TOWARDS A WEB-ENABLED GEO-SAMPLE WEB: AN OPEN SOURCE RESOURCE REGISTRATION AND MANAGEMENT SYSTEM FOR CONNECTING GEO-SAMPLES TO THE WEB

Anusuriya Devaraju[1], Jens Klump[1], Victor Tey[1], Simon Cox[2] and Ryan Fraser[1]

International Conference for Free and Open Source Software for Geospatial (FOSS4G 2017), August 14-19, 2017, Boston, MA.

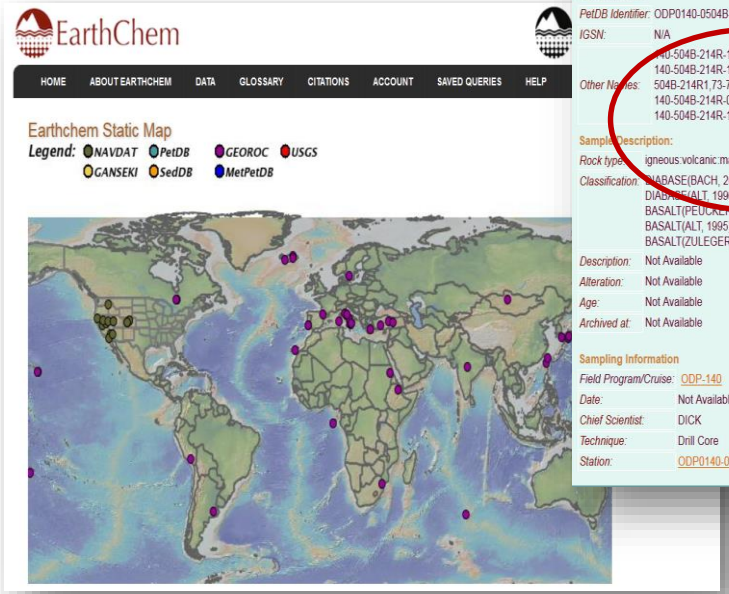
[1]MINERAL RESOURCES, [2]LAND AND WATER
www.csiro.au

Presentation Outline

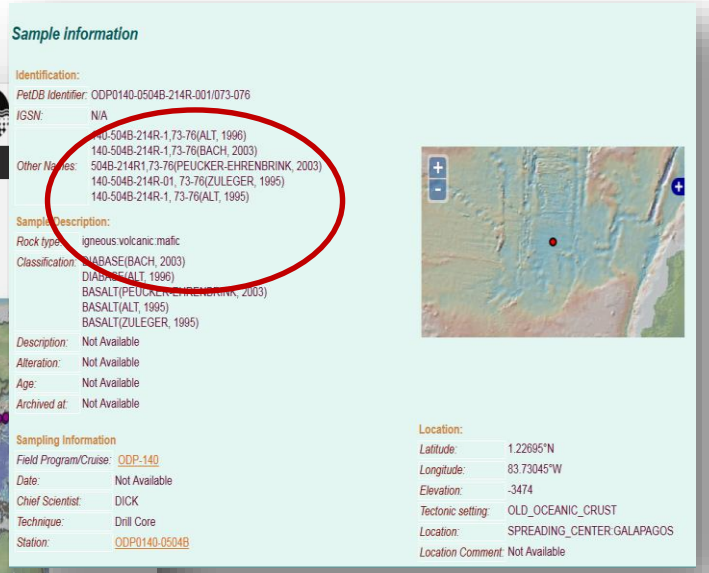
- Challenges in sample identification and discovery
- International Geo Sample Number (IGSN)
- IGSN Implementation in CSIRO
- Applications
- Lessons Learned and Conclusions

Challenges

- Inconsistent sample cataloguing practices.
- The lack of online catalogues for discovering physical samples.



The EarthChem data portal returns 71 samples called ‘A-1’.



Different names representing the same rock sample on the PetDB (Petrological Database)

Examples

Date: Three Weeks Ago

[redacted] (... There are 2 bags of Pulp Samples addressed to 'Blackham ... Wed 12/07/... 19 KB
Administration Support Officer Property (WA, SA, NT & Irymple) CSIRO Business &

Date: Last Month

[redacted] . Is anyone expecting 2 x roller racks 50kg? - please come to... Wed 14/06/... 16 KB
mann Administration Support Officer Property (WA, SA, NT & Irymple) | CSIRO Business &

Date: Older

[redacted] ... Looking for crude oil samples Thu 27/04/2... 10 KB
Hi All, I have a project for which I need a small batch of crude oil samples.. I need roughly 6 ranging

[redacted]

[redacted] (... A pallet of Mineral Samples in core trays, has been put in B... Thu 13/04/2... 520 KB
Administration Support Officer Property (WA, SA, NT & Irymple) CSIRO Business &

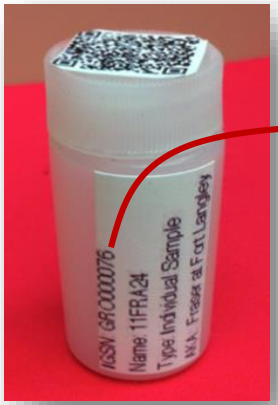
[redacted] ... Missing sample Wed 22/02/... 94 KB
MESSAGE FROM: [redacted] Hi folks, I am missing a pretty important sample. It is a zoned

Presentation Outline

- Challenges in sample identification and discovery
- International Geo Sample Number (IGSN)
- IGSN Implementation in CSIRO
- Applications
- Lessons Learned and Conclusions

International Geo Sample Number (IGSN)

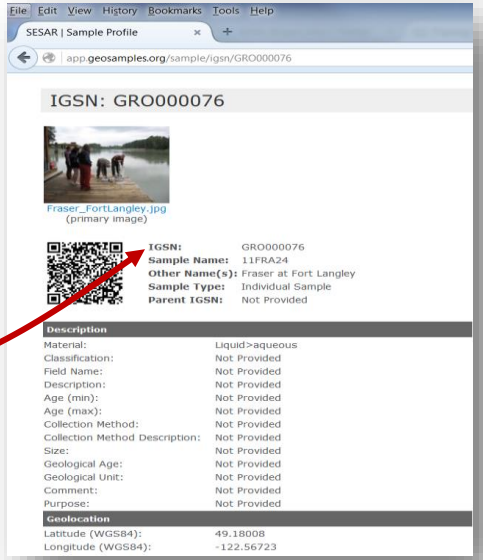
- Globally unique, persistent identifiers for physical samples (e.g., rocks, water, bio specimens), sample collections, and sampling features (e.g., boreholes).



Fraser River water sample
(image source: @igsn_info)

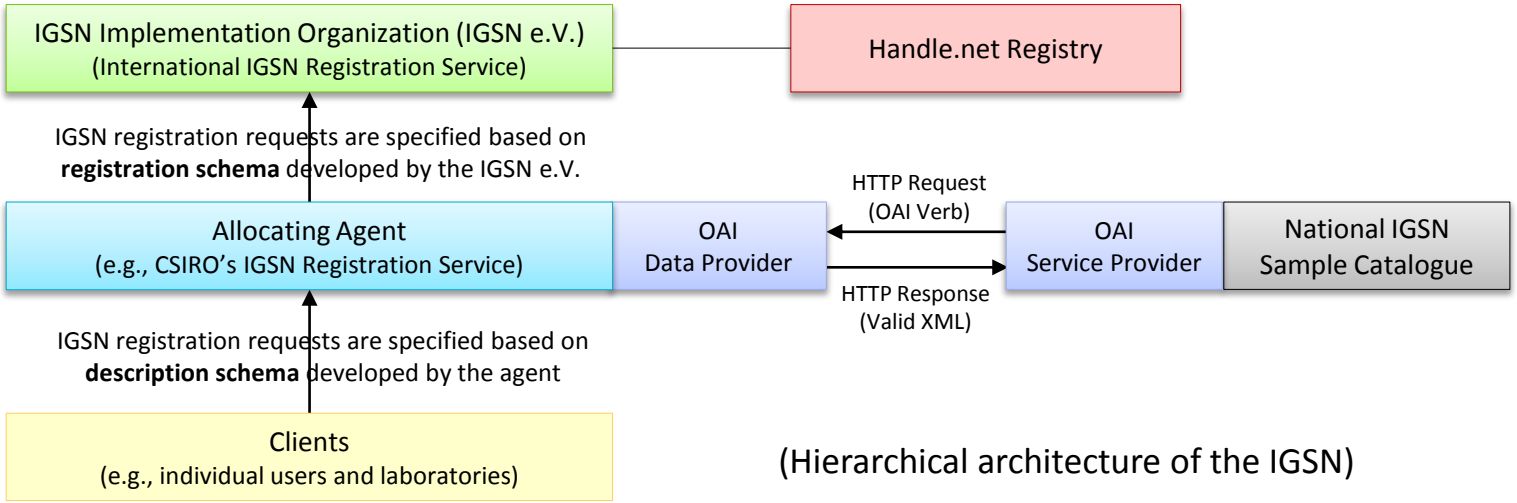
IGSN:GRO000076
http://hdl.handle.net/10273/GRO000076

IGSN ensures preservation and access of sample metadata



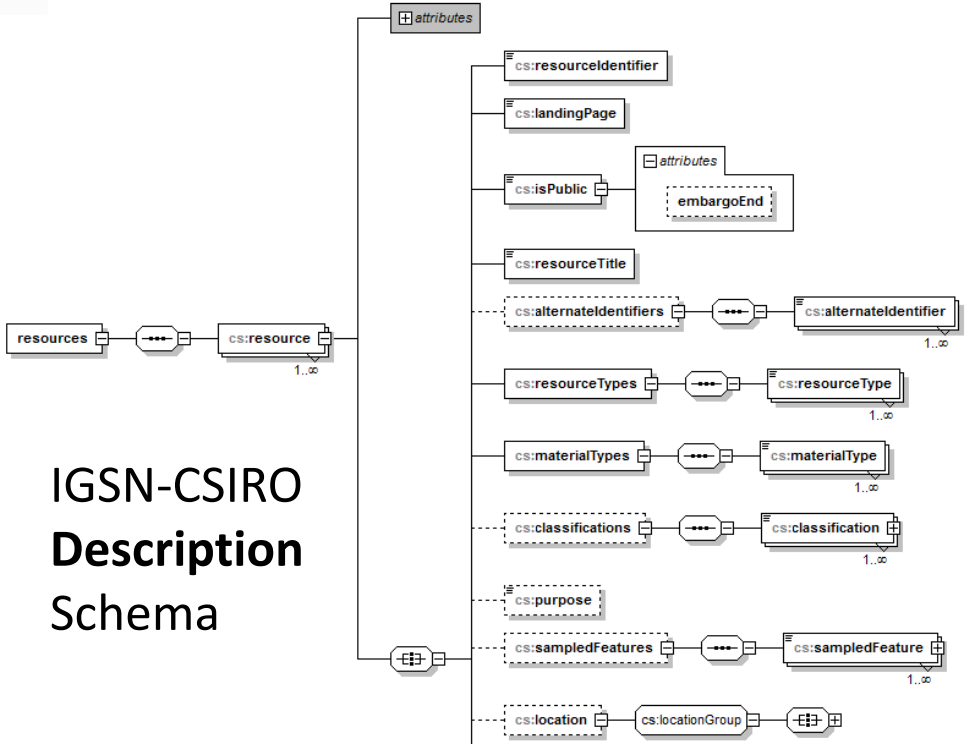
IGSN System Overview

- Clients send sample registrations to an allocating agent; the allocating agent forwards the registrations to the IGSN e.V.
- Australian IGSN allocating agents are CSIRO, Geoscience Australia, Curtin Uni.

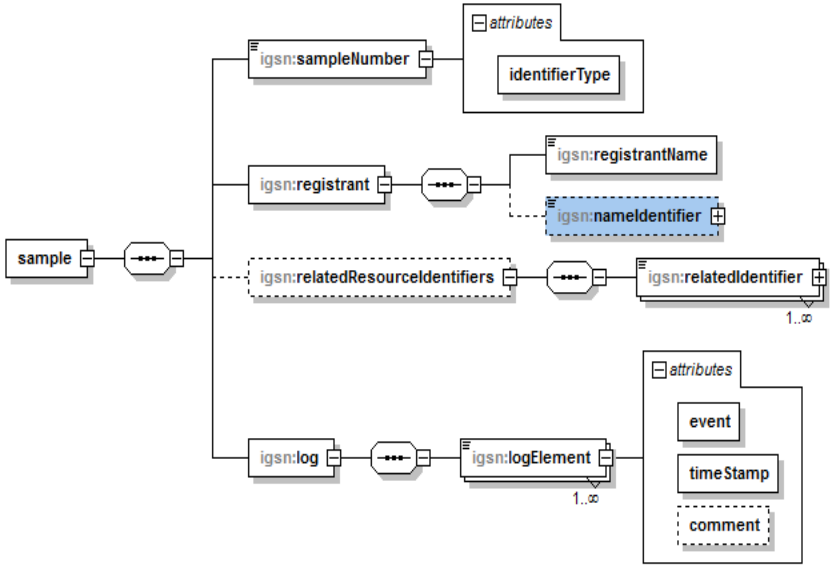


(Hierarchical architecture of the IGSN)

Metadata Schemas



IGSN-CSIRO **Description** Schema



IGSN **Registration** Schema

Presentation Outline

- Challenges in sample identification and discovery
- International Geo Sample Number (IGSN)
- IGSN Implementation in CSIRO
- Applications
- Lessons Learned and Conclusions

IGSN Implementation in CSIRO

- CSIRO became a member of IGSN in 2013, and the use of IGSN was initiated by CSIRO Mineral Resources.
- IGSN are currently used for:

Sample Data Repositories	Material Types	Registrations (as at 31.05.2017)
Repository of the Australian Resources Research Centre(ARRC)	rock, mineral, soil	25652
Capricorn Distal Footprints	rock, vegetation, water, regolith	4232
Reflectance Spectra Reference Libraries	mineral, rock, synthetic material	94

- Future use cases are National Soil Archive and Australian National Insect Collection.

Identifier Governance in CSIRO

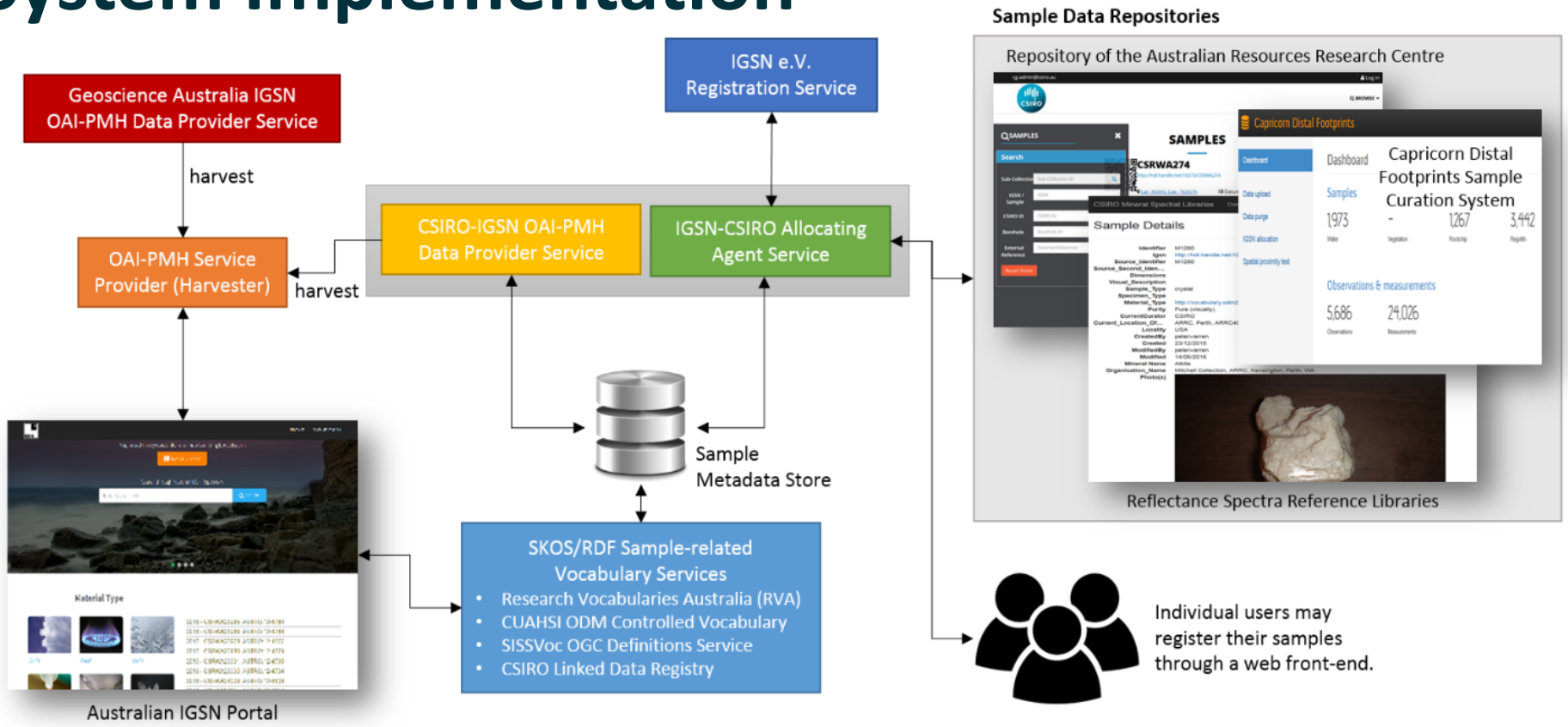
- Different namespace strategies may be used by other allocating agents.
- Following the IGSN Technical Specification, we established rules for allocating IGSNs in different parts of the organisation.

1. Allocating agent namespace ('CS' represents CSIRO)
2. Sub-namespace of a client
3. Sample code specified by the client

IGSN : CSRWASC00630

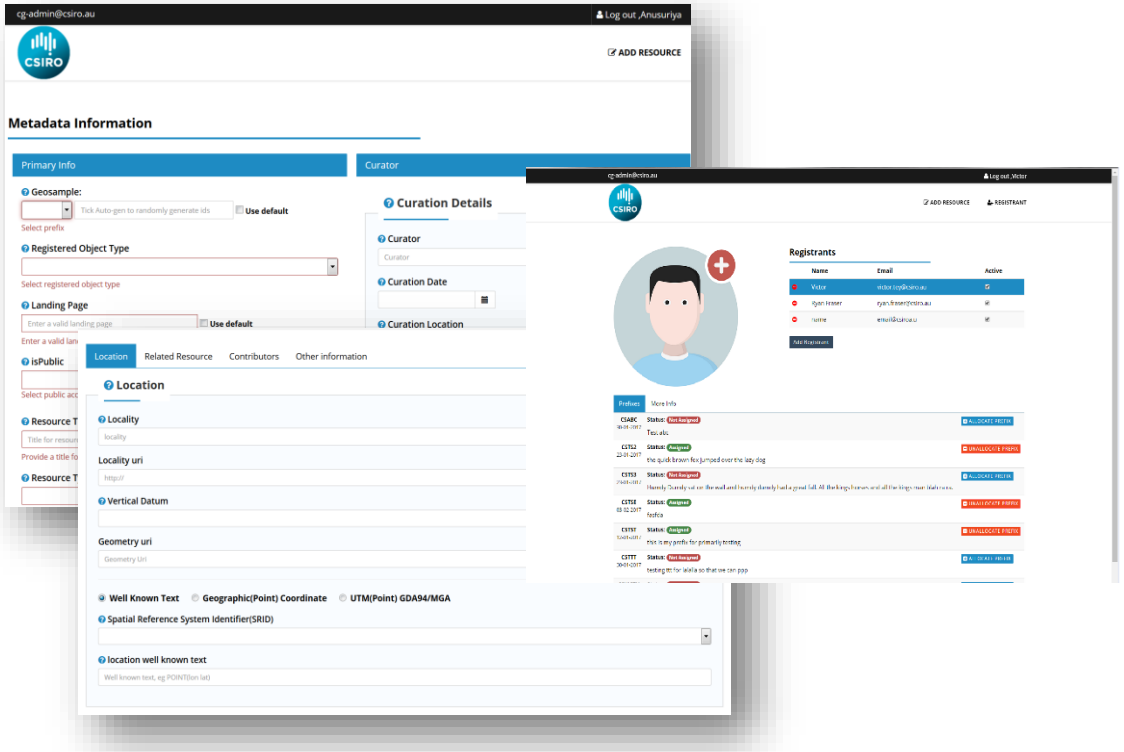
1 2 3

System Implementation



A Web Front-End for Sample Registrations

Individual users may register their samples through the GUI.



Prefix and user registration can be performed via the user interface from the Registrant menu option.

CSIRO-IGSN Description Metadata Schema

- Built upon the DataCite Metadata Schema (v4.0).
- An XML Schema which represents the core concepts of physical samples:
  - Sample identification, e.g., identifier, alternate identifiers, title, visibility, material, specimen type..
  - Sampling activity, e.g., location, time, purpose, method…
  - Sample curation, e.g., curator, curation location and time…
  - Related resources, e.g. publications, datasets…
- Supports batch sample registrations and has minimal restrictions on mandatory metadata elements.
- Captures the provenance of sample curation, and the relations between samples and related resources.
- Flexible in terms of representing spatial and temporal information.

SKOS-based Controlled Vocabularies

Vocabularies	Examples	Service	Provider
Material types	soil, rock, vegetation	Master Controlled Vocabulary Registry for ODM2	CUAHSI
Specimen types	thin section, grab, dredge, cuttings	Master Controlled Vocabulary Registry for ODM2	CUAHSI
Nil-reason types	missing, unknown, withheld.	OGC definitions of nil reasons CSIRO	CSIRO
Contributor types	originator, custodian, point of contact.	Linked Data Registry	CSIRO
*Registration types	physical sample, sample collection, sampling features.	ANDS Research Vocabularies Australia	CSIRO & Geoscience Australia
*Identifier types	DOI, IGSN, LSID, ORCID	ANDS Research Vocabularies Australia	CSIRO & Geoscience Australia
*Relation types	isDerivedFrom, hasDocument, hasDigitalRepresentation.	ANDS Research Vocabularies Australia	CSIRO & Geoscience Australia

A list of existing and newly developed SKOS vocabularies. New vocabularies are indicated by asterisks*.

CSIRO-IGSN Allocating Agent Service

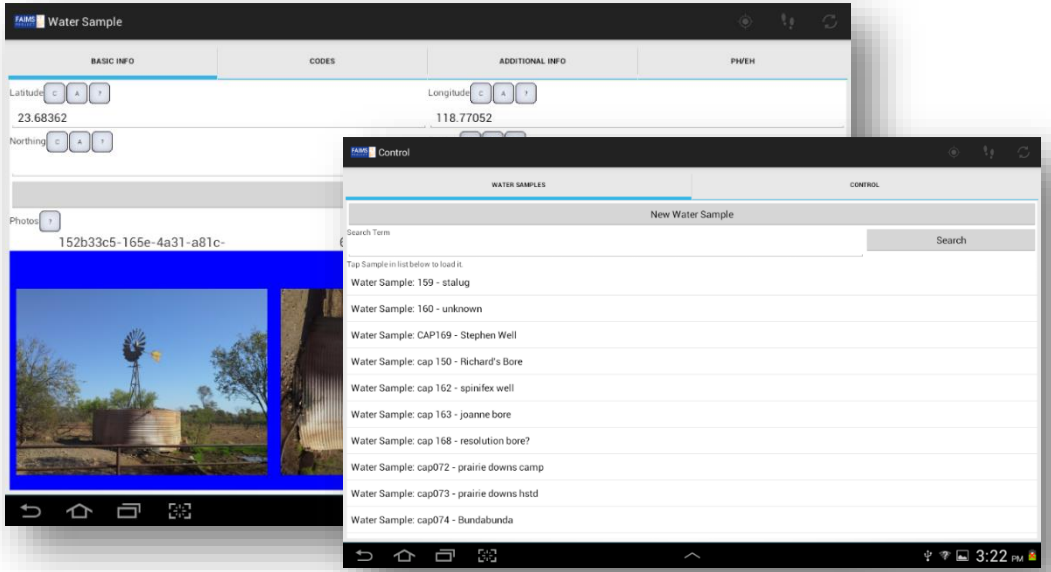
- Sample registrations must be specified based on the CSIRO-IGSN Description Schema.
- A RESTful web service that enable clients to
  - Register a sub-namespace
  - Get all registered sub-namespaces
  - Register samples
  - Get metadata of a sample or a list of samples
  - Mark a sample metadata as deprecated.

Presentation Outline

- Challenges in sample identification and discovery
- International Geo Sample Number (IGSN)
- IGSN Implementation in CSIRO
- Applications
- Lessons Learned and Conclusions

Applications

Tracking samples from the field to the sample repository.

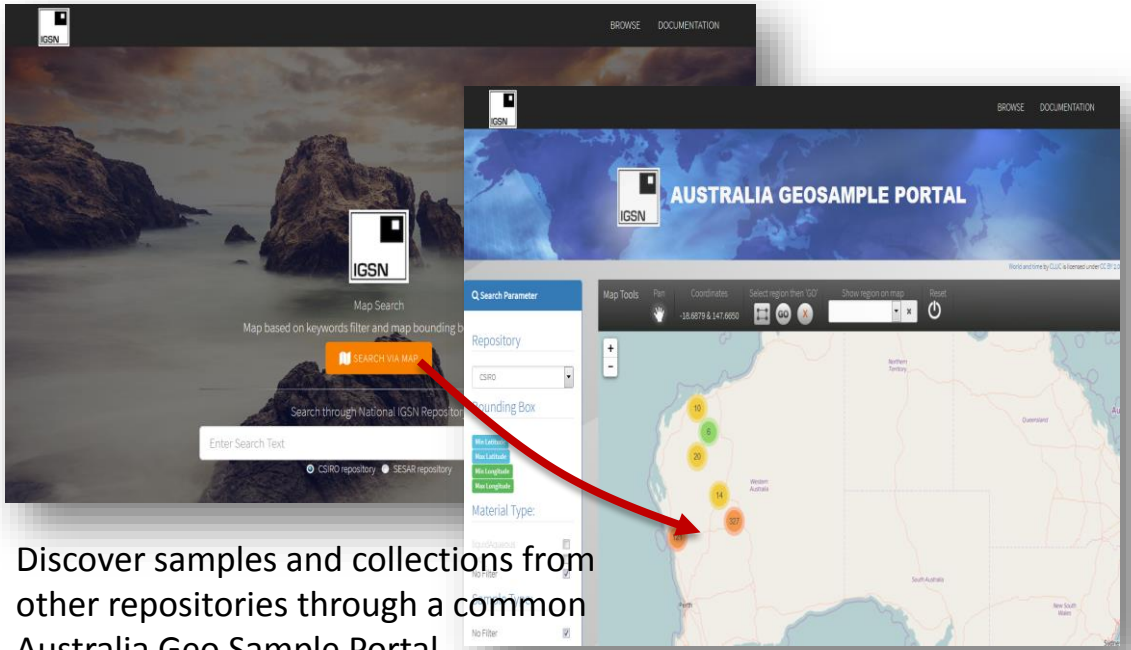


IGSN pre-allocation in mobile field data acquisition systems.

Images courtesy of Reid, N. (2015), Golodoniuc, P. (2017)

IGSN Applications

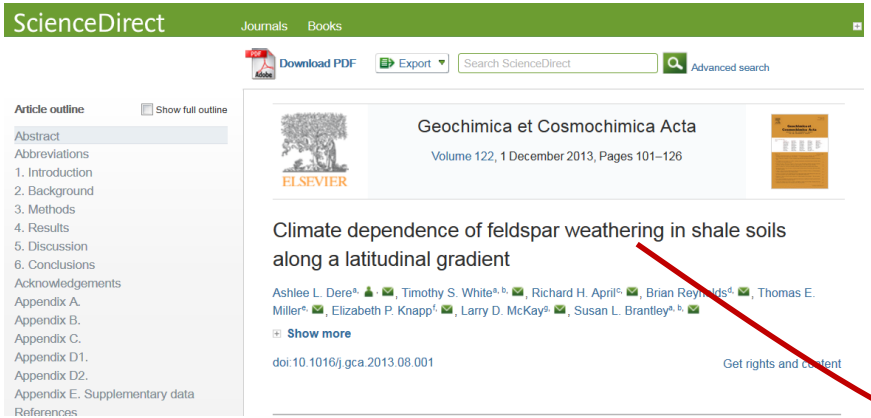
Smart labels for sample inventory management.

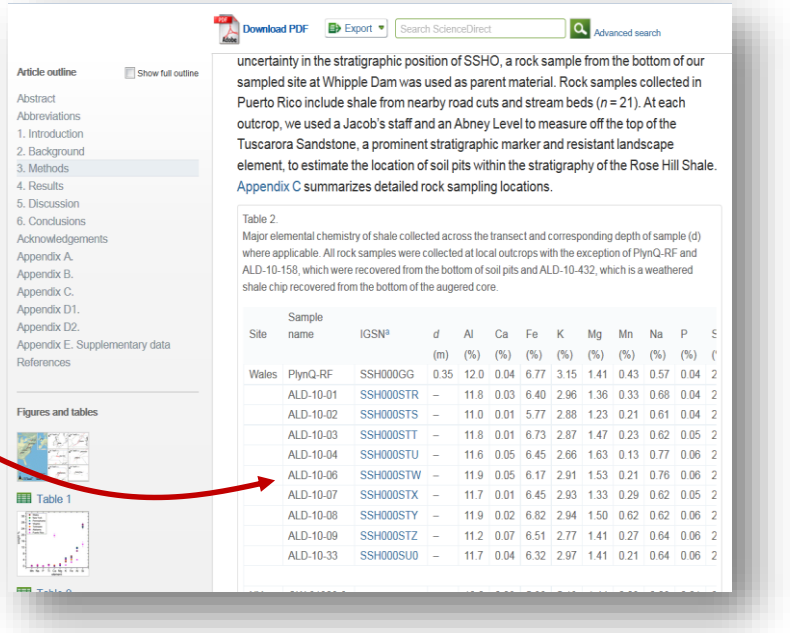


Discover samples and collections from other repositories through a common Australia Geo Sample Portal.

IGSN Applications

Reference samples by using IGSNs in publications, e.g., reports, articles, grant proposals.



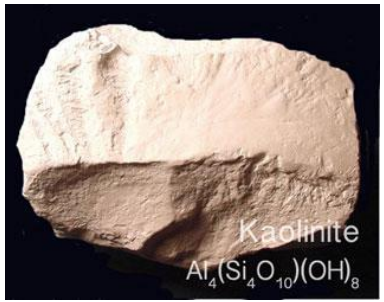


Publication URL :
http://dx.doi.org/10.1016/j.gca.2013.08.001

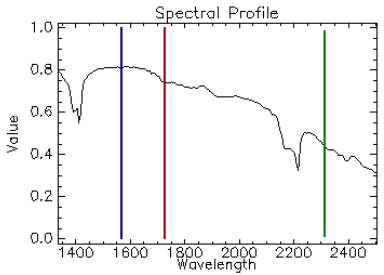
IGSN Applications

Cross-reference between the registered samples and their related resources (sub-samples, datasets, publications).

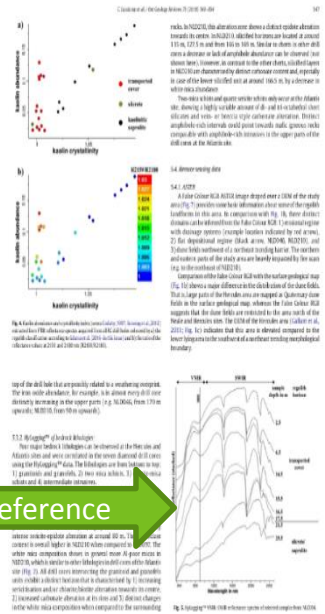
Specimen (IGSN)

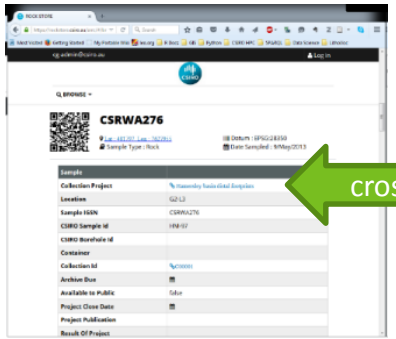


Spectrum (DOI)

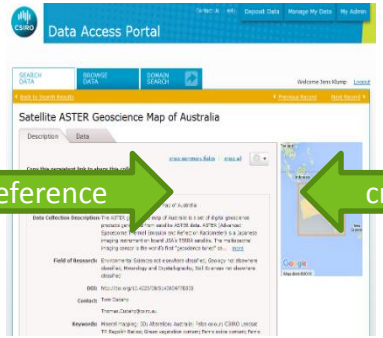


Publication (DOI)





cross-reference



cross-reference

Presentation Outline

- Challenges in sample identification and discovery
- International Geo Sample Number (IGSN)
- IGSN Implementation in CSIRO
- Applications
- Lessons Learned and Conclusions

Lessons Learned

- We ensure the global uniqueness of registered identifiers within the organization through the IGSN namespace governance.
- Integration of identifiers into new and existing systems
  - New sampling campaigns – adopt IGSN at the early stage of the sampling activity.
  - Existing sample curation systems – prepend the IGSN namespace to the local sample identifiers.
- What can be identified with IGSNs? We used IGSNs to identify not only physical samples but also sample collections and sampling features..
- Outreach, e.g., talks, meetings, and workshops.
- Documentations & public repositories.
- National collaboration, i.e., a joint project with the other allocating agents in Australia (e.g., Geoscience Australia and Curtin University).

Conclusions

- We described the implementation of IGSN to identify and publish physical samples and sample collections in CSIRO.
- This serves as a working example for implementing IGSN in a large organization.
- Benefits of implementing IGSN in CSIRO
  - Identify samples unambiguously and discover them easily.
  - Avoid duplicate sampling activities.
  - Promote re-use of the samples for new purposes.
- The use of IGSN in CSIRO may be expanded to other collections and operated as a central service by Information Management and Technology (IM&T) as part of CSIRO's National Facilities and Collections.

Related Links

Components	Link
CSIRO Allocating Agent Service	https://igsn.csiro.au/igsn30/api
Description Metadata Schema	https://igsn.csiro.au/schemas/3.0/
Metadata Store (source)	https://github.com/AuScope/igsn30/tree/master/sql
CSIRO-IGSN OAI-PMH Data Provider	https://igsn.csiro.au/igsn30/api/service/30/oai
National IGSN Web Portal	http://igsn.org.au
OAI-PMH Harvester and National IGSN Web Portal (source repositories)	https://github.com/AuScope/NatPortalIGSN
Sample Registration and Management GUI	https://igsn.csiro.au/igsn30

A list of the components developed and their links.

Mineral Resources
Anusuriya Devaraju
Postdoctoral Fellow
e anusuriya.devaraju@csiro.au

Mineral Resources
Jens Klump
Science Leader Earth Science Informatics
E jens.klump@csiro.au

Mineral Resources
Victor Tey
Senior Software Engineer
E victor.tey@csiro.au

Land and Water
Simon Cox
Chief Research Scientist
e simon.cox@csiro.au

Mineral Resources
Ryan Fraser
Research Manager
e ryan.fraser@csiro.au