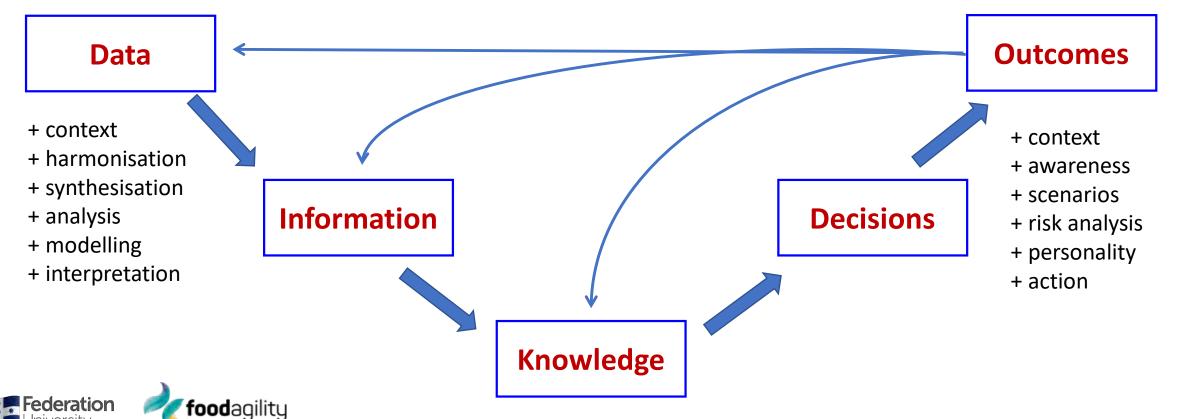


The Corangamite Soil Health

### **Better Data for Better Decisions Constellation**: The challenge

How to harness all the data, information, tools and knowledge

to power better decision making?



## CeRDI - Research 'pillars'

### Discoveries in the data

Federating multi-disciplinary and cross-disciplinary data can lead to new research discoveries

### Innovations in technologies

International implementations of Semantic Web technologies

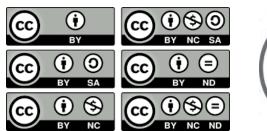
# Open Geospatial Consortium, Inc.



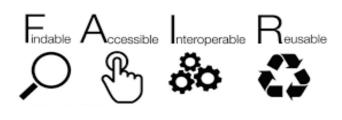
### Longitudinal societal impact research

Does making data FAIR change industry/government/community practice?











Dahlhaus et al. 2016 https://doi.org/10.2166/hydro.2015.169







Jan-16

Jan-17

Jan-18

Jan-15

Jan-13

Jan-14

The VVG portal is frequently used by a wide cross-section of society

## Research challenges



Value Proposition – clear use cases, obvious rewards for data providers and consumers, new research discoveries.



**Social Architecture** –Governance framework, data stewardship, plain language rules, organisation interoperability, rewarding, effortless and no disadvantage to sharing.



**Technical Architecture** – largely resolved but not widely adopted by data custodians, emerging changes such as Linked Data.

## Open access and the FAIR Guiding Principles

Findable, Accessible, Interoperable and Reusable

- Governments around the world have moved to Open Data
- Research funders increasingly require data to be published and available
- Many research journals want the data to be made available
- FAIR data  $\neq$  Open data

# SCIENTIFIC DATA

OPEN SUBJECT CATEGORIES " Research data " Publication characteristics OPEN Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson et al.#

Received: 10 December 2015 Accepted: 12 February 2016 Published: 15 March 2016

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically.



www.nature.com/scientificdata

## Value proposition – FAIR data

Current research practice represents an investment that disappears over time

Multiple benefits in making research data FAIR

- Data, information and knowledge is not lost, forgotten, ignored
- Verification of research outcomes
- Maximising potential for future research
- Enhancing societal outcomes





## Value proposition – FAIR data

### "Data is the new oil"

Basis for longitudinal research studies

- Decision support, Al and machine learning
- New discoveries
- Real world engagement and impact



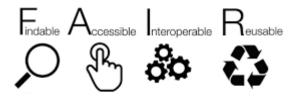


## Key elements to data management and use

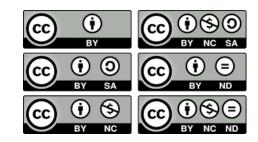
- Governance and stewardship
- Trust relationships and repositories
- Certification standards, vocabularies and core trust seal
- Licensing and ownership
- Education

An approach to assist data providers and consumers to have their data assessed and aligned for "FAIRness"









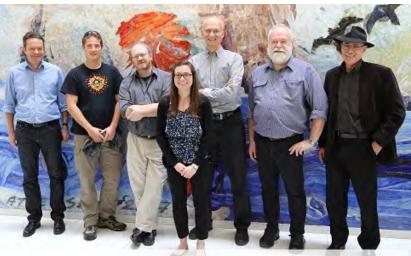


## International data interoperability standards

Groundwater Interoperability Experiment 2 (GW2IE) 2012 - 2016 OGC WaterML2: Part 4 – GroundWaterML2 (GWML2) 2016 -*World Meteorological Organisation (WMO) adoption of GWML2 is pending* Soil Data Interoperability Experiment (Soil IE) 2014 – 2016

Second Environmental Linked Features Interoperability Experiment (SELFIE) 2018 - 20

Mostly collaborations of participants from organisations in European Union, North America & Oceania



GW2IE meeting, Vienna, May 2014

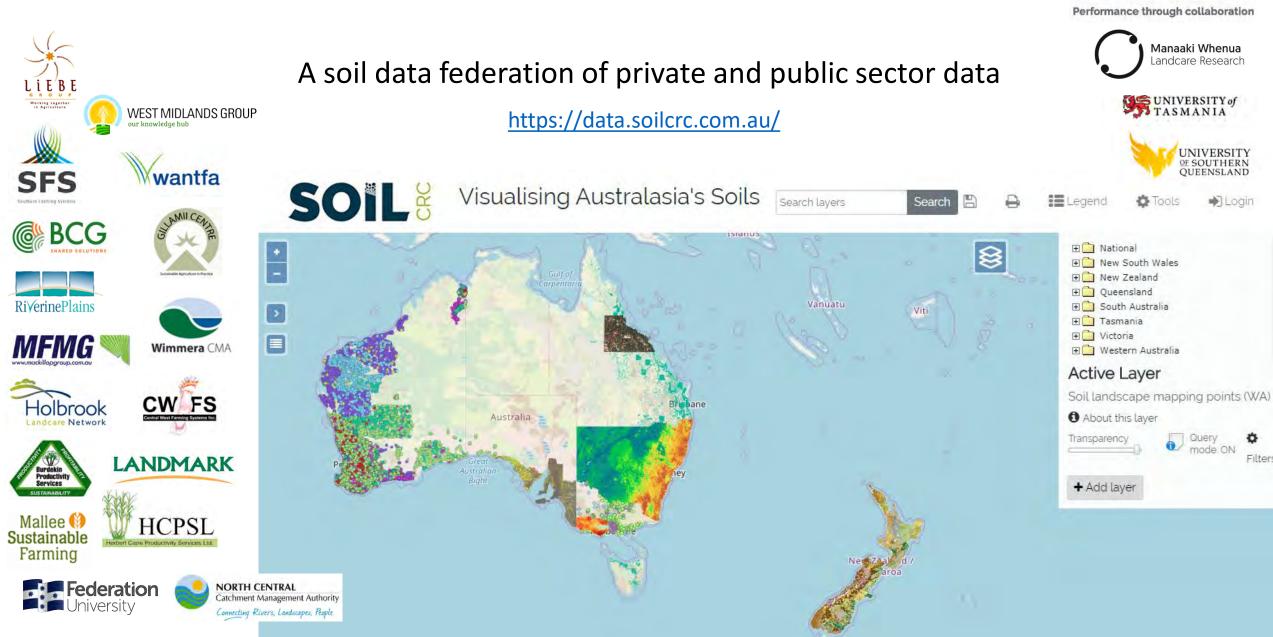






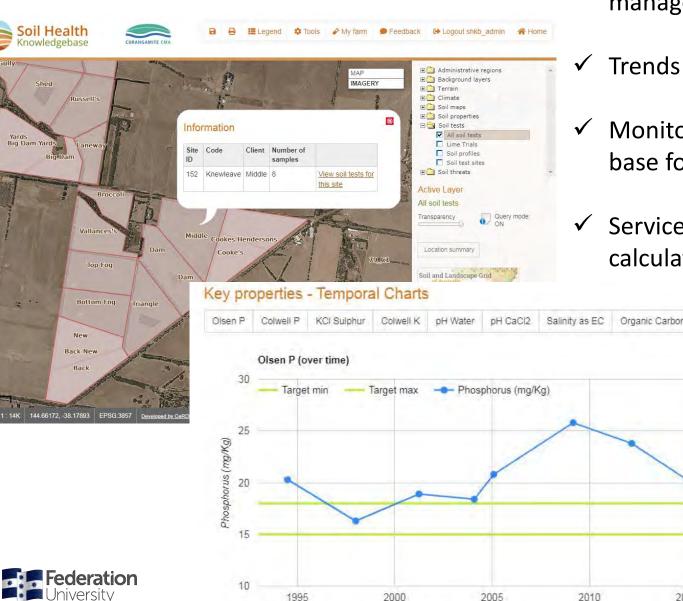
Interest Group on Agricultural Data (IGAD) meeting, Paris, Sept. 2015

## Visualising Australasia's Soils



SOIL

## Value propositions



- Online, free, trusted, supported, independent, spatial data management system
- ✓ Trends over time, benchmarking against local areas
- Monitoring, identify gaps, report to investors, evidential base for investment or social licence
- Services to group members, education, alerts, tools and calculators, decision support, combined data

#### Key properties - Temporal Charts

2015

Year

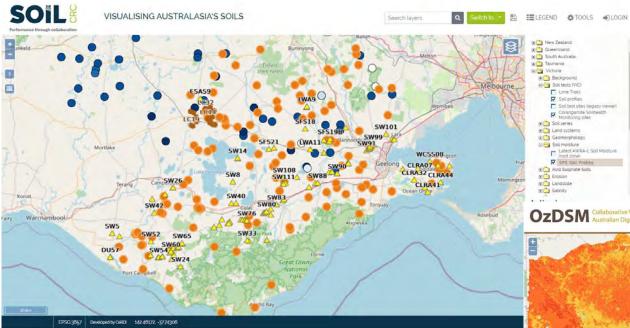


## Data custodian concerns

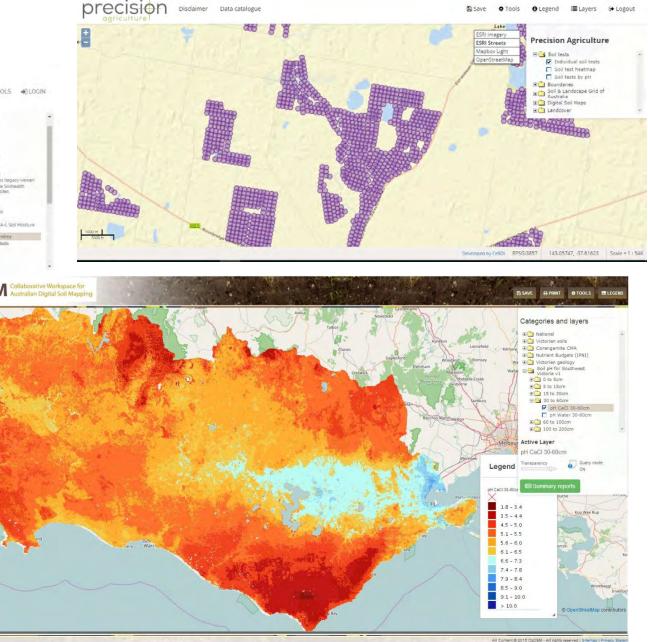
- data sharing privacy, security, trust, context
- resources time, skills, people
- appropriate respect for growers, project managers, and local researchers for time, effort and expertise that underpins the data they provide



## Data access controls

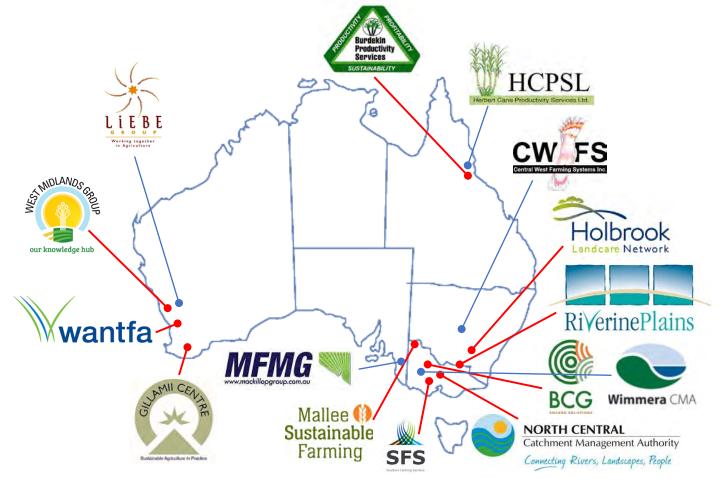


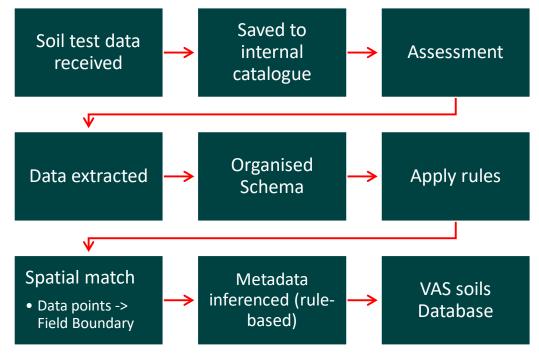
- ✓ Data access controls to allow visibility to the public, club or individual
- $\checkmark$  Can be anonymised to gridded data sets
- Embargoes for data sets, or fuzzy locations



## Private sector data

All project partners we visited have committed to some test data





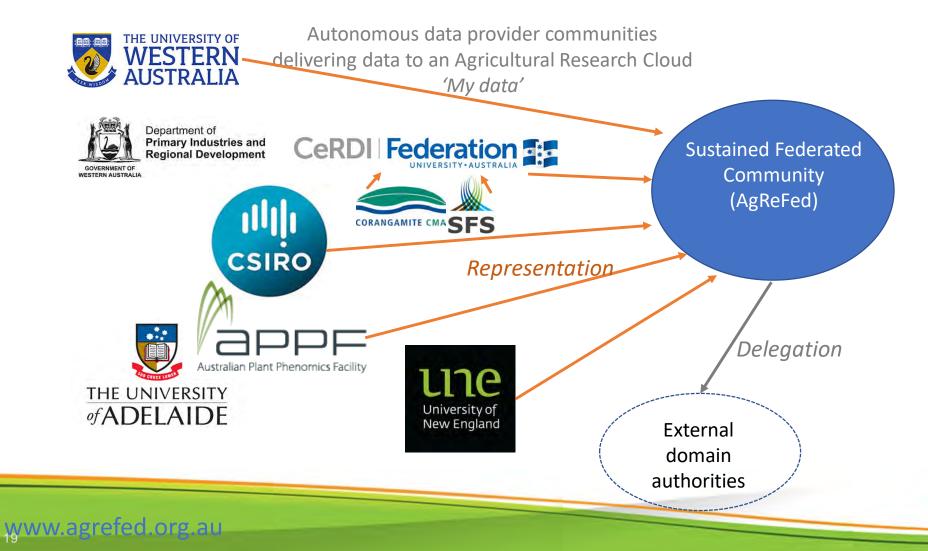
## Public sector data

- More than 24 thousand data sets are listed in open data catalogues with 'soil' as keyword
- Of these, 2166 are currently being assessed for inclusion in the VAS portal (point, polygon and raster data)
- Most are legacy data sets

Source	Scope	Home URL	datasets
Australian Government	National	data.gov.au	18351
Research Data Australia	National	researchdata.ands.org.au	2890
National Library of Australia	National	www.nla.gov.au	634
WA Government	WA	data.wa.gov.au	530
TERN	National	tern.org.au	453
NSW SEED	NSW	datasets.seed.nsw.gov.au	311
Google Datasets	National	toolbox.google.com/datasetsearch	
Queensland Spatial	QLD	<u>qldspatial.information.qld.gov.au</u>	254
NSW Government	NSW	<u>data.nsw.gov.au</u>	214
CSIRO	National	data.csiro.au	190
NT Government	NT	data.nt.gov.au	157
Geoscience Australia	National	services.ga.gov.au	114
SA Government	SA	data.sa.gov.au	103
Victorian Spatial Datamart	VIC	services.land.vic.gov.au	74
Tasmanian Government (The List)	TAS	www.thelist.tas.gov.au	73
Victorian Government	VIC	data.vic.gov.au	33
Queensland Government	QLD	www.data.qld.gov.au	28
ACT Government	ACT	www.data.act.gov.au	5
ASRIS	National	www.asris.csiro.au	5
SA waterconnect	SA	www.waterconnect.sa.gov.au	3
TOTAL			24677

Aim: to increase the value of Agricultural research data in Australia by making it more discoverable and re-usable (including analysis ready)

Access arrangements are an important part of moving 'My data' to 'Our FAIR data'



#### The data (so far)

- Rotational crop trials
- Frost nursey trials
- Soil moisture sensors and networks
- Weather station sensor network
- National soil data and services
- Soil sample dataset



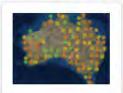
- Agricultural data being made more discoverable and re-usable in line with community expectations
- Data made accessible from multiple repositories, from across different organisations



The vision The approach Explore the data Use cases Get involved Who are we Resources FAQs Contact us Home

#### Explore the data

#### Featured Datasets



#### CSIRO - National Soil Site Database (NatSoil)

Contains descriptions of approximately 16,000 soil site investigations. The data includes morphological descriptions and chemical and mineralogical properties.





#### FedUni - SFS Soil Moisture Probe Network

A network of ~75 telemetered soil moisture probes across Victoria and Tasmania. Data typically includes soil moisture and soil temperature readings at depths 300-1000mm



#### FedUni - Corangamite Soil Health Monitoring Program

A set of 100 soil health monitoring sites from across the Corangamite CMA region of Victoria. Data includes pH, EC, nutrients (P,K,N,S), trace elements and Carbon fractions

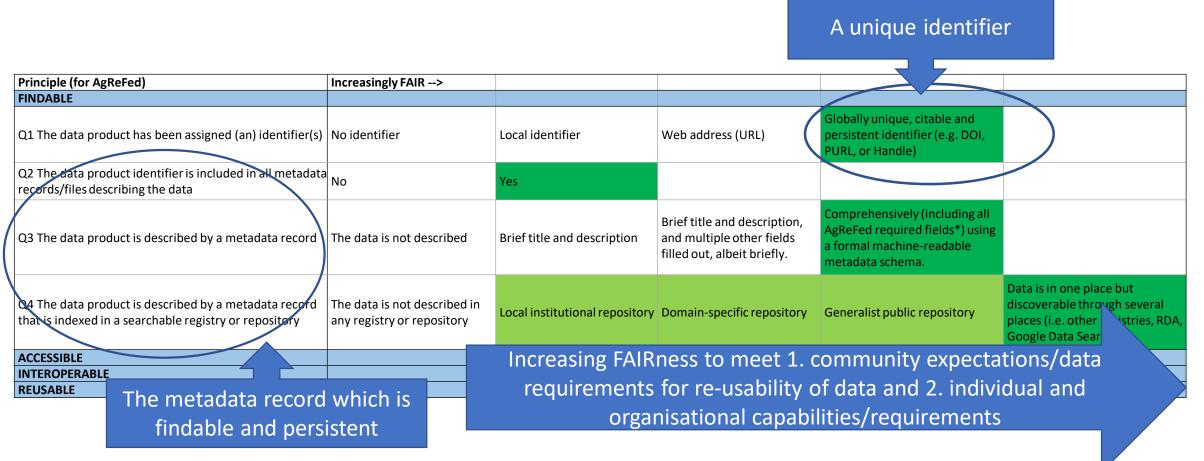


#### University of Adelaide - Waite Field Trials

The dataset brings together yield information from rotational crop trials together with weather and soil information covering multiple decades. It includes data from g parameters







#### AgReFed Community 'rules' (policy) relating to *access*:

\* Q3 - Minimum metadata requirements were specified. See <u>https://doi.org/10.25919/5cf179ba35db9</u>



www.agrefed.org.au

#### AgReFed Community 'rules' (policy) relating to *access*:

A clear access statement 'As open as possible, as closed as necessary'

Principle (for AgReFed)	Increasingly FAIR>				
FINDABLE					
ACCESSIBLE					
Q5 How accessible is the data? The access method(s) must be explicitly stated in the metadata record, e.g. if any authentication is needed, or there are any restrictions to access.	No metadata record	Access to metadata only	Unspecified access conditions e.g. "contact the data custodian to discuss access"	specified date; or A deidentified version of the data is publicly	ully accessible public, or to ersons who meet and follow cplicitly stated conditions and rocesses, e.g. ethics approval or sensitive data
Q6 Data are available for reuse via a standardised communication protocol, such as file download over https, or a web service.	No access to data	By individual arrangement	File download from online location	Non-standard web service (e.g. St OpenAPI/Swagger/informal API) O	randard web service API (e.g. GC)
Q7 The repository/registry agrees to maintain the persistence of the metadata record, even if the data product is no longer available.	No (or not applicable, if no metadata record exists)	Unsure	Yes		
INTEROPERABLE					
REUSABLE					
				It has to be accessible (technically c	

the web)



www.agrefed.org.au

#### AgReFed Community 'rules' (policy) relating to *access*:

## Standard licence readable to humans and machines

Principle (for AgReFed)	Increasingly FAIR>	numans and machines				
FINDABLE						
ACCESSIBLE						
INTEROPERABLE						
REUSABLE						
Q12 Machine-readable data licenses are assigned to each data product, and are stated in the metadata record.	No license is applied	without a license deed URL encoded in a machine-readable	Non-standard license applied, WITH the license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record	Standard license applia Creative Commons), w license deed URL enco machine-readable forr RDF/XML) in the meta	vithout a ded in a mat (e.g.	Standard license applied (e.g. Creative Commons), WITH the license deed URL encoded in a machine-readable format (e.g. RDF/XML) in the metadata record
Q13 The provenance of the data product is described in the metadata, i.e. project objectives, data generation/collection (including from external sources) and processing workflows.	No provenance information is recorded		Comprehensively recorded in a text format (i.e. TXT or PDF)	Comprehensively reco machine readable form metadata record's schu or in RDF, JSON, NetCE	nat (i.e. in ema or PROV,	
Q14 The preferred citation for the data product is provided in metadata record	No	Citation does not include identifier	Citation includes identifier			

citation



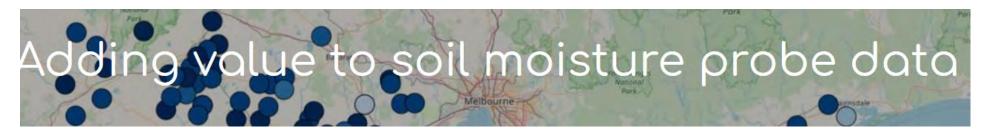
www.agrefed.org.au

# Challenges for providers of agricultural related research data in making data accessible

- Understanding roles, rights and obligations around providing (or restricting) data access.
  - Honouring contractual agreements between data originators (e.g. farmers) and service providers (e.g. contractors), funders and data provisioners
- Understanding institutional/organisational policies around data/IP access agreements (e.g. embargo periods on thesis data)
- Support/extension required re. licencing application
  - Understanding the pros and cons of different standard licencing (Open vs more restrictive)
  - Education required around FAIR does not = Open.
- Different data storage and data access policies, processes and information and technology support and priorities across agricultural research related organisations, examples -
  - Legacy database use/preference
  - Different repository choices
  - IT data access protocols Application programming interface APIs
- Close collaboration required between research groups, IT, library, IT and research data experts (example through Australian Research Data Commons ARDC)



### Data access use case: Soil sensor networks



### The data

Southern Farming System's (SFS) 🗹 soil moisture network across Victoria and Tasmania was stored in a closed database. The dataset was difficult to discover and the access conditions were not clearly specified to potential users of the data. Each individual request for data needed to be executed by research or technical staff, which was a poor use of time.

### The challenge

The challenge was make data discoverable and accessible for research, but only under conditions specified by the data provider to protect the rights of the data owners.

SFS was interested in providing data to researchers for predictive modelling of soil moisture.



### Data access use case: Brokered access – soil moisture probes

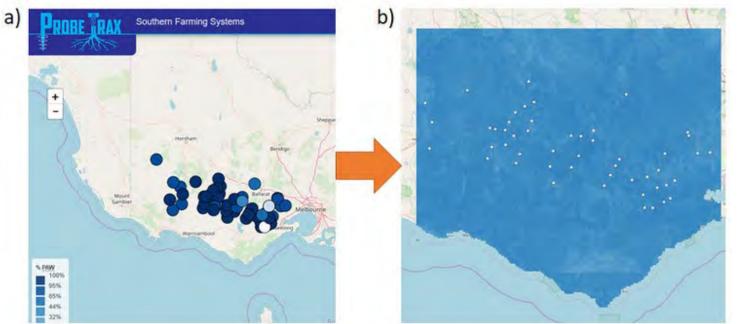
Systems Analyst Programmer Scott Limmer, CeRDI:

"We now have machine learning on services set up over the soil moisture probes. This is producing interpolated maps, that is predictions of soil moisture between the locations of probes"

"Providing the data as a service makes it much easier for researchers to provide these maps - the models can run automatically and be updated overnight"

"These maps will provide added value to the data provider (Southern Farming Systems) through making their data more FAIR through AgReFed"

The data is now more easily discoverable via the internet, and accessible directly to users via an authentication system. This means the soil probe data is more likely to be re-used whilst protecting the rights of the data owners.



Soil moisture observations at specific soil probe locations (a) is now being used for dynamic modelling to predict soil moisture levels between probe locations (b).

Licence & Rights: Non-Commercial Licence hide details Non-Commercial Licence https://creativecommons.org/licenses/by/4. 0/

Access to this collection data is partly restricted. The most current data (last 3 months) is not publicly available. Additionally, the exact location of the individual probes is obfuscated within a 5km radius of the actual location. Access to the full dataset may be granted following approval of a written application. Access:

Conditions apply Contact Information Street Address: 23 High Street, Inverleigh VIC 3321 Ph: 03 5265 1666

office@sfs.org.au



### Data access use case: Frost trial data

### The data

Partners through the ARC Centre of Excellence in Plant Energy Biology collect a wide range of datasets to investigate the desirable plant traits and growth conditions for crop yield under frost, including plant growth trait, hyper- and multi-spectral data, proteomics and metabolomic data.

### The challenge

Datasets were all processed and stored in separate locations. After publication or report writing, the data effectively became "lost" to other researchers. The researchers wanted the data they generated to be findable and useable by others in their team and more widely to advance crop science.







Department of Primary Industries and Regional Development



### Data access use case: Frost trial data

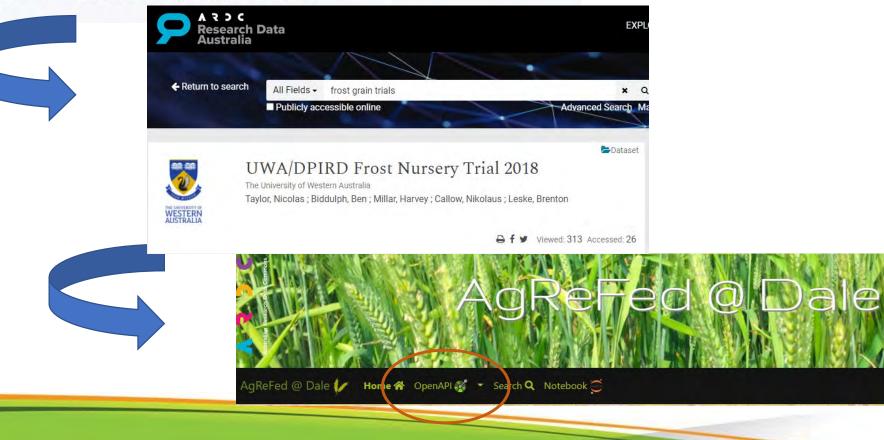


#### UWA/DPIRD Frost Nursery Trial 2018

Dataset

NICOLAS TAYLOR (DATA MANAGER), BEN BIDDULPH (DATA MANAGER), HARVEY MILLAR (DATA MANAGER), NIKOLAUS CALLOW (DATA COLLECTOR), BRENTON LESKE (CREATOR)

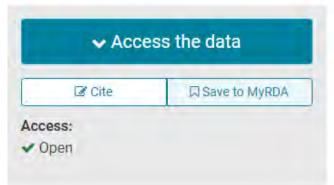
School of Molecular Sciences, ARC Centre for Plant Energy Biology, UWA School of Agriculture and Environment



Discover PEB >

## Data access use case: Open access - Frost trial data

- Discoverable through Reseach Data Australia, UWA repository, AgReFed
- Access: To the broader scientific community via an OpenAPI <u>http://webapps.plantenergy.uwa.edu.au/agrefed\_dale/</u>and through AgReFed



•The team, and other researchers are are now able to select, combine and download by other co-variates of interest including include plant growth/grain yield, ProteoMetabolite data and on-site weather data, as well as factors such as sowing time,

variety



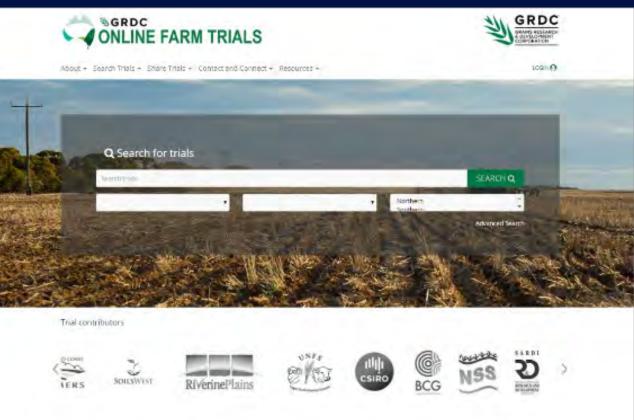
Senior Lecturer Dr Nicolas Taylor

"This has been a really great test case for how we can integrate all these types of data. For example, yield has been integrated with drone data. We are starting to see some really interesting things."





## Online Farm Trials An example/experience of Open Access agricultural data





### What is Online Farm Trials (www.farmtrials.com.au)

- Online Farm Trials beginnings in 2013/14; 3 initial contributors of trial research data and reports to the system.
- Focus on the system was to make trial data and information discoverable (*findable* and *accessible*) to the grains community.
- Australia rich history of trial research for over 150 years.
- There are now over 80 contributing organisations to OFT thousands of publicly accessible and available trials in 1 location.



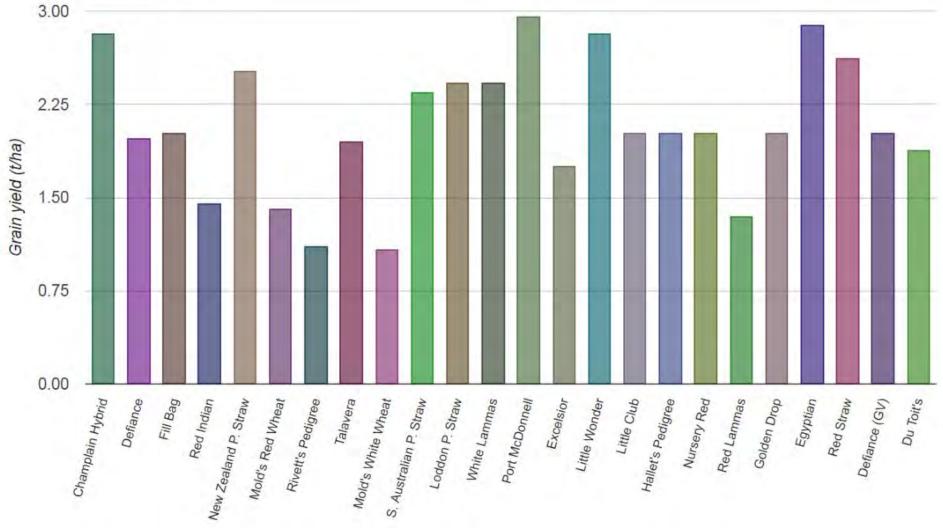
## Yield trials – Dookie (1884)

*Victoria Agriculture Report for the Year 1884* (1885); Department of Agriculture.

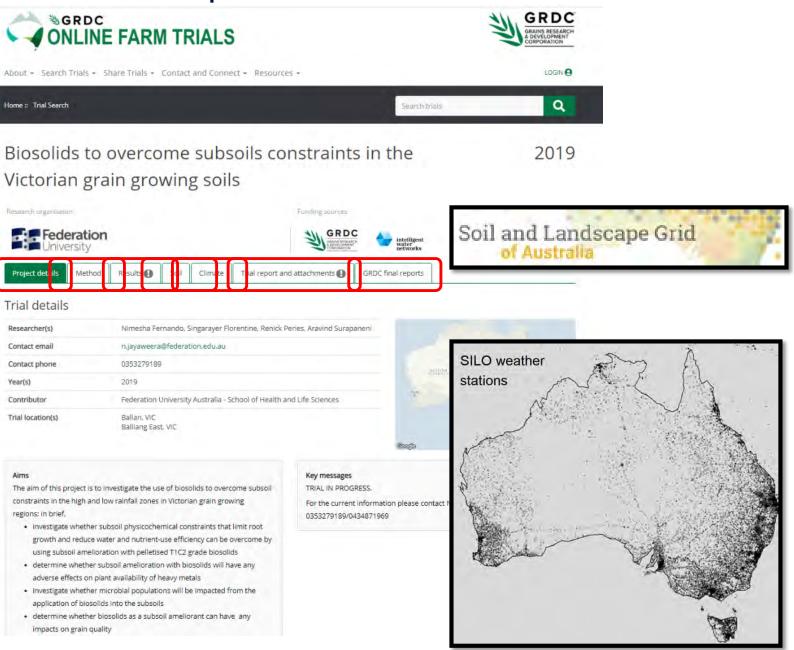
-	No.	Variety of Wheat.	Where procured.	Amount sown.	Date when sown.	Size of plot.	Rainfall while maturing.	No. of days maturing.	Date when harvested.	Weight of grain, 1st.	Weight c grain, 21	Rate por acre.
	$1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \\ 16 \\ 17 \\ 18 \\ 19 \\ 20 \\ 21 \\ 22 \\ 23 \\ 24 \\ 10 \\ 11 \\ 12 \\ 13 \\ 14 \\ 15 \\ 16 \\ 17 \\ 18 \\ 20 \\ 21 \\ 22 \\ 23 \\ 24 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 10 \\ 1$	Champlain Hybrid Defiance Fill Bag Red Indian New Zealand P. Straw Mold's Red Wheat Rivett's Pedigree Talavera Mold's White W S. Australian P. Straw Loddon P. Straw White Lammas Port McDonnell Excelsior Little Wonder Little Wonder Little Club Hallet's Pedigree Nursery Red Red Lammas Golden Drop Egyptian Red Straw Du Toit's	South Australia ,, France South Australia France South Australia Farm , California Katandra , Katandra , Katandra , Katandra Goulburn 	2 lbs. 2 lbs. 2, 1, 1, 2,, 2,, 2,, 2,, 2,, 2,, 2,, 2,, 2,, 2,, 2,, 2,, 2,, 2,, 3,, 2,, 3,, 2,, 3,, 2,	1884. 16 May ", ", ", ", ", ", ", ", ", ",	80th of acre 40th of acre 30th of acre """"""""""""""""""""""""""""""""""""	$\begin{array}{c} 10 \cdot 63 \\ 10 \cdot 68 \\ 10 \cdot 87 \\ 10 \cdot 68 \\ 10 \cdot 87 \\ 11 \cdot 02 \\ 11 \cdot 02 \\ 11 \cdot 02 \\ 10 \cdot 87 \\ 10 \cdot 68 \\ 10 \cdot 87 \\ 10 \cdot 68 \\ 10 \cdot 68 \\ 10 \cdot 87 \\ 11 \cdot 02 \\ 10 \cdot 68 \\ 10 \cdot 68 \\ 10 \cdot 87 \\ 10 \cdot 68 \\ 10 \cdot 87 \\ 10 \cdot 01 \end{array}$	$\begin{array}{c} 202\\ 213\\ 210\\ 202\\ 213\\ 226\\ 226\\ 213\\ 226\\ 213\\ 202\\ 213\\ 202\\ 213\\ 202\\ 210\\ 213\\ 226\\ 226\\ 226\\ 226\\ 226\\ 226\\ 226\\ 202\\ 210\\ 192 \end{array}$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	$\begin{array}{c} \mathrm{lbs.}\\ 84\\ 59\\ 45\\ 43\\ 75\\ 42\\ 33\\ 58\\ 32\\ 70\\ 72\\ 72\\ 88\\ 52\\ 84\\ 60\\ 15\\ 15\\ 10\\ 15\\ 86\\ 78\\ 60\\ 56\end{array}$	$\begin{matrix} \text{lbs.} \\ 6 \\ 5 \\ 3 \\ 3 \\ 2^{\frac{12}{3}} \\ 2 \\ 5 \\ 3 \\ 4 \\ 4 \\ 3 \\ 5 \\ 4 \\ 4 \\ 1^{\frac{12}{3}} \\ 1^{\frac{12}{3}} \\ 1^{\frac{12}{3}} \\ 3 \\ 4 \\ 5 \\ 4 \\ 5 \\ 4 \\ 5 \\ 4 \\ 5 \\ 4 \\ 5 \\ 4 \\ 5 \\ 4 \\ 5 \\ 5$	$\begin{array}{c} \text{bush.} \\ 42 \\ 291 \\ 30 \\ 211 \\ 371 \\ 21 \\ 161 \\ 29 \\ 16 \\ 35 \\ 36 \\ 36 \\ 36 \\ 44 \\ 26 \\ 42 \\ 30 \\ 30 \\ 30 \\ 30 \\ 30 \\ 30 \\ 20 \\ 30 \\ 3$







### A more recent example ...





# OFT – coalition of Open Access data relevant to Agriculture, and agricultural data context

- Initial Terms of Use v. Terms of Contribution.
  - **C C**.**G**. The data and information accessed through OFT may also be subject to Terms of Use specified by the Trial Contributor, in which case the Terms of Use of a Trial Contributor take precedence over other information provided on OFT. The Terms of Use for a Trial Contributor or for OFT may change without notice.
- 2017-18: began revision of <u>Terms of Use</u> and <u>Terms of</u> <u>Contribution</u>
- Dispute resolution procedure developed.





### **Terms of Use**

- Initial focus on getting trials online accessible to users.
- Providing clarity on: Terminology, Disclaimer/s, Use of Information and Licensing.
- Aware of trademarks, etc: Plant Breeders Rights, Registered products/names/IP.
- Citations aware of contributor and how to cite OFT as a general source, a trial in OFT or when using data exported from OFT.





### Terms of Contribution (T's&C's)

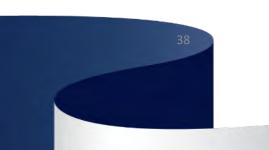
- Involvement of GRDC legal provided guidance and support.
- Priority was on ensuring that trial metadata (details about the trial) was open (Freedom to Operate GRDC priority).
- Asked organisations contributing trials to OFT to make their trials **FAIR**.
- New T&C's apply the Creative Commons Attribution 4.0 International Licence (CC BY 4.0, https://creativecommons.org/licenses/by/4.0/) to data made available via OFT.





### **The Experience/Learning**

- Plethora of Open Access data
- Understanding T's&C's is critical contributing the data and using the data.
- Support of domain experts (data wrangling, sharing, metadata, services, etc).
- •The data should be FAIR





# Open Access Week – Thanks for watching!

## Open access in research practice: the CeRDI experience

A/Prof Peter Dahlhaus, Principal Research Fellow
Dr Megan Wong, Research Associate
Dr Nathan Robinson, Senior Research Fellow
Centre for eResearch and Digital innovation (CeRDI)

Wednesday 21 October 2020 – Online Research Seminar