What resources are available Internationally and what are the new drivers?

Lesley Wyborn[1, 2, 3]

[1]*National Computational Infrastructure, Australian National University,*

[2]*Australian National Data Service;*

[3]*AuScope,*

ED 32B-03

Contact: lesley.wyborn@anu.edu.au

Lesley is.....

- Chair of the Academy in Science, Data in Science Committee
- On the AGU Data Management Advisory Board
- On the Steering Committee of the ICSU/CODATA Commission on Scientific Standards for Integration of Research Data for Transdisciplinary Science
- On the Steering Committee for the AGU/Research Data Alliance Project for FAIR and Open Data in Research Publications
- On the AuScope Strategic Reference Panel
- AGU Meetings Committee

lesley.wyborn@anu.edu.au

Topics

- FAIR Publishing in the Earth and Space Sciences
- The Linked Open Publication
- IGSN
- AuScope Virtual Environment

lesley.wyborn@anu.edu.au

The AGU-led FAIR Data Publishing Project

- Funded by the Laura and John Arnold Foundation (LJAF)
- LJAF makes strategic investments in criminal justice, education, evidence-based policy, public accountability, and **research integrity**
- Aims to develop a collaborative solution across researchers, journals and repositories that will evolve the Earth and Space Science (ESS) publication process to include all research inputs into a publication to help develop a unified process that supports researchers from grant application through to publishing.

The Problem…

Optional

At the time of publication:

- Are the data (samples, software/services) that support the paper properly documented and stored in a persistent, sustainable repository?

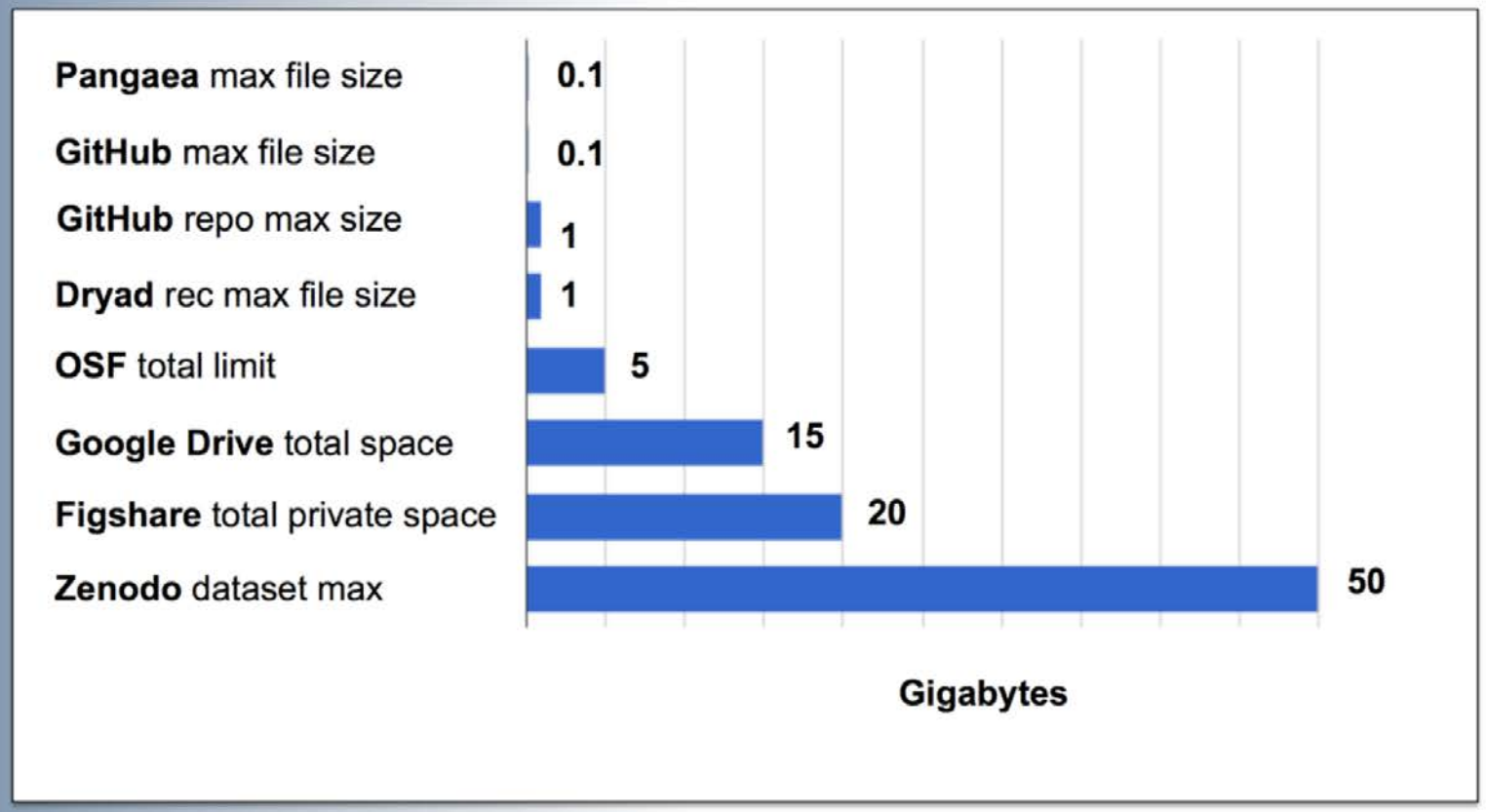
Optional

- Are the data (samples, software/services) citable with a persistent identifier (e.g., DOI), and support the Nature FAIR Guiding Principles for scientific Data management and Stewardship (https://www.nature.com/articles/sdata201618)

No

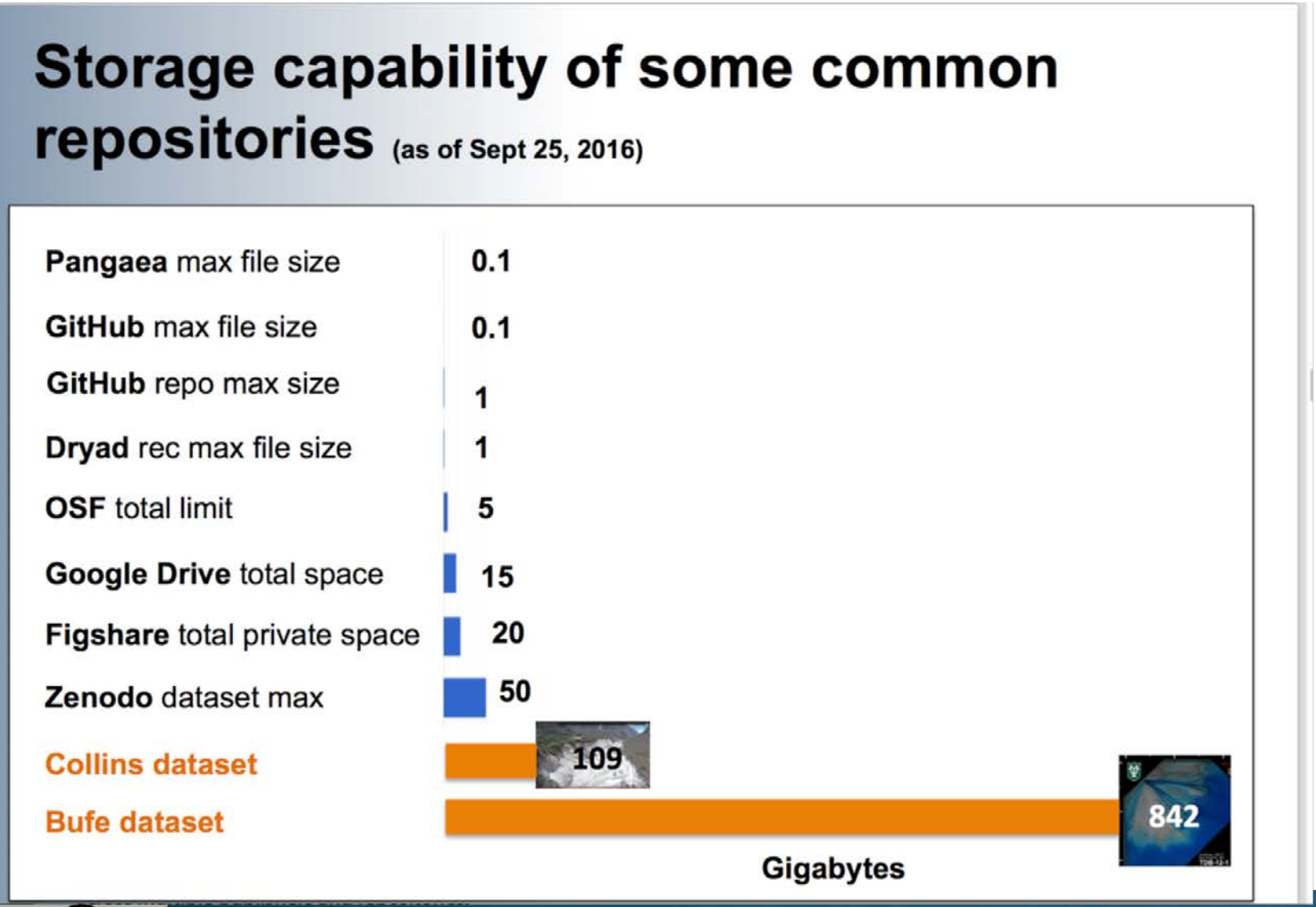
- Do researchers have a similar experience with submitting their paper and supporting data (and software/services) no matter the journal?

Emerging issue: storage capacity of publisher repositories



Slide Courtesy of Dr Leslie Hsu, USGS

Reality of the size of geoscience data sets



Slide Courtesy of Dr Leslie Hsu, USGS

The Solution (1): Proper data documentation and storage

- In support of a publication…there is proper **data documentation and it is** stored in and accessible from a trustworthy repository.
- Need to:
  - require data be Accessible from a repository with the paper as the default option (No data as Supplements to papers)
  - Need to engage domain repositories to ensure proper curation as much as possible

The Solution (2): Proper data citation with a persistent identifier

- In support of a publication…proper **data citation with a persistent identifier** supporting the FAIR Guidelines.
  - Need to require use of repositories that use persistent identifiers.
  - Need to require use of repositories that have landing pages that support citation.
  - Need to require use of repositories that can support community agreed standards

lesley.wyborn@anu.edu.au

The Solution (3)
Provide a similar experience for a researchers

- In support of a publication…provide a **similar experience for a researcher** when submitting their paper and supporting data (and software/services) no matter the journal.
  - Journals and repositories need to define and adopt recommendations and align policies.

Partners

- Earth Science Information Partners
- Research Data-Alliance
- *Science*
- *Nature* Research
- *PNAS* (NAS)
- Center for Open Science (COS)
- COPDESS
- **AuScope**
- **National Computational Infrastructure, ANU**
- **Australian National Data Service (ANDS)**

Focus is on the Earth and Space science journals, repositories, and researchers.

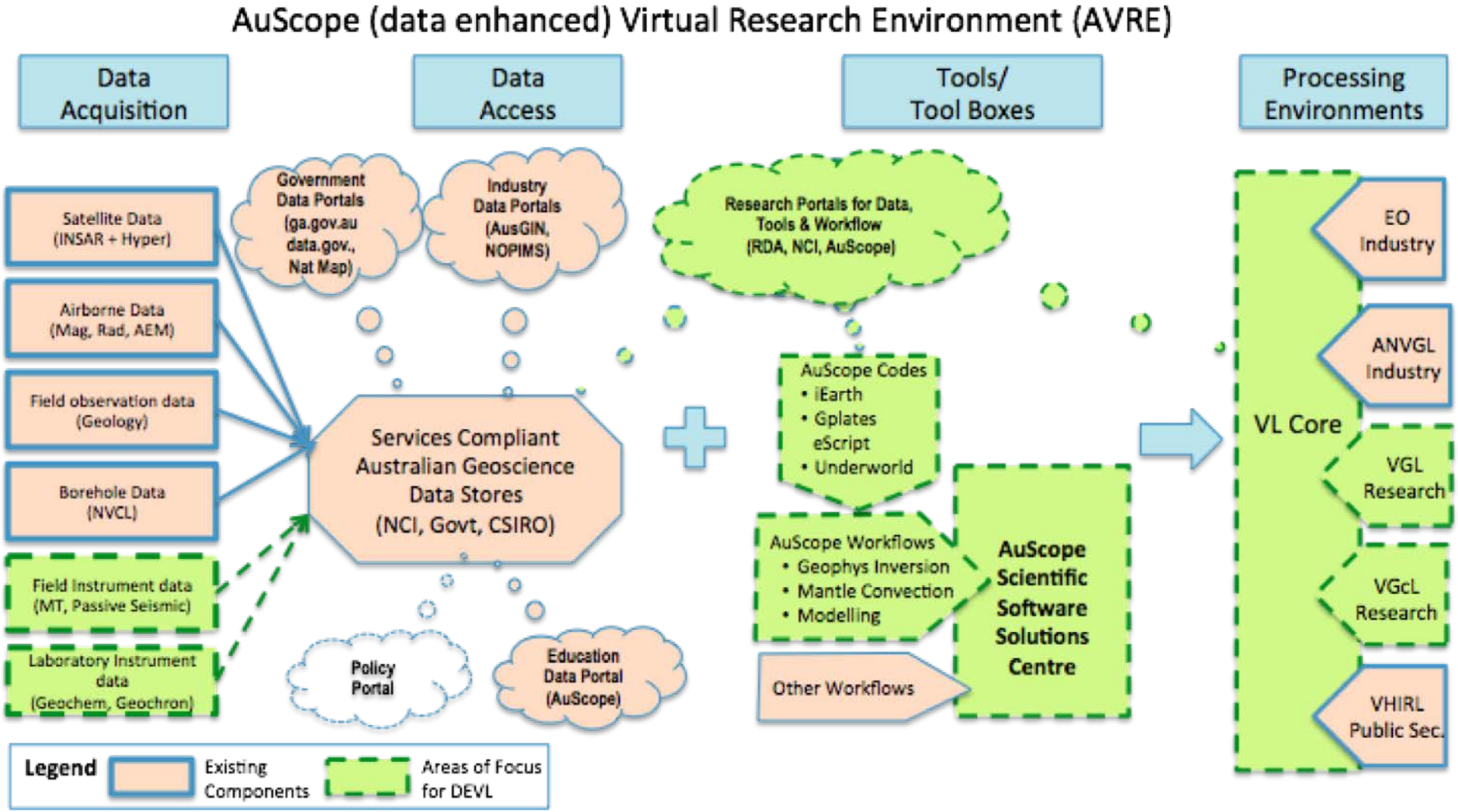
Data standards required for Publications in Earth Sciences

- Across Earth and Environmental areas they are patchy
- More common in Geophysics
- For Machine Learning/Deep Learning we need more consistent and agreed vocabularies, semantics and ontologies?
- Enter the ICSU/CODATA Commission on Scientific Standards for Integration of Research Data for Transdisciplinary Science

But internationally there is movement at the station For Geochemistry

- Growing attempts on standards for geochemistry
- Certified domain repositories are coming on line (e.g., NSF IEDA)
- Proposed online registry of reference materials for isotopes
  - Proposed meeting at Goldschmidt
- US and Germany are combining to share developments and develop standards
- Will be done through International Geochemistry Society
  - Proposed splinter meeting at EGU Vienna

Australian Virtual Research Environments



lesley.wyborn@anu.edu.au

And we just got $425,000 from NCRIS

- National **Collaborative** Research Infrastructure Strategy
- AuScope Geosciences DeVL (Data-enhanced Virtual Laboratory
  - IGSN identifiers for the Australian ~~Geoscience~~ NCRIS Research community
  - A FAIR MagntoTelluric (MT) repository for research data at NCI
  - Enable a software registry that indexes code in GitHub repositories
  - Revamp the AuScope Virtual Laboratory infrastructure to enable it to make connection to more geochemistry data sets and more software – i.e. support the Virtual GeoChemistry Laboratory (VGCL)

What is a Virtual Research Environment (VRE)?

A Virtual Research Environment, Science Gateway or Virtual Laboratory is an on-line system supporting collaborative research that enables harnessing of the power of the Internet to support a more dynamic, online approach to collaborative working

Key features:

- Provide access to data resources that are accessible online
- Enable online use of discipline-specific tools, such as data analysis, visualisation, or simulation management
- Online access to compute resources
- Collaboration support (Web forums and Wikis)
- May include publication management and teaching tools

VRE's are important in fields where research is primarily carried out in teams spanning multiple institutions and even countries



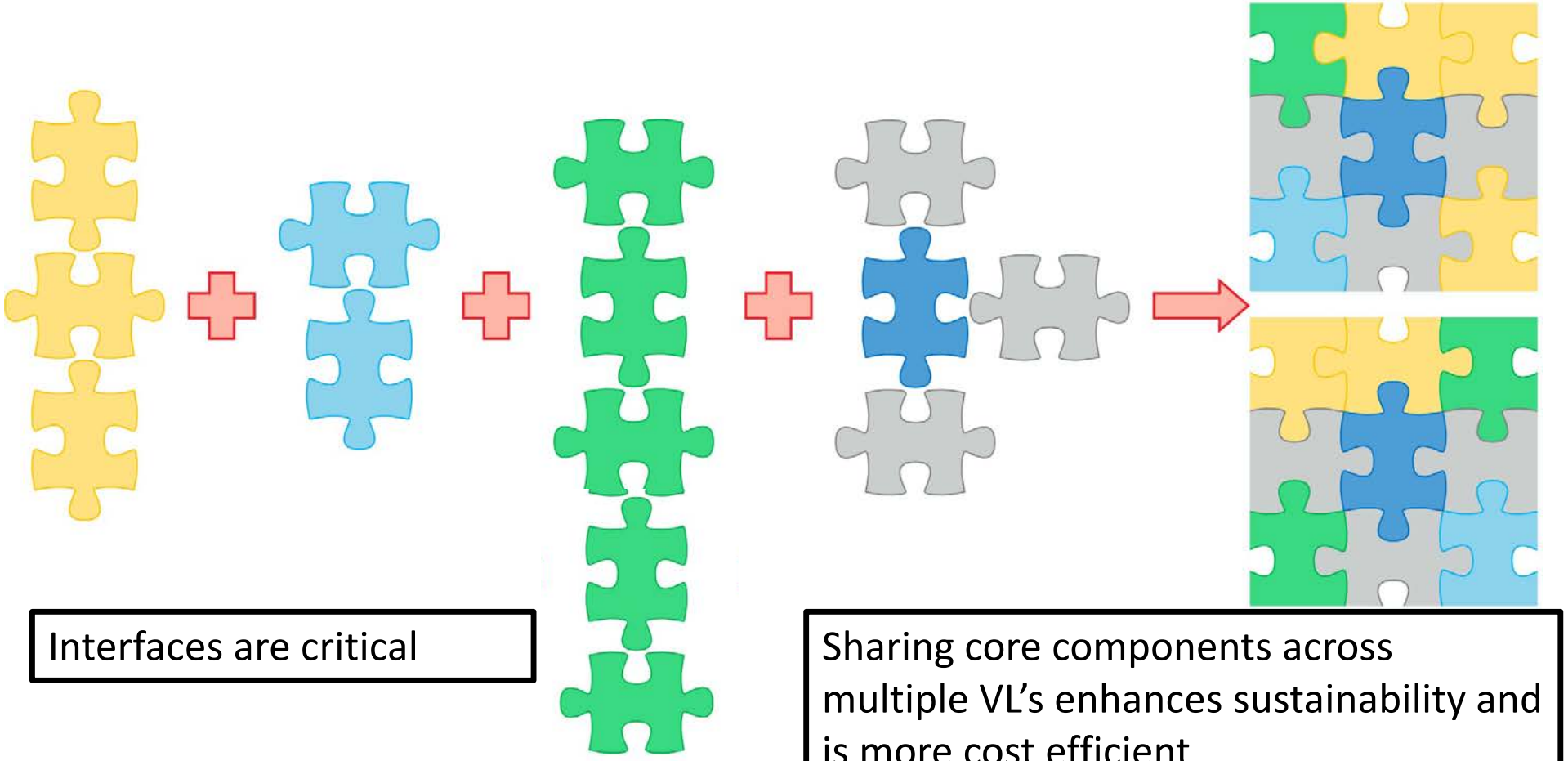


lesley.wyborn@anu.edu.au

Elements of A Virtual Research Environment

- Text goes here

Data Services + **Processing Services** + **Compute Services** + **Enablers (eg. OGC "Glue")** → **Virtual Laboratory**

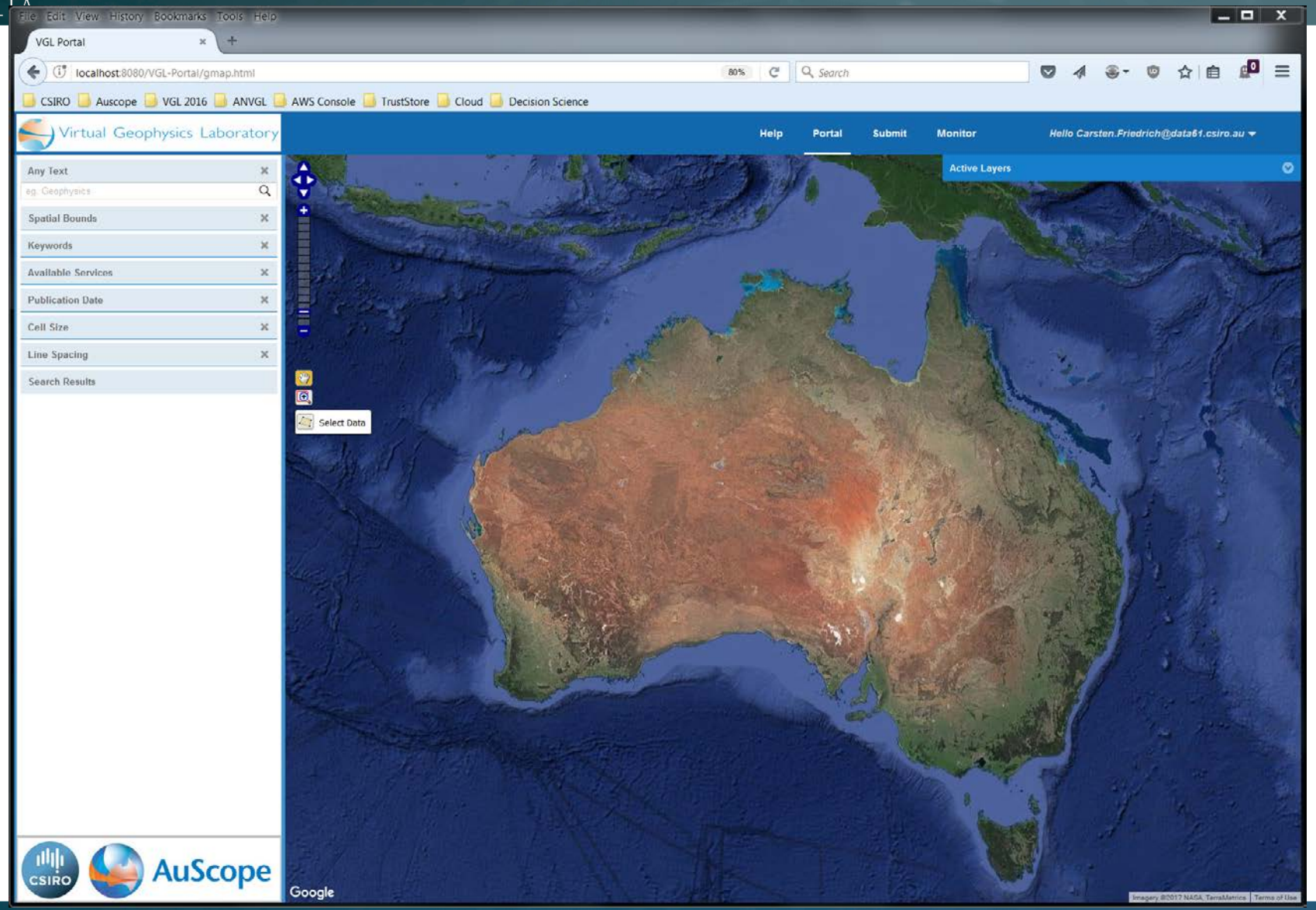


Interfaces are critical

Sharing core components across multiple VL's enhances sustainability and is more cost efficient

lesley.wyborn@anu.edu.au

Introducing the Virtual Geophysics Laboratory



lesley.wyborn@anu.edu.au

Data Select Screen

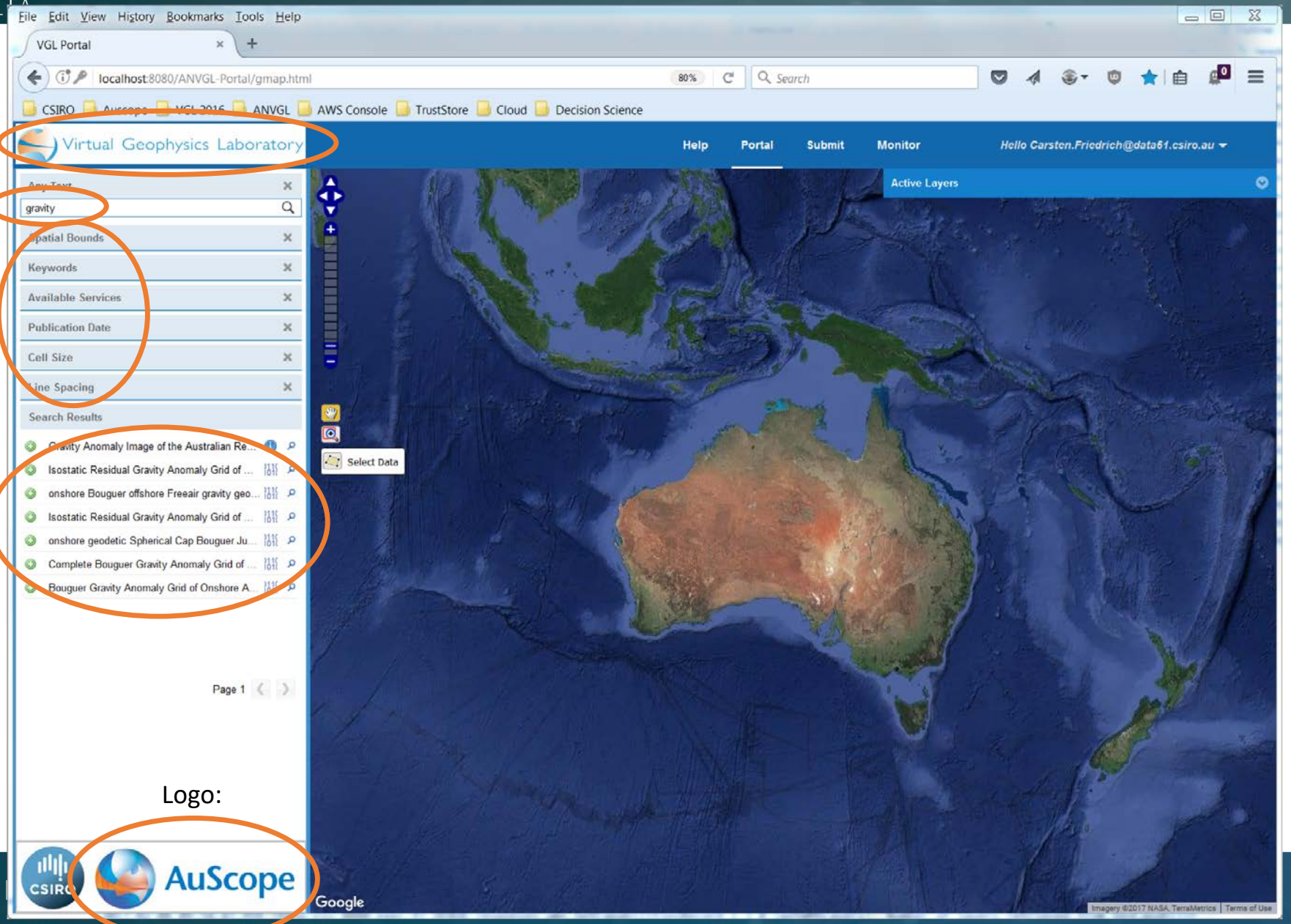
Logo:

Search term:

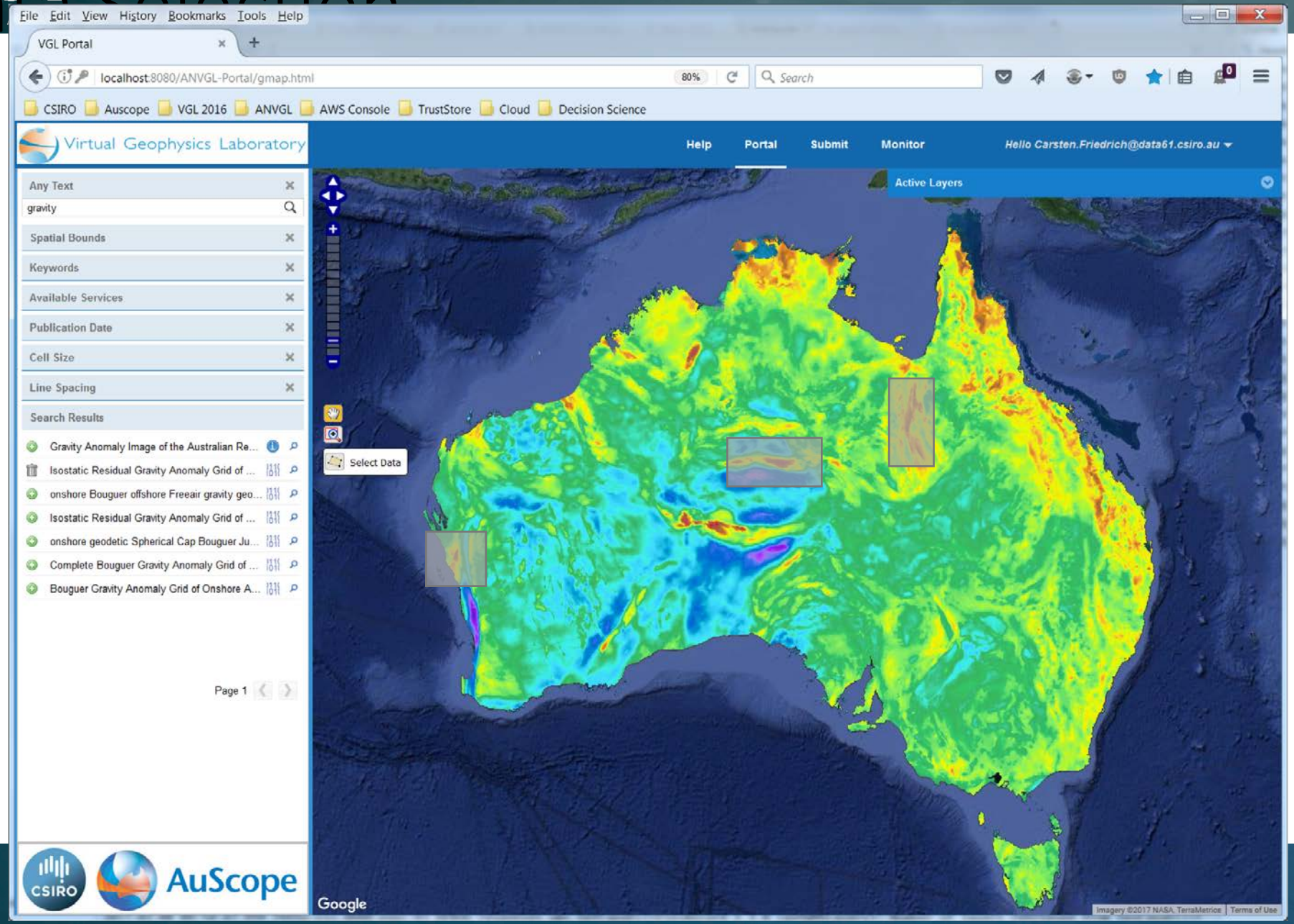
Search facets:

Result datasets:

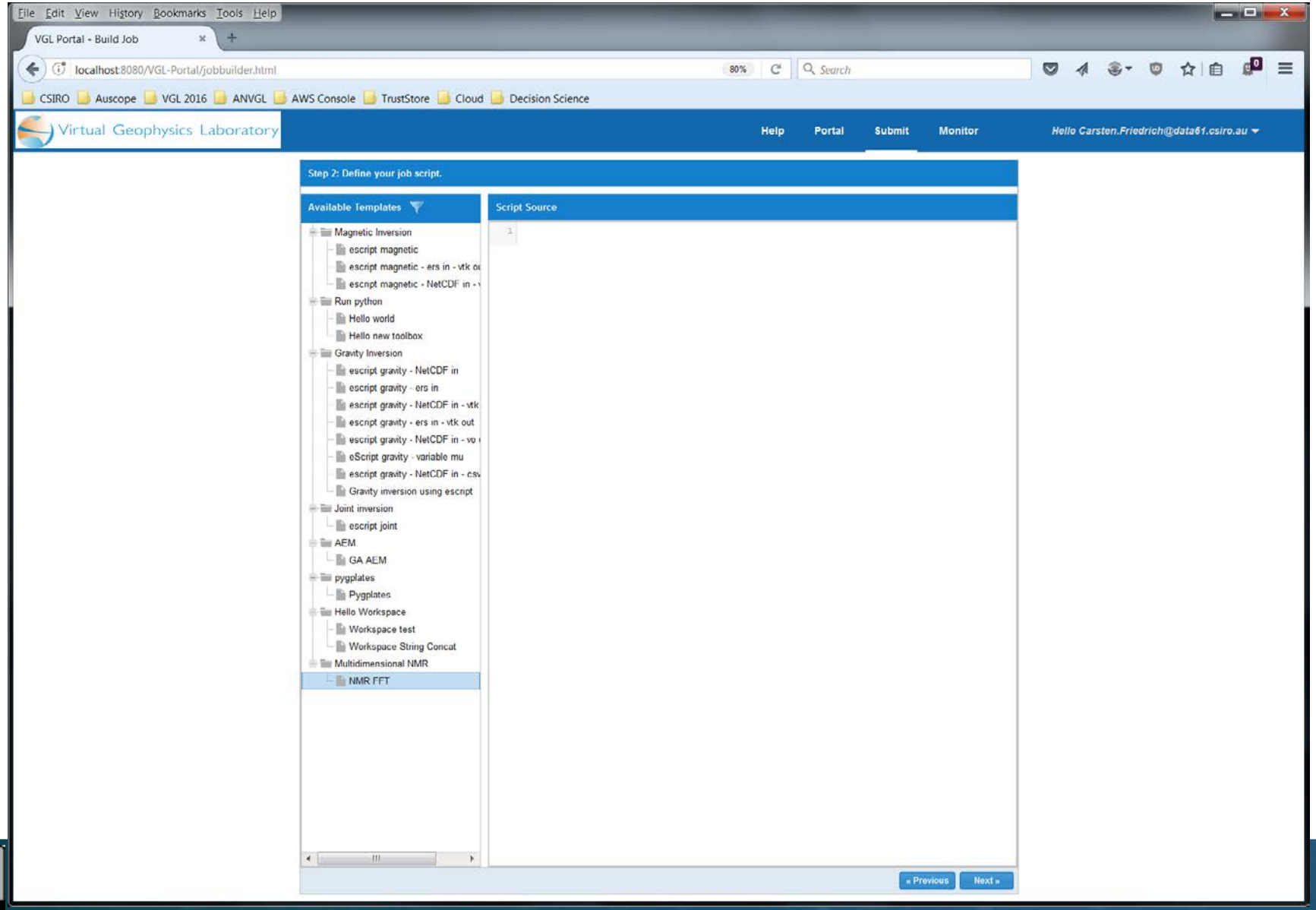
Logo:



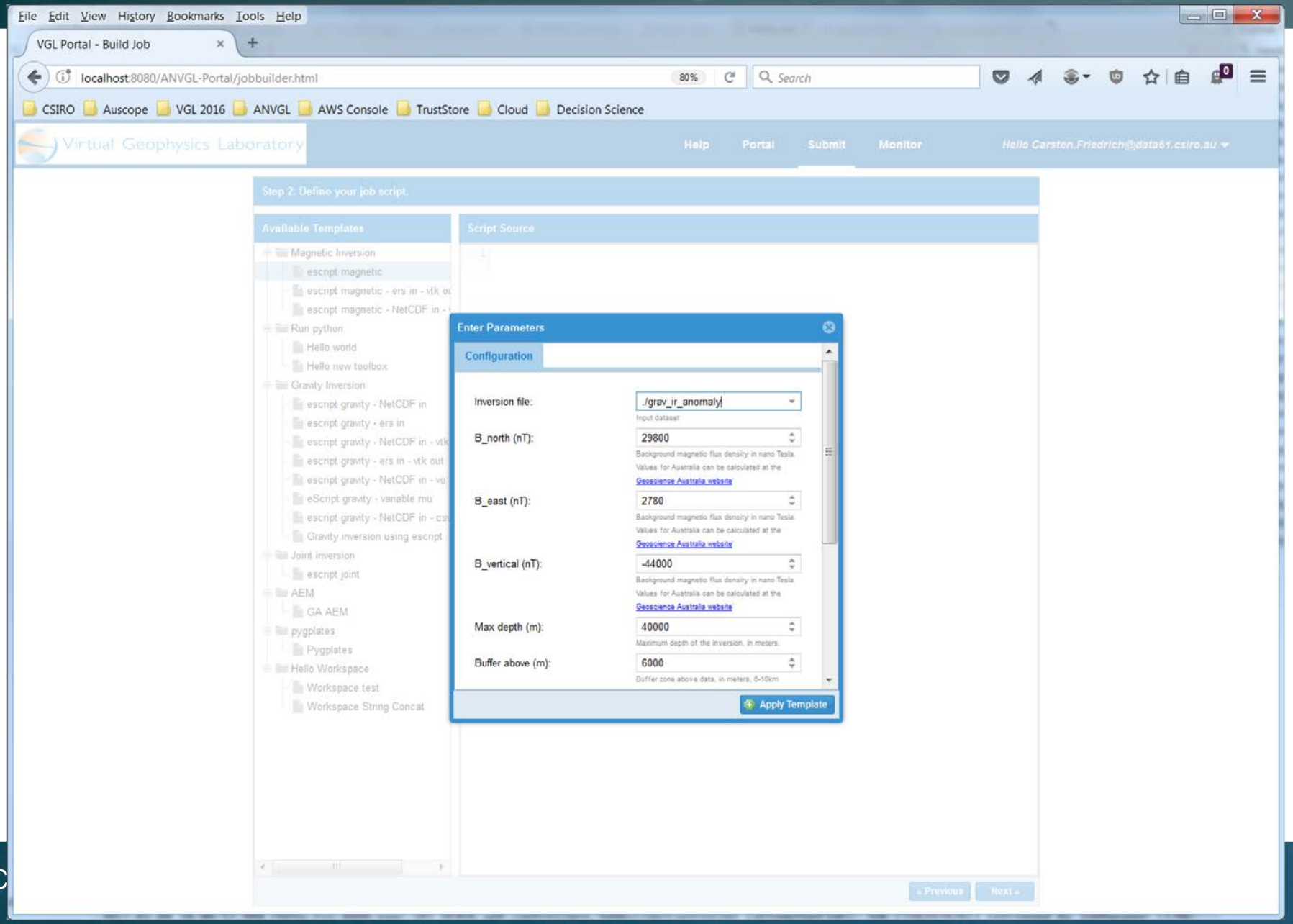
Data Selection



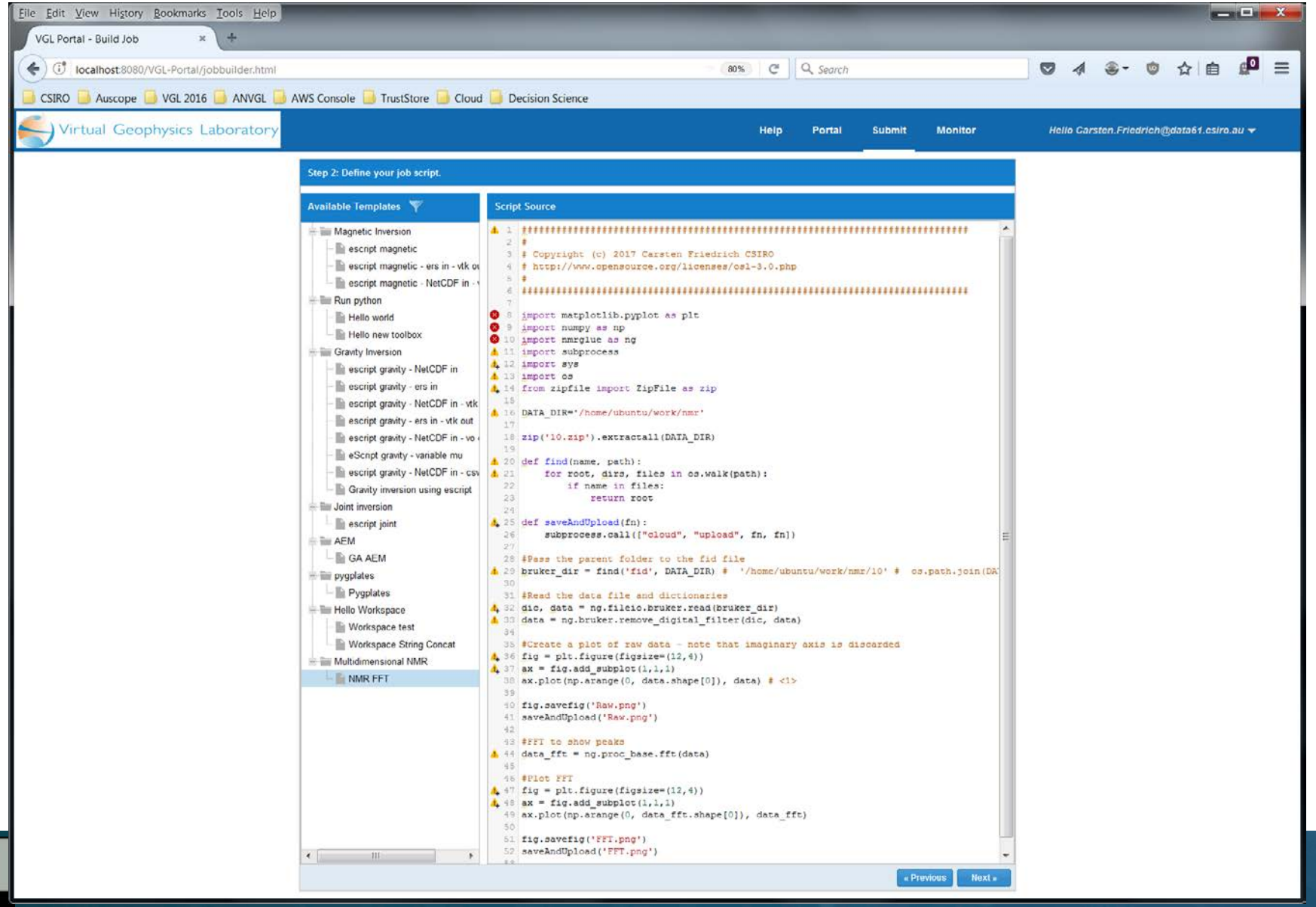
Browsing available science codes



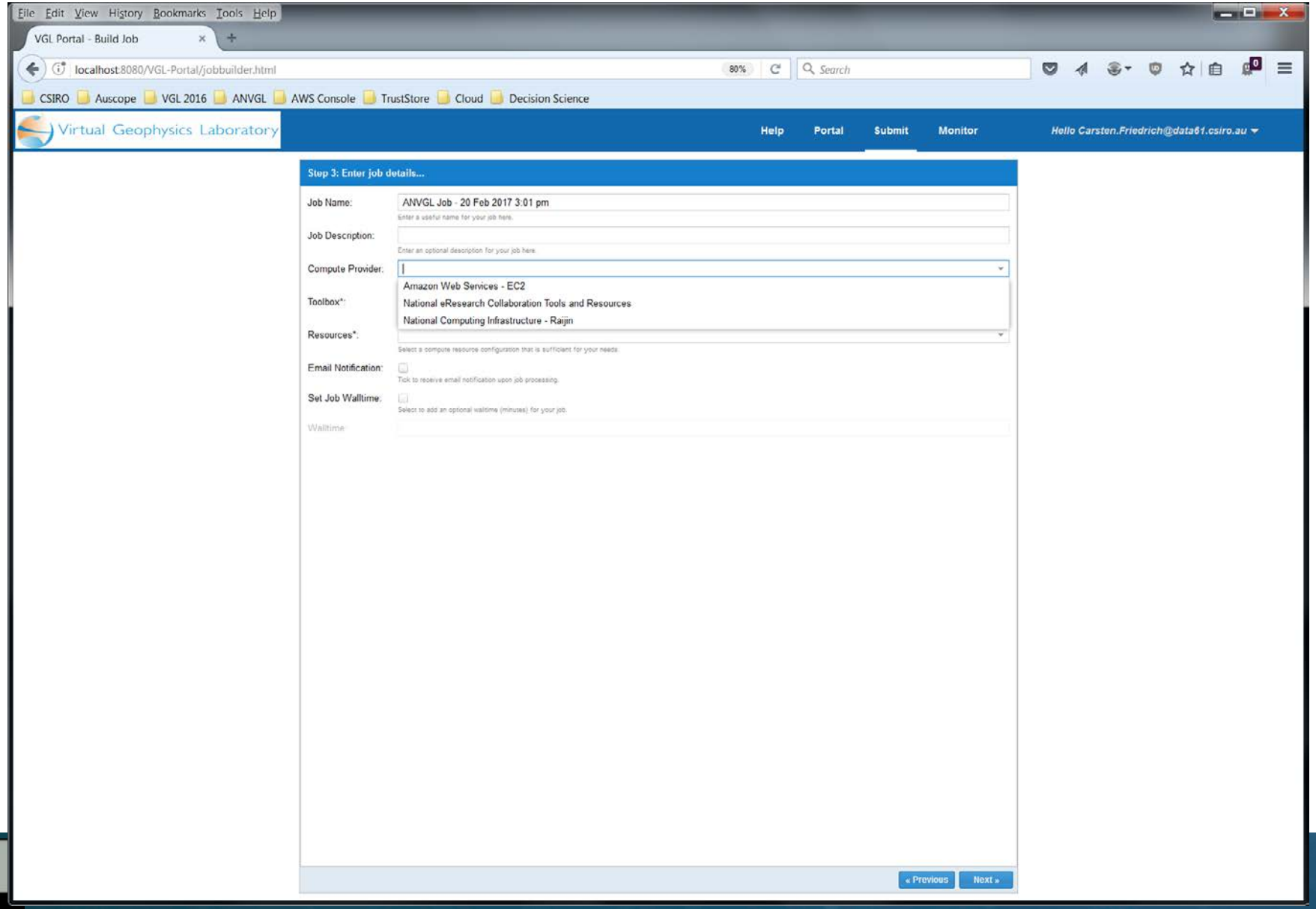
Task Parameterization



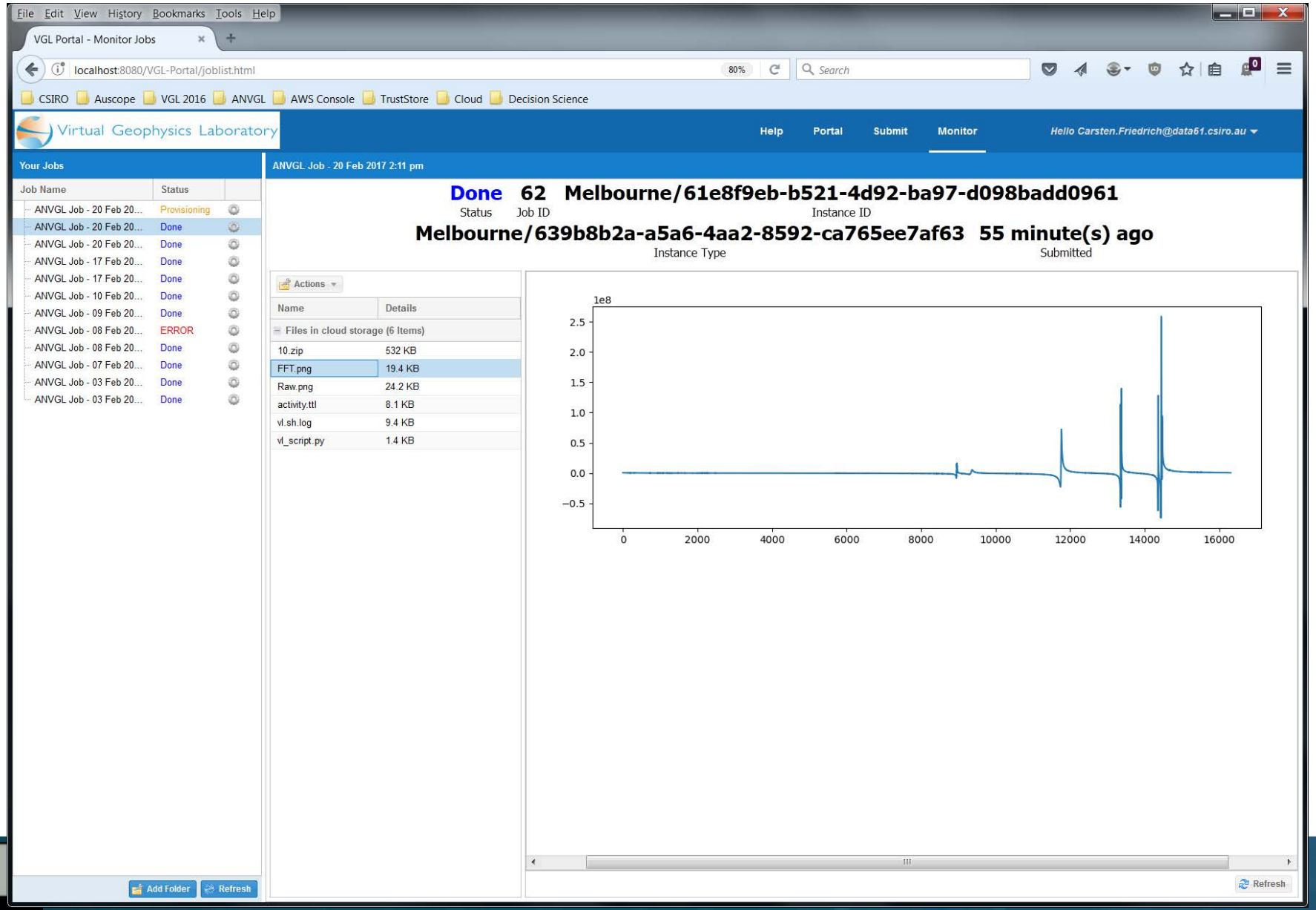
Allowing power users to make last minute adjustments



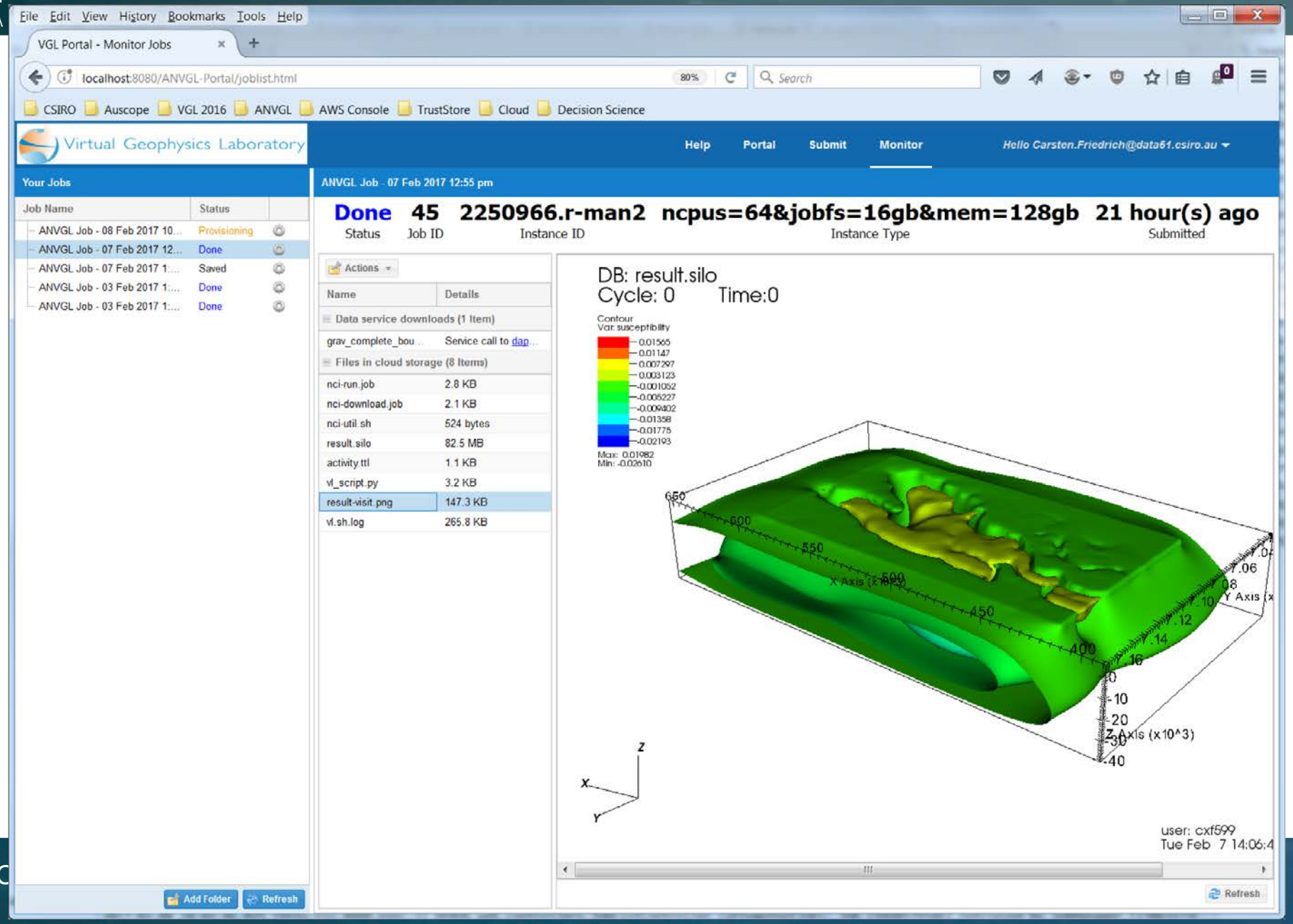
Selecting compute platform and resources



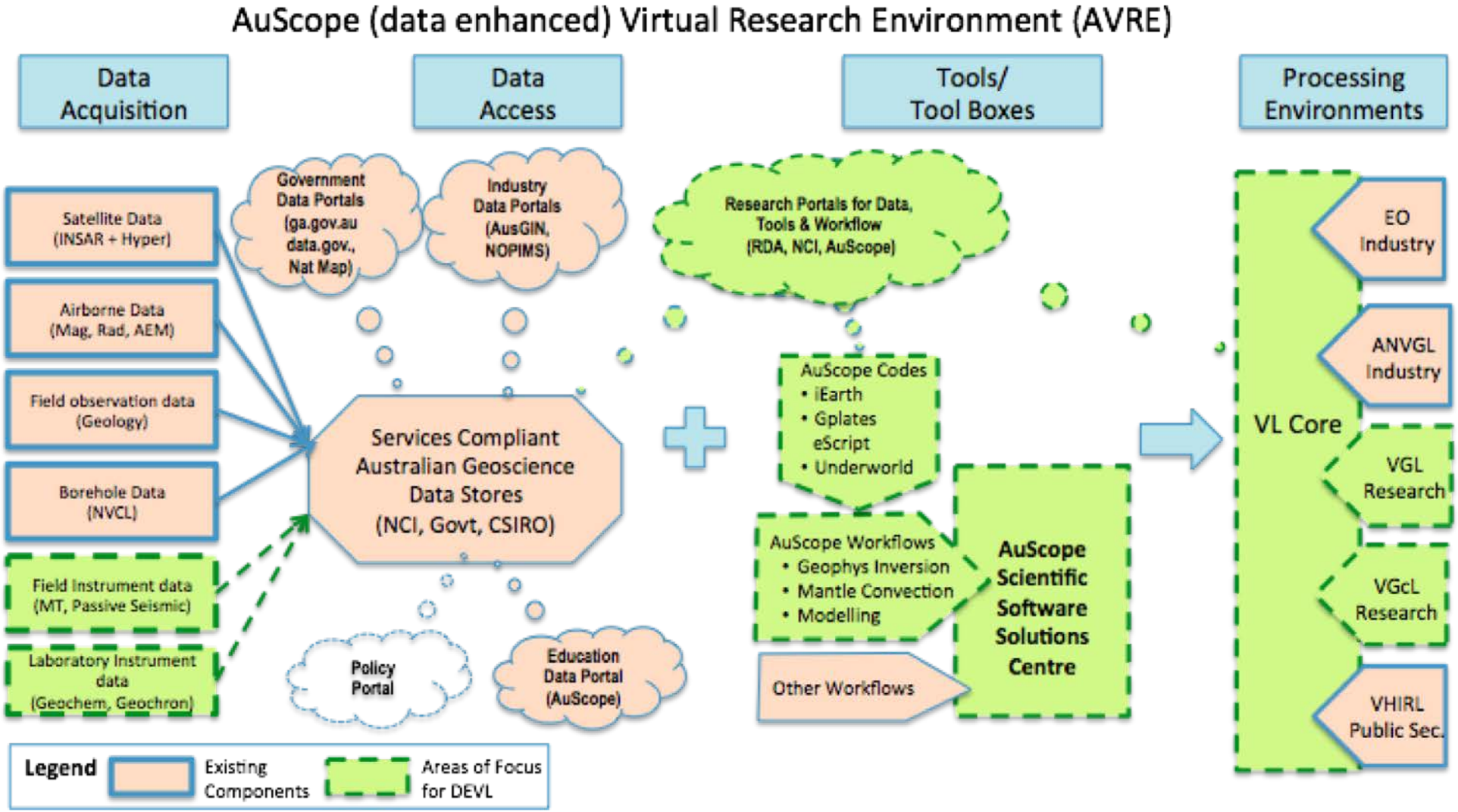
Monitoring progress and retrieving results



Task Results

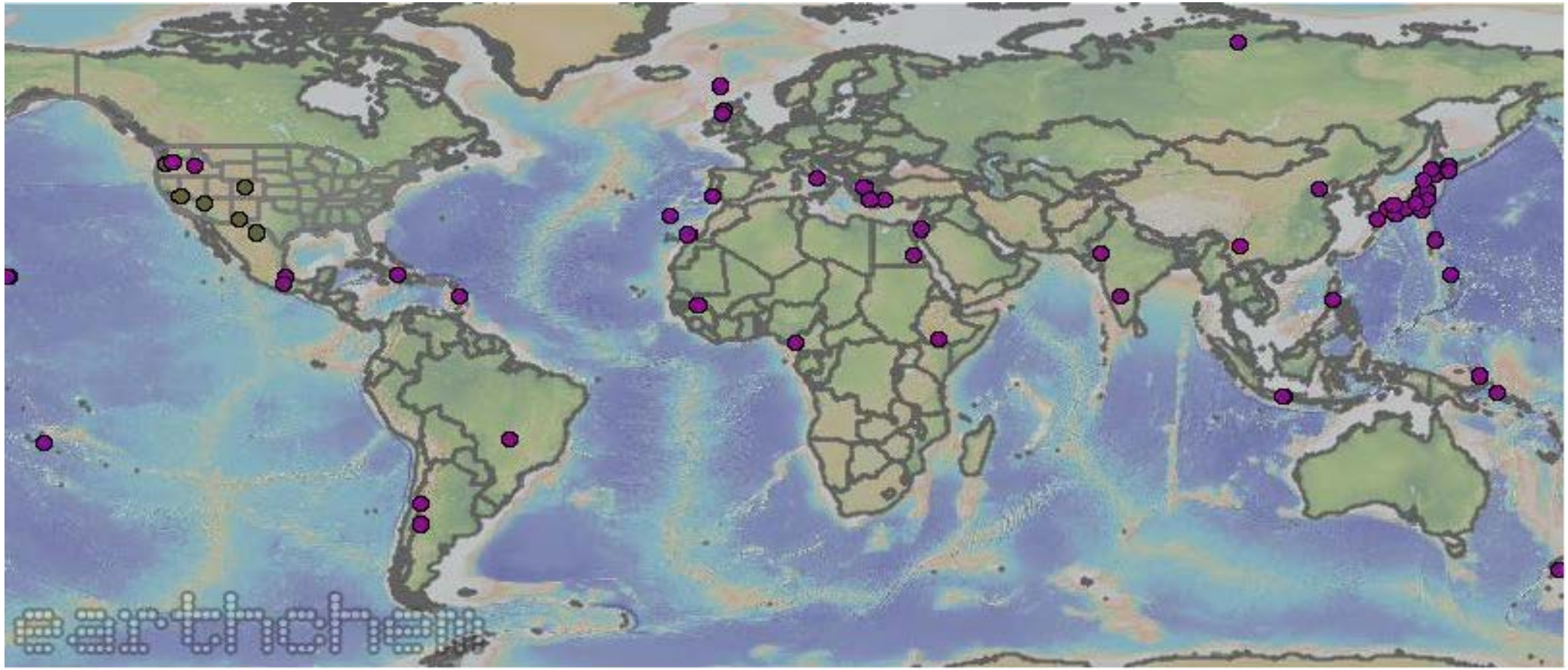


Australian Virtual Research Environments



lesley.wyborn@anu.edu.au

What is the IGSN?



"M-1" is obviously not a unique name.

lesley.wyborn@anu.edu.au

What Does IGSN Offer?

- A registry of identifiers that are guaranteed to be **globally unique** and **persistent**.
- It facilitates **online discovery and access** to physical samples
  - Web application and programmatic access to sample **metadata catalogues**
  - **Network** with sample repositories and data centres
- Ensures **preservation and access** of sample metadata (but not samples)
- Aids in the identification of samples in the literature
- Currently has **6 million IGSN's** registered from agents in 8 countries

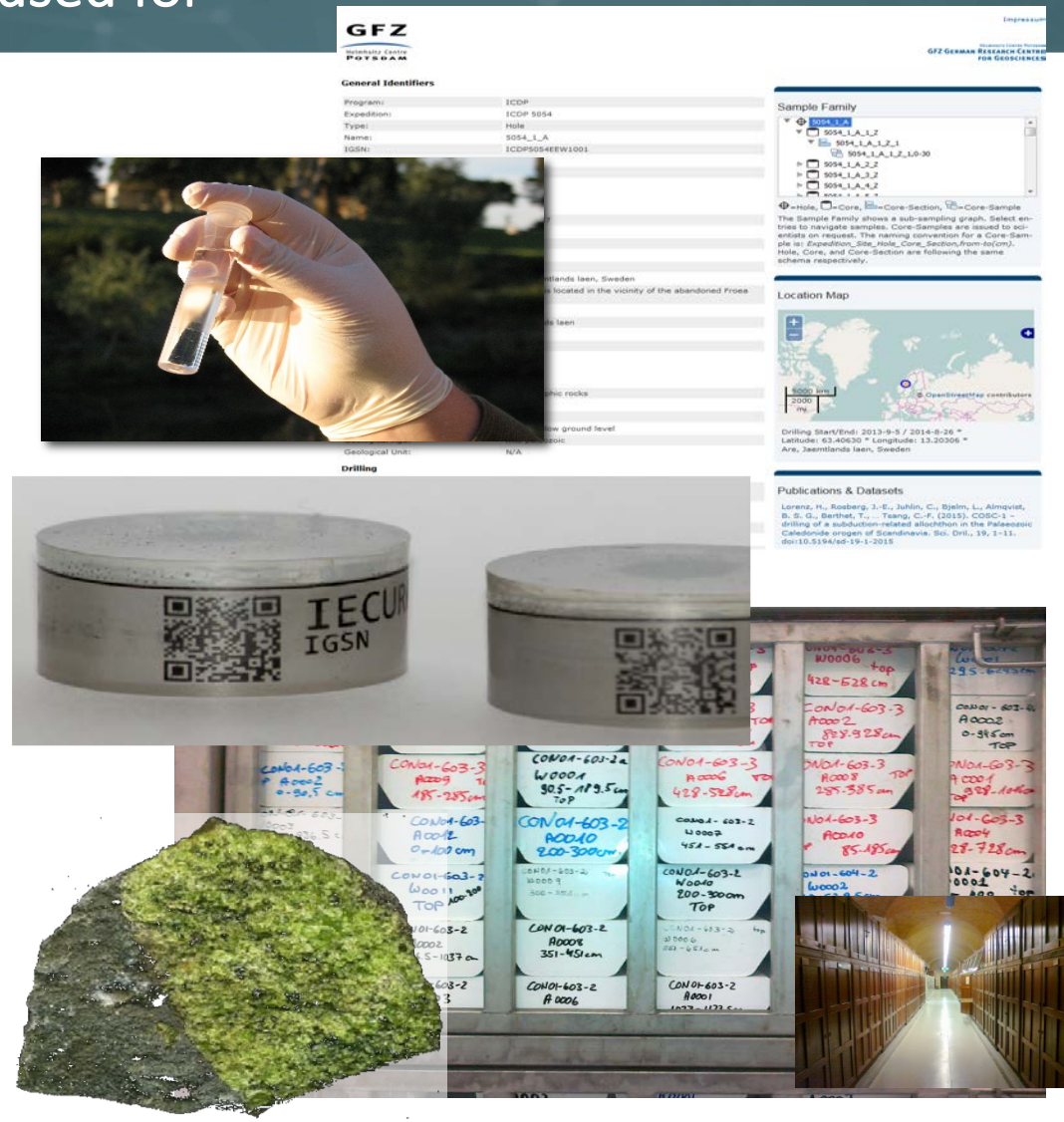
Try it out: http://doi.org/10273/AU288

lesley.wyborn@anu.edu.au

What IGSN can be used for

- Geological samples and other materials (rocks, water, biological materials, …)
- Collections (groupings of samples)
- Sampling features (boreholes, outcrops, …)

- Samples can be linked to each other through the “related identifier” metadata element.



Where to next for IGSN in Australia

- The three current IGSN members in Australia:
  - CSIRO – Government Research Organisation
  - Geoscience Australia – Government Geological Surveys
  - Australian National Data Service – Research Sector

- Involvement with Australian National Data Service
  - Help provide infrastructure for the research community
  - Currently provides a catalogue for data, software
  - Mints DOI's
  - Assisted with community building (e.g. workshops) to introduce sample identifiers to other institutions and other science disciplines
  - In 2018 will be working with Curtin, Macquarie and Melbourne Universities

Lesley is.....

- Chair of the Academy in Science, Data in Science Committee
- On the AGU Data Management Advisory Board
- On the Steering Committee of the ICSU/CODATA Commission on Scientific Standards for Integration of Research Data
- On the Steering Committee for the AGU/Research Data Alliance Project for FAIR and Open Data in Research Publications
- On the AuScope Strategic Reference Panel

lesley.wyborn@anu.edu.au