MAXIMISING THE VALUE OF SOIL DATA: CHALLENGES AND SOLUTIONS WE EMPLOYED

FOCUS: USER CASES OF FARMING GROWER GROUPS AND CATCHMENT MANAGEMENT AUTHORITIES

Wong, M; MacLeod, A; Simons, B; Robinson, N; Gillett, H; Sexton, A; Medyckyj-Scott; Thompson, H; Neyland, A; Milne, R; Corbett, J; Dahlhaus, P

Presenter: Research Assoc. - Dr Megan Wong, mr.wong@federation.ede.au,

PI – Assoc Prof Peter Dalhaus p.dahlhaus@federation.edu.au,
Tech lead/services: Andrew MacLeod a.macleod@federation.edu.au

Federation University Centre of eResearch and Digital Innovation, Australia

Federation University

SOiL CRC
Performance through collaboration

Project/initiatives of farmer and catchment level:

https://data.soilcrc.com.au/map/about

Soil monitoring network
and soil health knowledge base project

https://www.ccmaknowledgebase.vic.gov.au/soilhealth/

Development through related projects:

Vocabulary assistance,
much thanks CSIRO

FOCUS USER CASE: VISUALISING AUSTRALASIA'S SOILS

A SOIL CRC INTEROPERABLE SPATIAL KNOWLEDGE SYSTEM

Federation University

SOIL CRC

Performance through collaboration

To meet scoped end-user experiences:

Search and discovery of data for decision making

- For underpinning on-farm decision making
  - See trends across space and time
  - View data in context with other data at the local, regional, state and national level

Data for communication, with context, that is trusted

- Finding out what other growers and groups are doing
- Avoid duplication, direct new work
- Connect projects – historical and current, on varying topics
- Get a big picture of farm and region activities
- For funding applications, banking, insurance
- Others need to understand the context of the data - Will data be interpreted correctly for decision making if there is no need to contact people with local expertise?
- Privacy, security, trust – will the data be used in ways that benefit growers?

Interface functionality for data management

- Fill skills gap in data management time, skills, expertise
  - Knowledge transfer that means data is not lost when staff leave

Brief: An "Interoperable spatial knowledge system providing Soil CRC participants, and the broader agricultural industry, with access to data, information and knowledge on Australasian soils."

Federation University

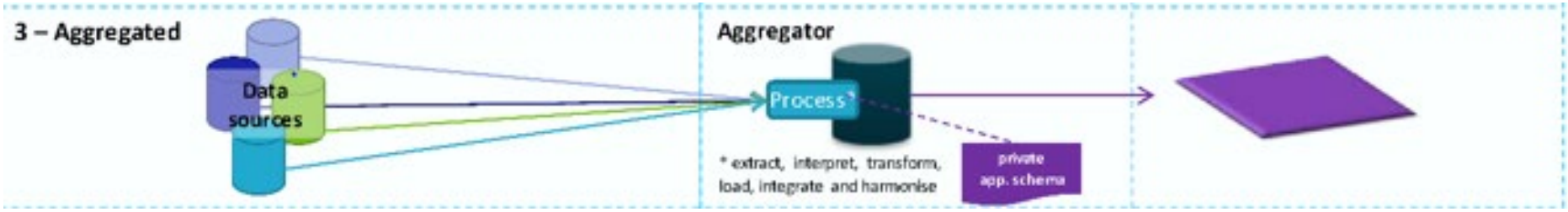
SOIL CRC
Performance through collaboration

Current state: Data providers do not all have the systems and capacities to store and serve data for discovery and re-use through VAS in a well described, consistent format and structure to support delivery of end user experiences

Our Task: To take varying data content and formats from different data providers and make it available to potential users in a standard format, with standard content, via a standard mechanism. That is, to make it more FAIR.

The Solution:

- An 'Aggregator' within the broader VAS system for Grower Group and CMAs wishing to utilise
    - Operationalisation and functionality is backed by a standards based approach
    - Data holdings are submitted, interpreted, transformed, loaded, integrated and harmonised by the aggregator (VAS)



Extract from figure Box et al 2015 – data supply models

Federation University

SOiL CRC
Performance through collaboration

ISO19156 and Open Geospatial Consortium (OGC) and **Observations and Measurements (O&M)** based model to store field and laboratory environmental data.

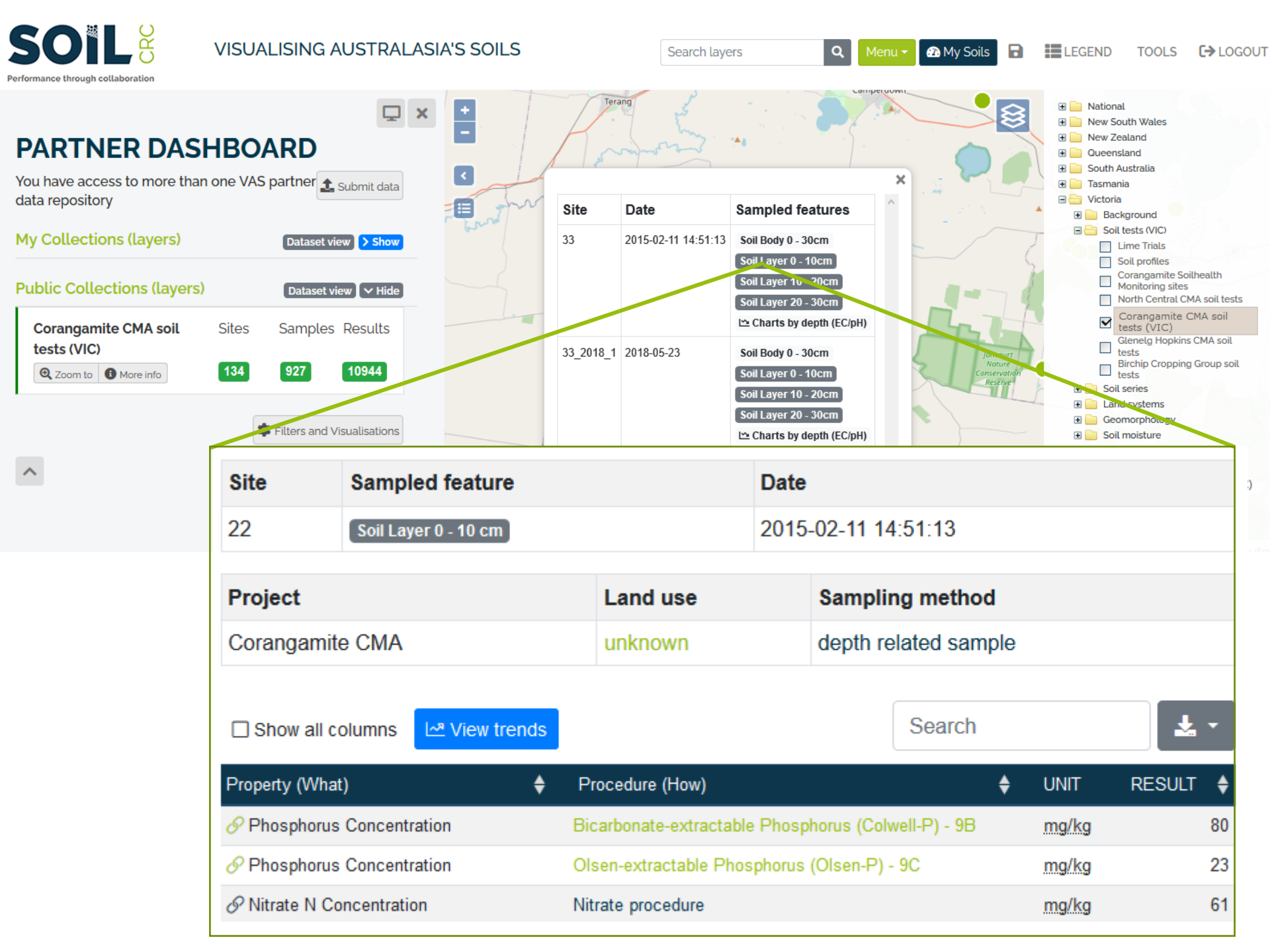
- **Results** (value, term, ranges of values and terms, descriptions)
- **Procedure** used to make observation (e.g. Olsen Phosphorus, loss on ignition)
- **Feature of Interest** observation is made on (e.g. "SoilBody", SoilHorizon", "SoilLayer“)
- **Spatial sampling features (**e.g."SoilBody", SoilHorizon", "SoilLayer", "SoilProfile”)
- **Property being measured** (e.g. Phosphorus concentration)
- **Time** of observation
- **The Substance or species** (e.g. ‘phosphorus’)
- **Specimen** (e.g. a core with accession number)

Data delivered by API endpoints as **JSON-LD** using **SSN/SOSA** (experimental, not fully compliant), RESTful API documented using OpenAPI specification (**OAS3**). Delivery by -

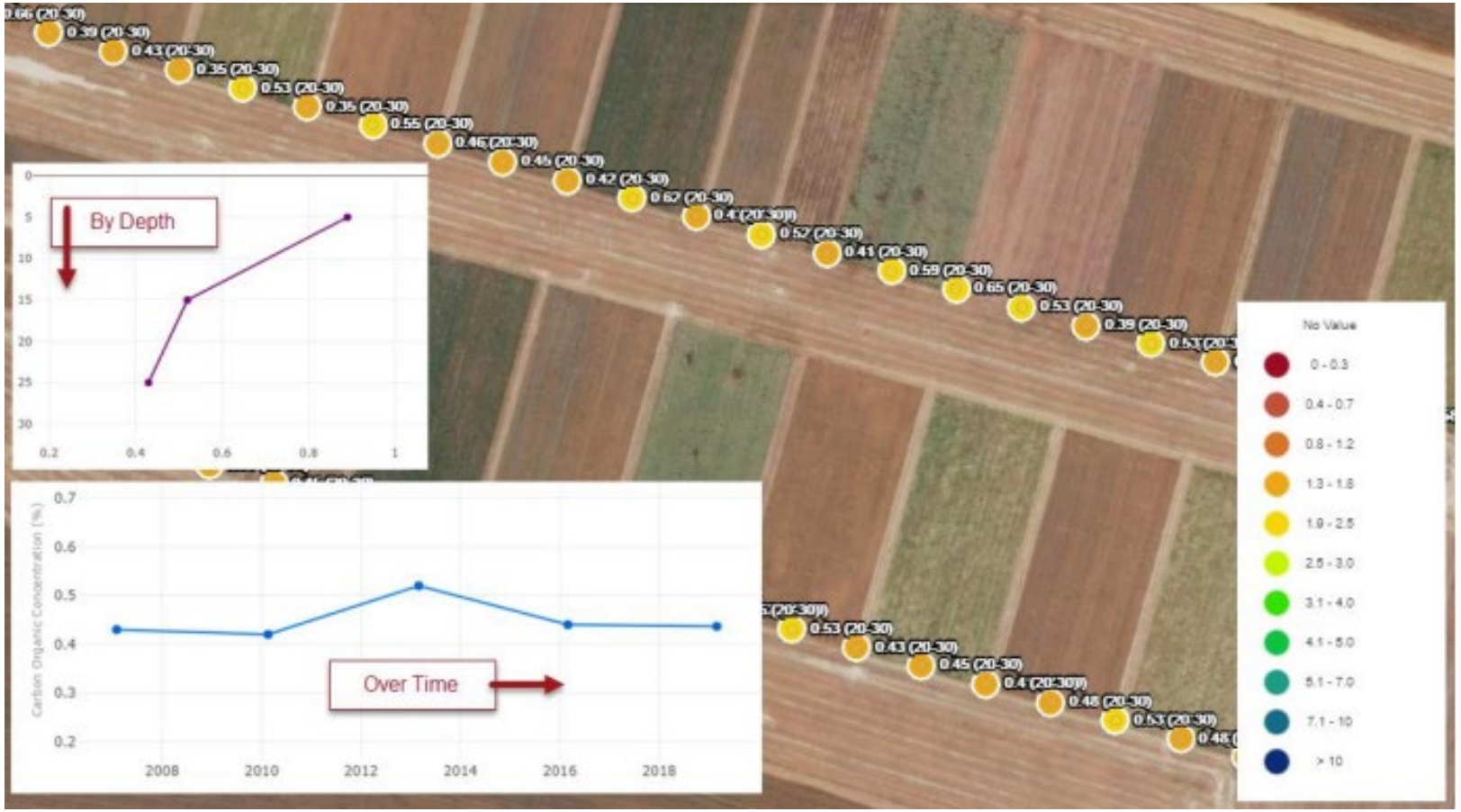
- **Observations** (based on the observed property, procedure used or feature of interest)
- **Sites** (such as plots, pits, paddocks)
- **Specimens**
- **Soil features** (layers, horizons, profiles, bodies)

Federation University

SOiL CRC
Performance through collaboration


SOIL CRC
Performance through collaboration
VISUALISING AUSTRALASIA'S SOILS
Search layers
Menu
My Soils
LEGEND
TOOLS
LOGOUT
PARTNER DASHBOARD
You have access to more than one VAS partner data repository
Submit data
My Collections (layers)
Dataset view
Show
Public Collections (layers)
Dataset view
Hide
Corangamite CMA soil tests (VIC)
Sites
Samples
Results
Zoom to
More info
134
927
10944
Filters and Visualisations
Terang
Site
Date
Sampled features
33
2015-02-11 14:51:13
Soil Body 0 - 30cm
Soil Layer 0 - 10cm
Soil Layer 20 - 30cm
Charts by depth (EC/pH)
33_2018_1
2018-05-23
Soil Body 0 - 30cm
Soil Layer 0 - 10cm
Soil Layer 10 - 20cm
Soil Layer 20 - 30cm
Charts by depth (EC/pH)
National
New South Wales
New Zealand
Queensland
South Australia
Tasmania
Victoria
Background
Soil tests (VIC)
Lime Trials
Soil profiles
Corangamite Soilhealth Monitoring sites
North Central CMA soil tests
Corangamite CMA soil tests (VIC)
Glenelg Hopkins CMA soil tests
Birchip Cropping Group soil tests
Soil series
Land systems
Geomorphology
Soil moisture
Site
Sampled feature
Date
22
Soil Layer 0 - 10 cm
2015-02-11 14:51:13
Project
Land use
Sampling method
Corangamite CMA
unknown
depth related sample
Show all columns
View trends
Search
Property (What)
Procedure (How)
UNIT
RESULT
Phosphorus Concentration
Bicarbonate-extractable Phosphorus (Colwell-P) - 9B
mg/kg
80
Phosphorus Concentration
Olsen-extractable Phosphorus (Olsen-P) - 9C
mg/kg
23
Nitrate N Concentration
Nitrate procedure
mg/kg
61

Organic carbon concentration. Procedure: Walkley-Black. UoM: %



Federation University

SOiL CRC
Performance through collaboration

The Challenges, and some of our current solutions

Variety:

Varieties of dataset format .pdf reports, excel, local databases **and structure** e.g. excel layouts/tables

Various data vintage - legacy data, recent data, recent sampling campaigns, ongoing monitoring sites

Wide Variety of information from different soil domains - soil physicochemical, physical, biological, profile observations

Sampling regimes –

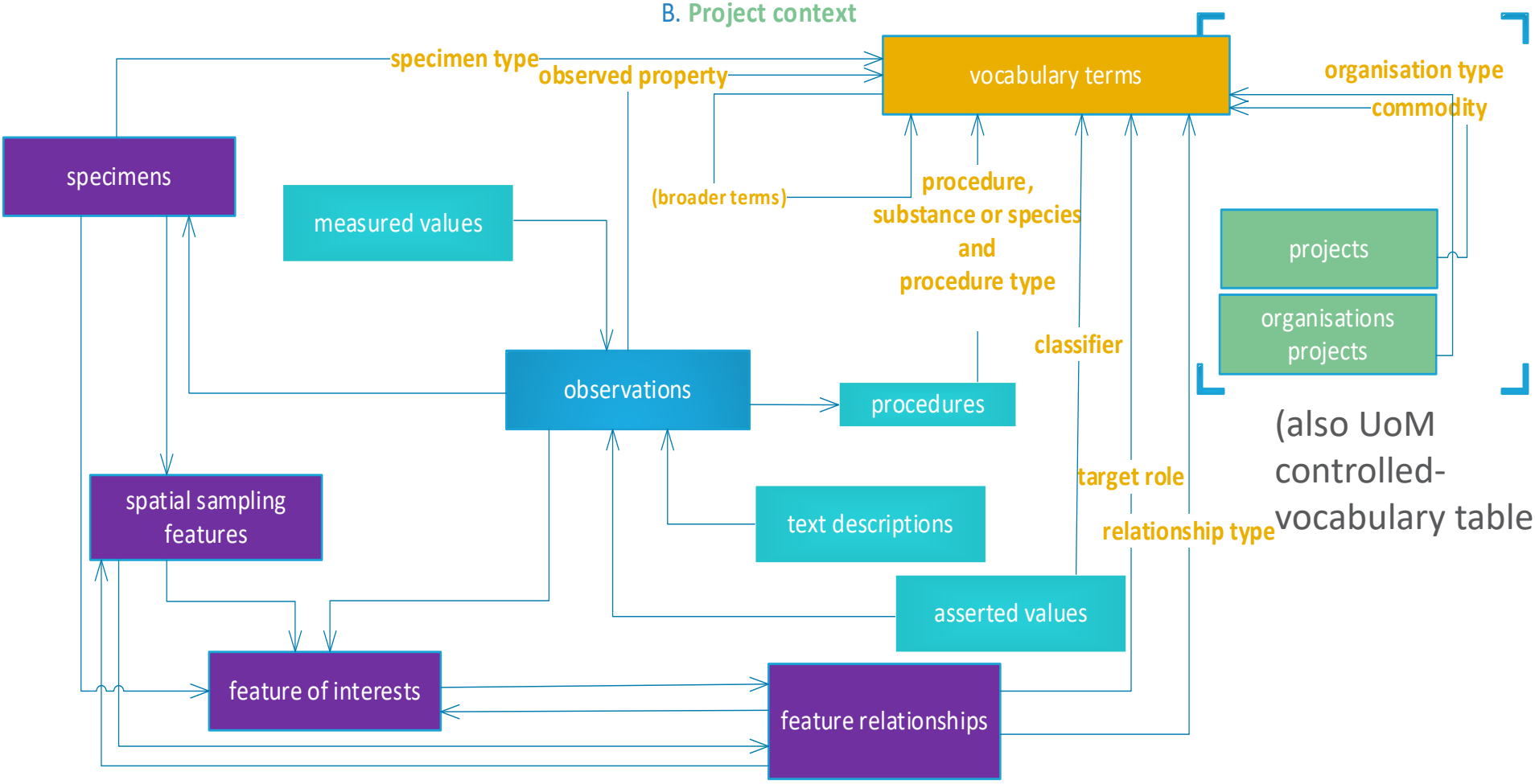
- Various **experimental designs and factors** - Documentation hard to interpret
- **Repeated measurements** over time
- Most sampling was soil layer at **depths** (e.g. 0–10cm, 10–20 cm, 20–30 cm) – no standard ranges
- Most often no detail on **sampling method** (e.g. disturbed, undisturbed, core diameter etc)
- Most often no great detail on **sampling regime** (e.g sub-sampling regime and geo-references for)
- No detail on **samples** provided aside from lab numbers (e.g sample type, pre-treatment, storage)

Part of solution: O&M design pattern, leveraging of controlled-vocabulary terms for harmonisation

Vocabularies are used to describe

A. From observation model: 1. **observations**, **measurements** made on **specimens** and **features** 2. **Feature relationships**

B. **Project context**



(also UoM controlled-vocabulary table)

SOiL CRC

Performance through collaboration

Challenge – Relationships and identity Mapping the observations within and between datasets

Needed to map observations within and between datasets – to compare data over time, space, within and between datasets

- **Identity issues** - Inconsistent naming of sites, samples and geo-references
- **Complex relationships** (temporal and spatial, within and between datasets) needing to be mapped
  - **Temporal** - Relationship sometimes specified (e.g. by site IDs), sometimes not
  - **Spatial** – e.g. soil layers to soil bodies to soil sites, relationship between paired sites
  - Difficult to map complex one to many relationships **using excel spreadsheet template**

Solution – Vocabularies for relationship types and roles, and database analysis

- Vocabularies used to describe spatial and temporal relationships at upload where possible or
- Database analysis of depth, spatial and context data

Federation University

SOiL CRC
Performance through collaboration

Challenge: Resolution of data to maximise re-use potential

Many related to **the procedures used to make observations** –

- Unknown, not included
- One dataset, multiple laboratories
- Historical data – laboratories change their methods over time, standards change over time
- Multiple rows of results for one observable property, with sometimes no procedures or UoM specified
- Inconsistent or ad-hoc calculations, from lab or provider? Eg CEC, eCEC

Solution – Map to highest resolution possible using controlled-vocabulary

- National standard procedure
- Citable, publicly available, published on-line
- Laboratory specific procedure, preferably published online
- Not ideal - Undocumented procedure

Challenge – Controlled-vocabulary creation, management and governance

- **Discoverability. Solution:** Begin resources such as OLS and AgroPotral, AGROVOC, AgroO, QUDT, INSPIRE terms….

- **Limited National and International terms** (or not easily discoverable!?) particularly for observed properties and procedures. **Solution** – working with National agency (CSIRO) to fill gaps.

- **Procedures, Non standard or poorly documented. Solution** – created controlled-vocabulary concepts with as much provided information as possible. *On-going governance and usefulness of these is problematic*

- **Ongoing maintenance and governance** – some vocabularies have been established without determining roles, rights and responsibilities to publish, maintain and govern. **Solution** – ongoing collaboration with authorities/eResearch community, updates to vocab metadata etc. *VAS project vocabulary collections require governance establishment.*

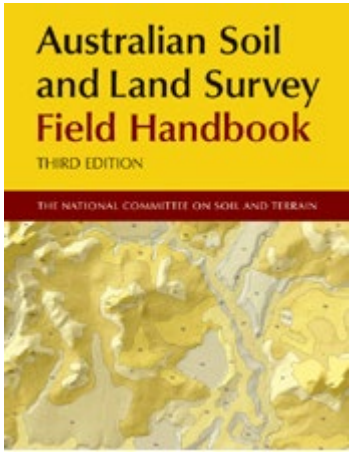
- **Upskilling** of research staff in semantics and semantic technologies. **Solution:** Begin with currently supported user interfaces with semantic expert support

Federation University

SOiL CRC
Performance through collaboration

Controlled-vocabularies in use

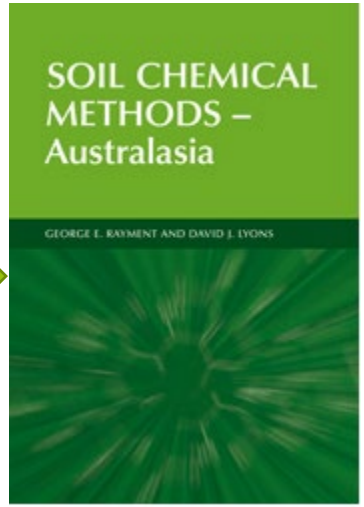
- **Re-use existing** controlled-vocabularies where they exist e.g. ENVO and AgrO Ontologies, Units of Measure (QUDT) and INSIPRE features
- **National standards** e.g. the 'green book' – soil chemistry procedures and the 'yellow book' – Soil Profile, land Surface, Landform, Substrate classifiers. *Much thanks for assistance and vocab work: Linda Gregory, Simon Cox (CSIRO)*
- **Aust/NZ soil related persistent identifiers** using **http://anzsoil.org .** Currently located in CSIRO Linked Data Registry http://registry.it.csiro.au/def/soil/au and see https://github.com/ANZSoilData
- **Last resort - New controlled-vocabulary terms** are created only if necessary (mostly for procedures) 'federation university' created.



Contents

Name	Notation
Australian soil and land survey field handbook codes	asls
edu	edu
Soil chemical methods - Australasia	scma

Federation University

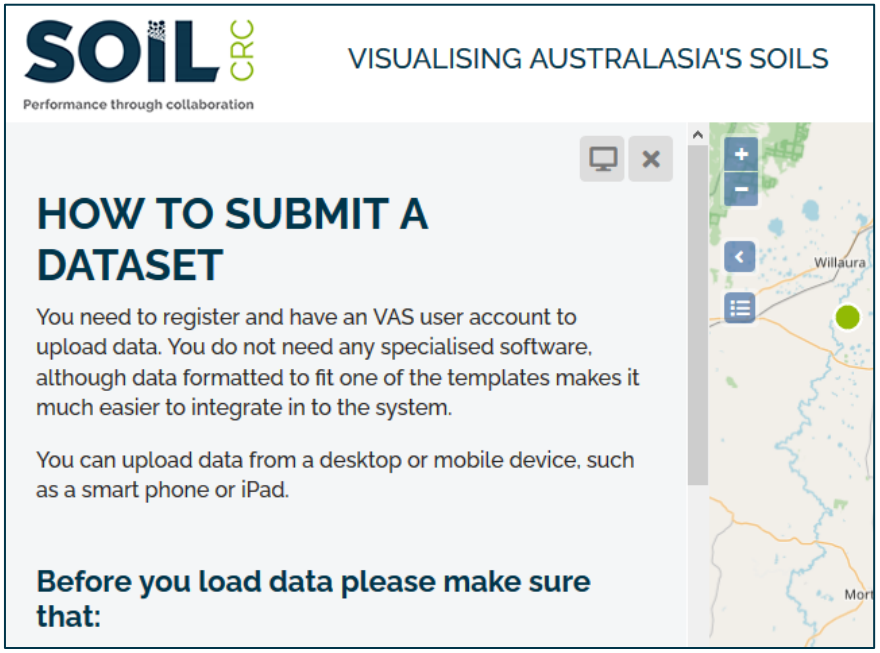


SOiL CRC
Performance through collaboration

Ongoing work and considerations -

Ongoing continued outreach and communications, Pilot self-serve interface with metadata questionnaire/DB integration, spreadsheet, licencing education. To help address -

Lack of basic metadata for FAIR – generally not received from data providers
Trust and access requirements
Transaction cost – data description for ingestion/re-usability – currently on service provider/research staff,



Data quality and provenance considerations

Continuing to strengthen value proposition

Solution: Phase 2 foci 2021 – 2024:
Platform usage Education
Engagement re value added data re-use tooling

Federation University

SOiL CRC
Performance through collaboration


Thank you ESIP
we look forward to the conversation!
SOIL CRC
Performance through collaboration