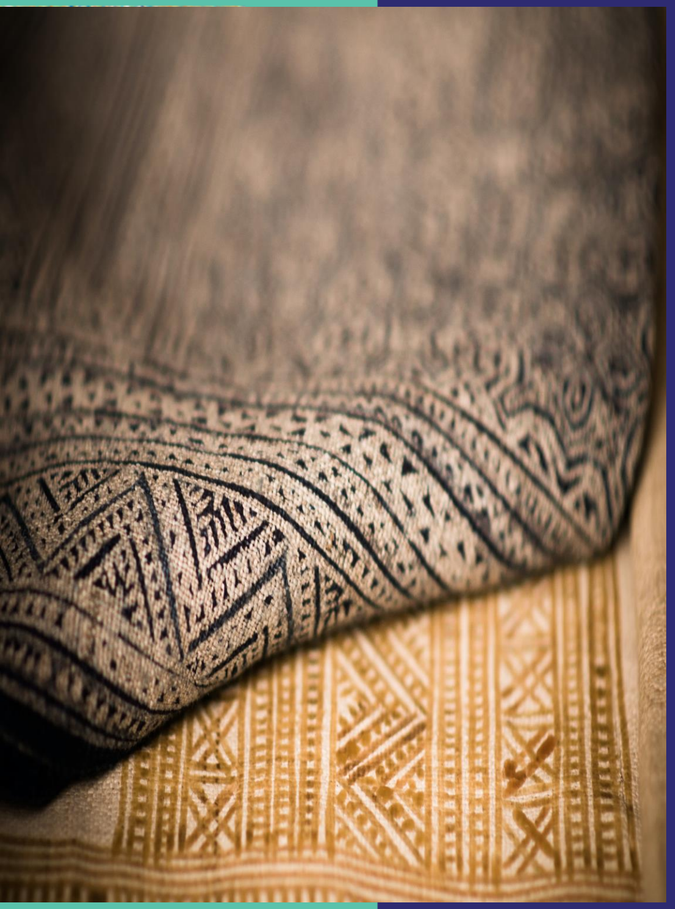
AuScope

Answering Australia's Geoscience Questions



Source: https://unsplash.com/@trang3019

Rolling back the carpet to reveal the challenges of improving FAIR compliance of the AuScope Data Infrastructure

Vincent Fazio, CSIRO Mineral Resources
Jens Klump, CSIRO Mineral Resources
Lesley Wyborn, ANU

eResearch Australasia 2022, Brisbane, 20th October 2022

NCRIS National Research Infrastructure for Australia
An Australian Government Initiative

What is FAIR?

Guiding principles that aid in knowledge discovery
Helping users of scientific data, both humans and machines (e.g., workflows, algorithms)

Improving discovery, access, integration and analysis

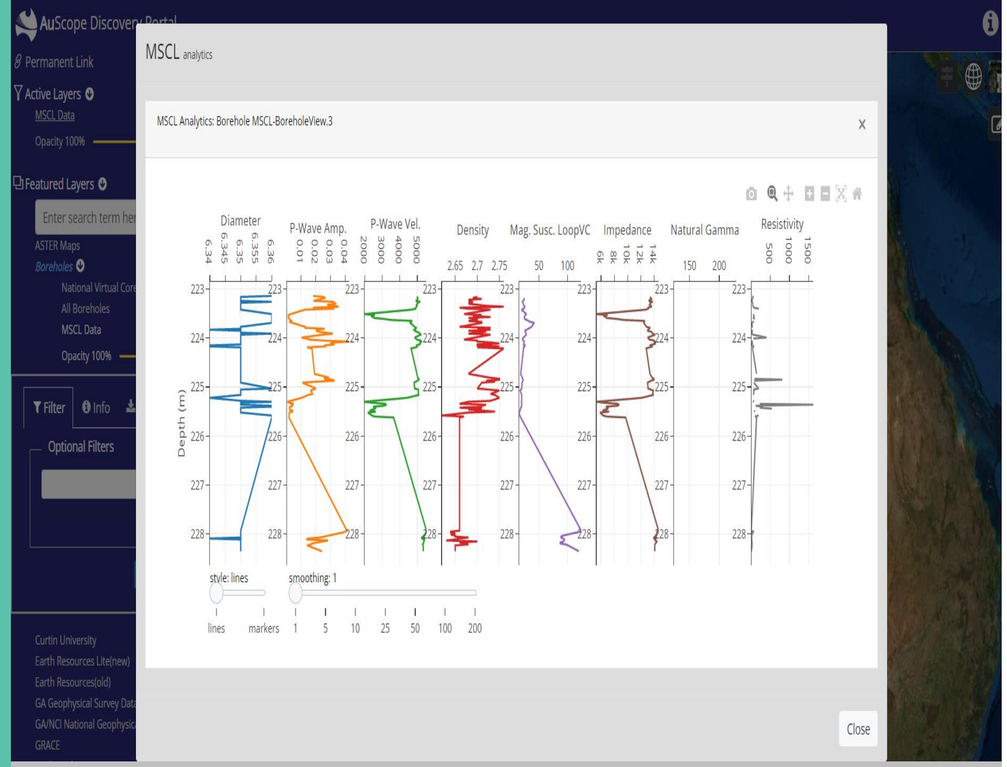
Four facets of FAIR: Findable, Accessible, Interoperable, Reusable

https://force11.org/info/the-fair-data-principles/

https://www.go-fair.org/fair-principles/

Journal Article
https://www.nature.com/articles/sdata201618

AuScope Discovery Portal



https://www.auscope.org.au/

https://www.auscope.org.au/avre

http://portal.auscope.org.au/

AuScope Discovery Portal

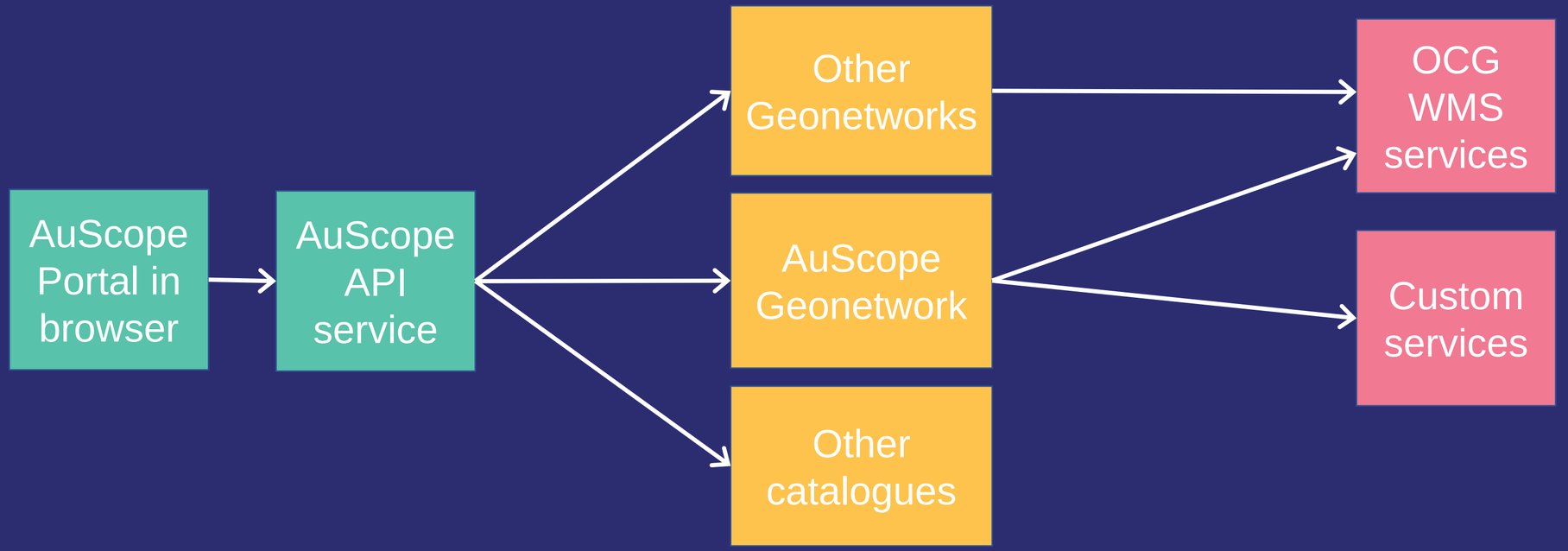
- AuScope – Australia's premier provider of research tools, data, analytics and support to Australia's geoscience community
- AuScope Discovery Portal is part of AuScope's AVRE (Australian Virtual Research Environment)
- Explore, display and download Australian science data
- Gravity, magnetics, seismic, borehole mineralogy, satellite, published reports …

Work Areas of FAIR Conformity

01 End Consumer

02 Metadata Catalogue

03 Dataset Provider



Area 01: End Consumer

Improving FAIRness for End Consumers

“F2. data are described with rich metadata”[1]

“R1. (meta)data have a plurality of accurate and relevant attributes.”[1]

. . .

[1]https://www.nature.com/articles/sdata201618

Area 01: End Consumer

AuScope Portal – need to upgrade from old metadata standard

ISO19115-3:2018[1] XML standard has many improvements over old ISO19139:2007
e.g. specific licensing types, project sponsor

There are software development implications of upgrading metadata standard & increasing FAIRness:

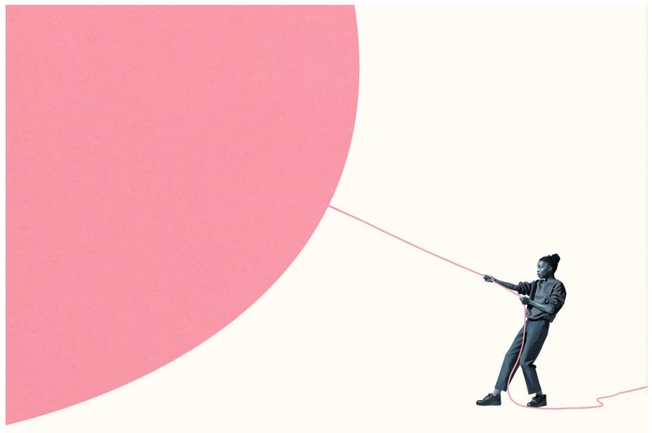
- User interface to display new richer metadata fields in a more prominent position
- Update internal data structures
- Update communication paths between GeoNetwork[2] and API service

[1] https://www.geonetwork-opensource.org/manuals/trunk/eng/users/annexes/standards/iso19115-3.2018.html
[2] https://www.geonetwork-opensource.org/

Area 03: End Consumer

AuScope Portal – an internal tug of war

- Previous trend was to migrate functions from back-end to browser, as browsers became more capable
- Desire to support APIs & Python notebooks requires careful software design
- New metadata features must be fully supported from the back-end API



Area 02: Metadata Catalogue

Upgrading and Managing Geonetwork Instances

“F2. data are described with rich metadata”[1]

“F4. (meta)data are registered or indexed in a searchable resource.”[1]

“A1. (meta)data are retrievable by their identifier using a standardized communications protocol”[1]

“A1.1 the protocol is open, free, and universally implementable”[1]

“I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. ”[1]

. . .

[1] https://www.nature.com/articles/sdata201618

Geonetwork – upgrading records

AuScope AVRE runs a geospatial GeoNetwork catalogue[2]

To help with FAIR compliance

1. Upgrading GeoNetwork[2] records from ISO19139:2007 XML standard to ISO19115-3:2018[1]
2. Adding richer metadata

[1] https://geonetwork-opensource.org/manuals/trunk/en/annexes/standards/iso19115-3.2018.html
[2] https://www.geonetwork-opensource.org/

Geonetwork – back to the future

- Most of our GeoNetwork records are created via harvesting of OCG services of other organisations, e.g., state geological surveys, NCI
- Harvesting was intended to lower costs over storing & maintaining duplicate copies of the data
- But to create ISO19115-3 records the harvester must be configured with an ISO19115-3 template
- Contributions have been made to GeoNetwork’s source code
- Non-harvested records easily converted via export & re-import

Other geonetwork instances

1. Need to support all versions of GeoNetwork
2. Catalogue quality and standards do vary between suppliers
3. Need to maintain backward compatibility of earlier versions

Non-harvestable services

What to do?

1. Hand maintained GeoNetwork records
2. Try a lower barrier to entry e.g. configure ‘pycsw’ (https://pycsw.org/) to read database and provide a basic CSW interface

Non-geospatial hurdles

- Many institutions use metadata catalogues which have machine readable interfaces but do not support geospatial coordinates e.g. figshare (https://figshare.com/)
- Requirement on the owner to implement custom fields

  e.g. 2D bounding box
- Advocacy needed:

  1st hurdle: convincing owners to implement custom fields

  2nd hurdle: standardise these customisations across institutions

Improving machine readable APIs?

What about new OGC Records APIs[1] ?

- Still a work in progress
- Available as a fledgling microservice in GeoNetwork[2]

Others (e.g. pycsw) do not support ISO19115-3:2018

There is interest in updating pycsw to support ISO19115-3

[1] https://github.com/opengeospatial/ogcapi-records & https://ogcapi.ogc.org/records/

[2] https://github.com/geonetwork/geonetwork-microservices/blob/main/modules/services/ogc-api-records/README.md

Area 03: Dataset Provider

Dataset Providers

“F2. data are described with rich metadata”[1]

“A1. (meta)data are retrievable by their identifier using a standardized communications protocol”[1]

“R1.1. (meta)data are released with a clear and accessible data usage license.”[1]

“I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. ”[1]

“R1. (meta)data have a plurality of accurate and relevant attributes.”[1]

. . .

[1] https://www.nature.com/articles/sdata201618

Managing Server Instances

- Most OCG-compliant server instances accessed by the AuScope Portal are owned by third parties
- Need to convince and educate their owners of the importance of FAIR:
  1) Keep their metadata up to date
  2) Confirm licensing conditions
  3) Provide citations and identifiers

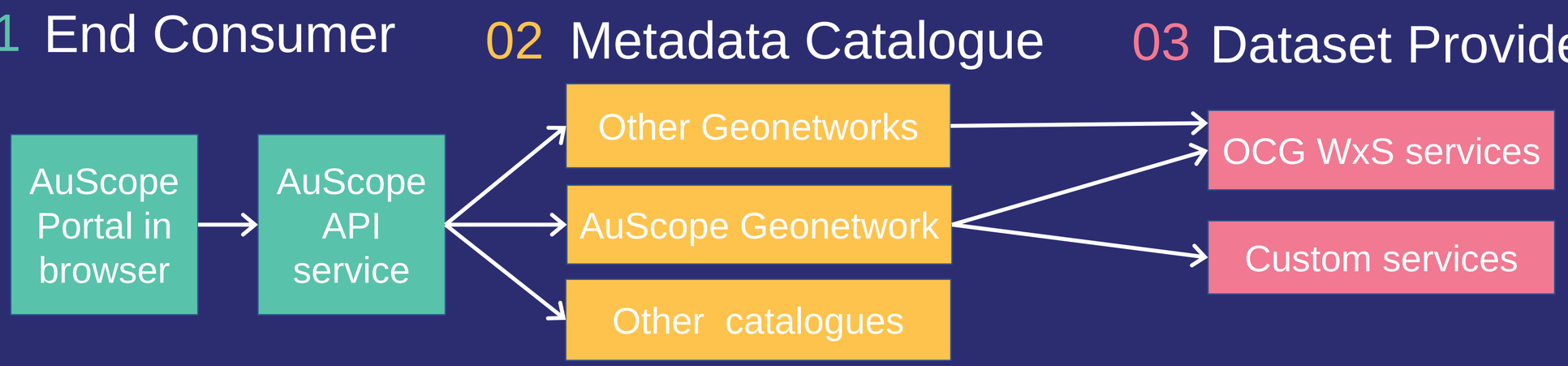
Improving machine readable APIs ?

- New OCG APIs[2] not utilised as yet Why?
- Geoserver OGC API plugin still under ‘heavy development’[1]
- Others are not widely adopted among our data providers

  e.g. https://pygeoapi.io/

[1] https://docs.geoserver.org/latest/en/user/community/ogc-api/index.html
[2] https://opengeospatial.github.io/e-learning/ogcapi-features/text/basic-main.html

Summary



In Progress:

Develop metadata enriched software (e.g. F2)

Upgrade GeoNetwork catalogues to iso19115-3 (F2)
Advocacy for geospatial institutional catalogues (F4)
Co-opt non-OCG data sources (A1)

Be FAIR Aware (R1.1)

Later:

OCG API development (A1)

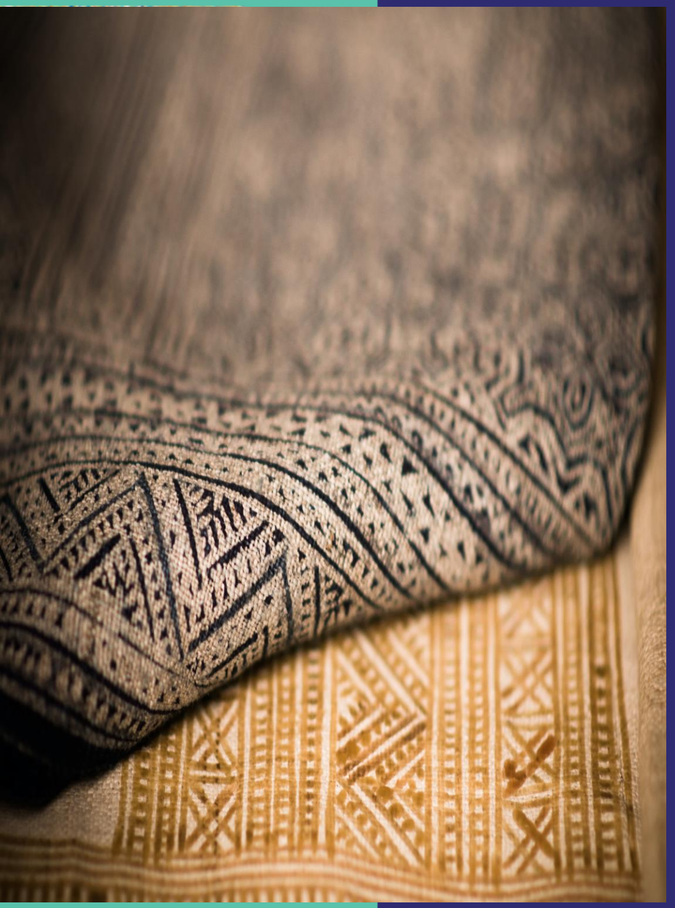
OGC API Upgrade (A1)

OGC API Upgrade (A1)

Reference: https://www.nature.com/articles/sdata201618

AuScope

Answering Australia's Geoscience Questions



Source: https://unsplash.com/@trang3019

Thank you

w auscope.org.au/avre

e vincent.fazio@csiro.au

CSIRO

NCRIS National Research Infrastructure for Australia

An Australian Government Initiative