Registering Samples with SESAR

Kerstin Lehnert and Sarah Ramdeen

CZ HUB Data Management Mini Workshop
February 2nd 2022

Agenda

- Motivations for sample management
- What is an IGSN
- Introduction to SESAR
- Demo
- Discussion/Questions

Why Talk About Sample Management?

Samples are fundamental research products and research resources

Samples need to be Findable, Accessible, Interoperable, & Reusable (=FAIR) for the same reasons that data and software need to be FAIR

- Open, transparent, and reproducible science
- Broader impacts and return on investment

Following best practices in sample management allows you to

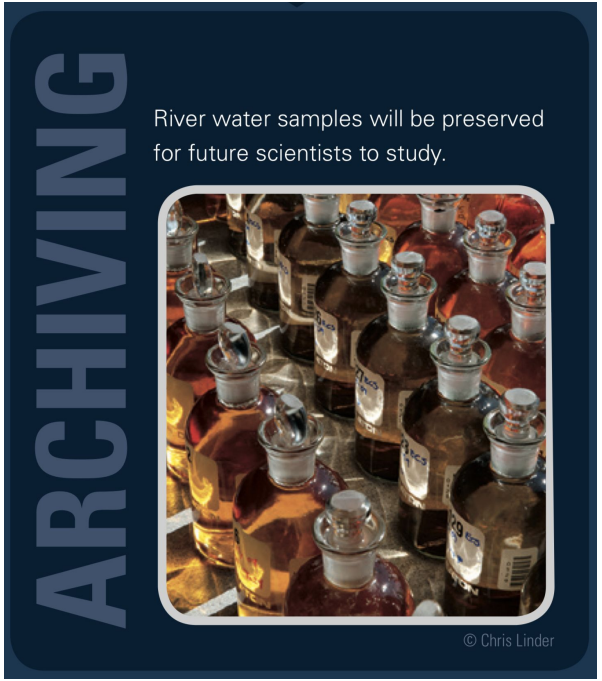
- Work more efficiently with samples (track, share, and cite samples)
- Grow the scientific impact of the samples

The Need for FAIR Samples

"Maybe ten years from now, you want a sample of 2010 Fraser River water. Where do you go? How do you get it? …

… I would give a lot for a 1930 or 1960 water sample from the Mississippi! But it's gone."

From "RiverQuest – Sampling the world's rivers to assess our planet's health" by Kate Madin



Why CZNet Needs to Talk about Samples

NSF data policy includes samples!

"All NSF proposals must include a document of no more than two pages uploaded under "Data Management Plan" in the supplementary documentation section of the proposal. This supplementary document should describe what data/samples will be collected, what analyses will be done, and how the project will provide open and rapid access to samples, data, derived data products (e.g., models and model output), and other information on the project during and after the project's completion."

Sharing samples across clusters can augment scientific insights

ENABLING FAIR DATA PROJECT

Individual Researchers will strive to:

- Make research outputs FAIR and, whenever possible, open by depositing research outputs (e.g., data, software, physical sample information, etc.) in trustworthy, community-accepted, FAIR-aligned repositories that support:
  - Persistent identifiers for data (and other research outputs as is possible) and consistently using these in citations.
  - Licenses for data (and other research outputs as is possible) that is as open as possible to enable the widest potential reuse.
- Cite data, software, physical samples, and other products created or reused for your research in your publications.
- Include a data availability statement in your publication to make it clear where the data (and other research outputs as is possible) that supports the paper can be accessed along with any other access information.
- Prepare, use, and manage data management plans for your data and other research outputs. Keep the plan updated as research progresses.
- Educate colleagues in practices that enable open and FAIR research outputs.
- Support development of open and FAIR standards and practices in your institutions and organizations, and in scholarly publishing as authors, reviewers, and editors.

- Cite data, software, physical samples, and other products created or reused for your research in your publications.

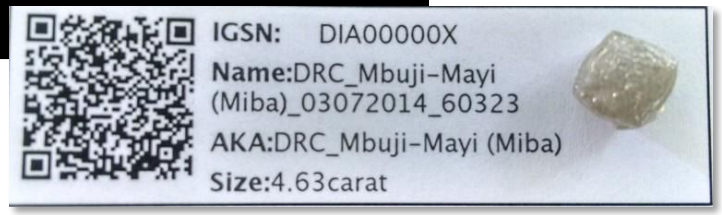
Requirements for FAIR Research Outputs

- unique and persistent identifiers
- metadata appropriate to assist discovery
- citation in a form equivalent to other scholarly outputs.
- accessible through a standard, web-based protocol (technical interoperability)
- provenance information
- usage license
- well curated & persistently accessible
- linked securely to associated publications and other resources.

Requirements for FAIR Research Outputs

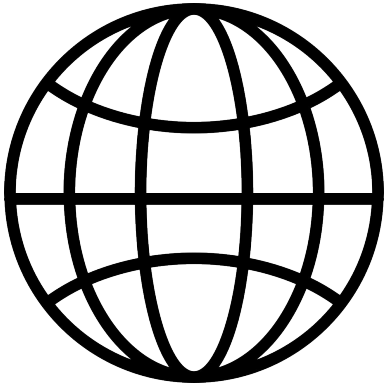
- unique and persistent identifiers ➪ International Generic Sample Number (IGSN)
- metadata appropriate to assist discovery ➪ SESAR metadata profiles
- citation in a form equivalent to other scholarly outputs.
- accessible through a standard, web-based protocol ➪ SESAR web services
- provenance information ➪ SESAR metadata profiles
- usage license
- well curated & persistently accessible ➪ SESAR metadata catalog & profile pages
- linked securely to associated publications and other resources. ➪ IGSN & EarthChem

IGSN is a globally unique, persistent identifier for samples



10M sample iDs issued

Community demand to scale to billions to support unique identification and discoverability of samples and collections



Global participation

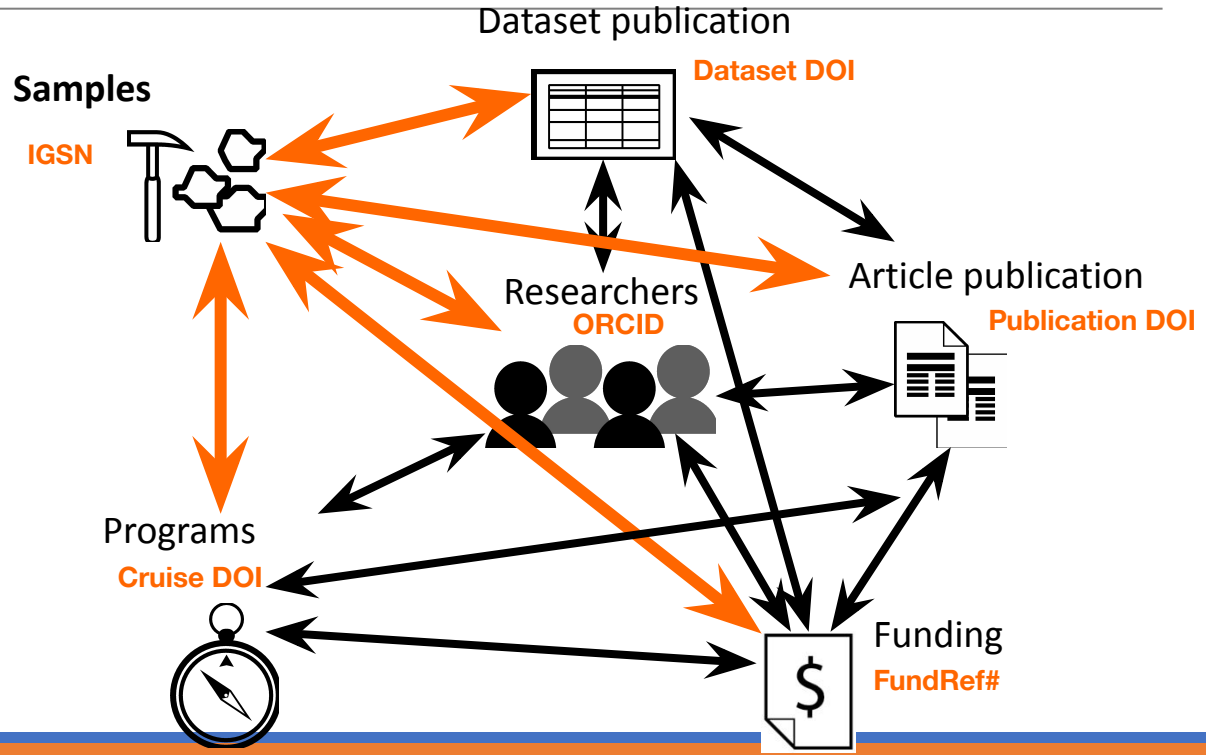
Adoption in geosciences driving interest in archaeology, biodiversity, materials science, genomics, planetary sciences

Value

Combine leadership in PID registration technology, core metadata, and communities of practice

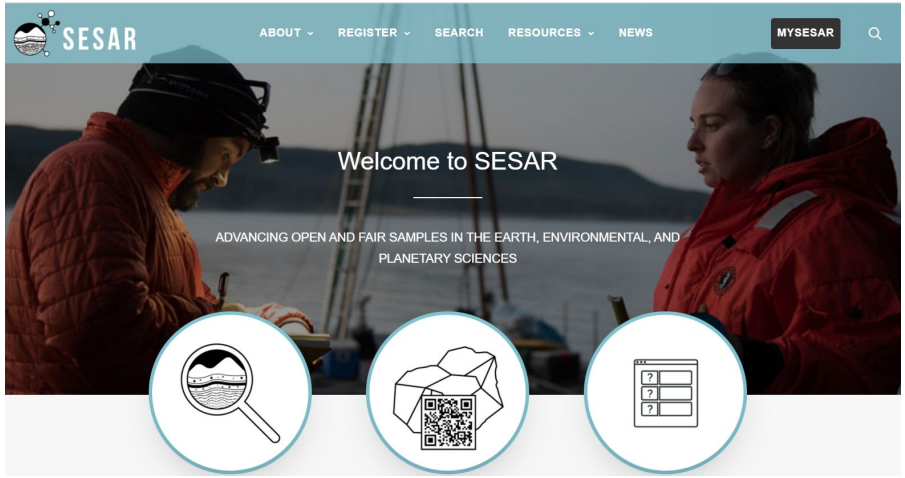
Best Practice for Sample Identification: IGSN

- Locate (Find)
- Access
- Link (Interoperate)
- Cite



What is SESAR?

SESAR is a community platform that helps make samples more discoverable, accessible, and reusable, and connects samples with the knowledge ecosystem derived from them.

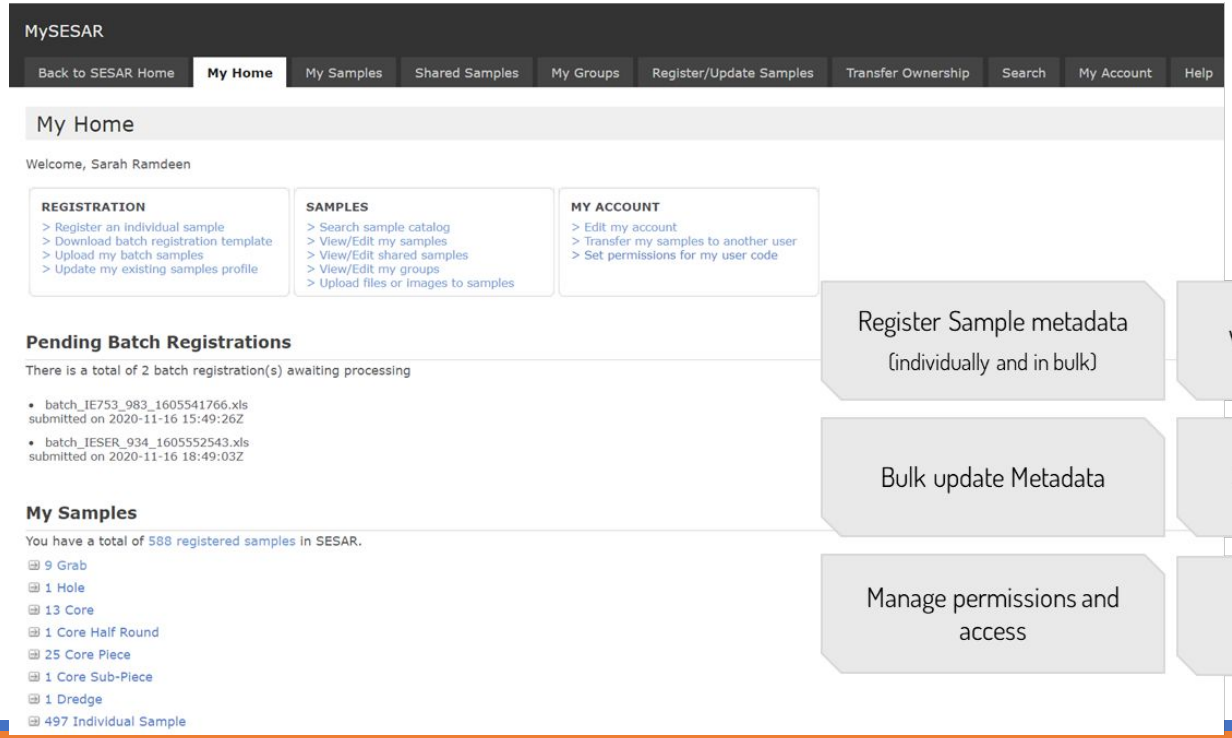


Primary services

- Metadata management system
- Registration of sample metadata and minting of IGSN
- Catalog for discovery/access

www.geosamples.org

Management System for Sample Metadata



Register Sample metadata (individually and in bulk)

View and Edit Metadata

Bulk update Metadata

Upload files or images

Manage permissions and access

Export metadata into EarthChem Library templates

IGSN registration

As part of our registration services we:

- Provide IGSN during the sample registration process
- Maintain metadata profile pages that IGSN ‘resolve’ to
- Collect and maintain descriptive metadata unique to the SESAR community

IGSN: IAC000009

KI-04 hand sample.JPG
(primary image)

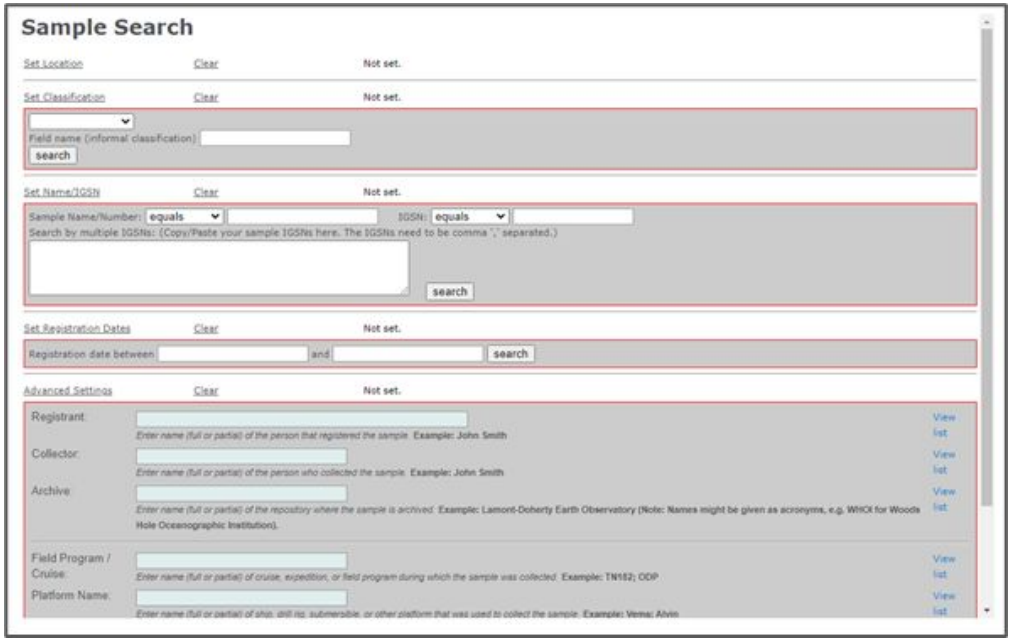
IGSN:	IAC000009
Sample Name:	KI-04-112710
Other Name(s):	
Sample Type:	Individual Sample
Parent IGSN:	Not Provided

Description	
Material:	Rock
Classification:	Igneous>Volcanic>Mafic
Field Name:	basanite
Description:	Xenolith-bearing mafic lava with large (1cm) crystals. Low vessicularity.
Age (min):	Not Provided
Age (max):	Not Provided
Collection Method:	rock hammer
Collection Method Description:	Not Provided
Size:	5 kg
Geological Age:	Not Provided
Geological Unit:	Not Provided

Catalog for Discovery and Acccess

The SESAR Catalog Search allows users to:

- Search the SESAR catalog by geospatial information, sample type, sample classification, archive, and other criteria.
- View individual sample profiles.
- View samples locations on maps.
- Download lists of samples.

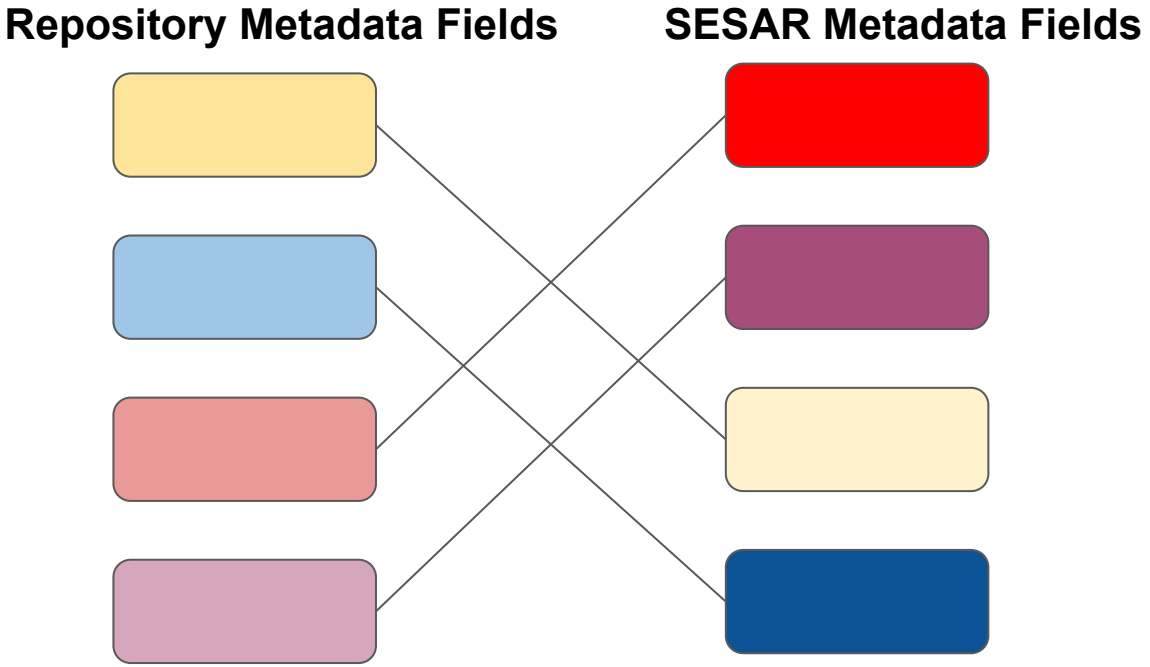


SESAR Demo

- Accessing MySESAR
- Account management
  - Creating a user code
  - Sharing a user code
- Batch template creation (for bulk registration)
- Registering a sample
  - Individual Sample
  - Batch Template
  - *Demonstration of web services can be scheduled*

Metadata mapping

AVO	SESAR
StationID	*NA*
SampleID	Sample Name
Latitude	Latitude
Longitude	Longitude
Geologist	Collector/Chief Scientist
DateVisited	Collection date
Volcano	Primary physiographic feature
Location Description	Location description
Sample Description	Sample description
SampleType	Field name (informal classification)
AT GMC	Current archive
	Current archive contact
Material	Material
NA	Classification
NA	Country
NA	State/Province
NA	Collection method



Metadata mapping

A	B	C	D	E	F	G	H	I	J	K	L
StationID	SampleID	Latitude	Longitude	Geologist	DateVisited	Volcano	Location Description	Sample Description	SampleType	AT GMC	Material
05LSCN001	05LSCN001A	51.9585	178.51265	Neal Christina A	09/28/05	Little Sitkin	Block and ash flow s	Dense clasts from inc	Lava	yes	Whole-rock
05LSCN001	05LSCN001B	51.9585	178.51265	Neal Christina A	09/29/05	Little Sitkin	Block and ash flow s	Pumiceous juvenile	Pumice	yes	Whole-rock
05LSCN001	05LSCN001C	51.9585	178.51265	Neal Christina A	09/30/05	Little Sitkin	Block and ash flow s	Dense clast from pyr	Pyroclastic flow	yes	Whole-rock
05LSCN007	05LSCN007A	51.95359	178.48427	Neal Christina A	10/01/05	Little Sitkin	West Cove lava flow	No description availa	Lava	yes	Whole-rock
05LSJL001	05LSJL001A	51.935217	178.518933	Larsen Jessica Faust	09/29/05	Little Sitkin	south flanks	Large dark black bon	Bomb	yes	Whole-rock
05LSJL001	05LSJL001B	51.935217	178.518933	Larsen Jessica Faust	09/29/05	Little Sitkin	south flanks	Sample collected fr	Lava	yes	Whole-rock
05LSJL002	05LSJL002	51.93495	178.5172	Larsen Jessica Faust	09/29/05	Little Sitkin	south flanks	Sample from a clear	Lava	yes	Whole-rock
05LSJL003	05LSJL003	51.934267	178.5144	Larsen Jessica Faust	09/29/05	Little Sitkin	south flanks	Juvenile sample fron	Bomb	yes	Whole-rock
05LSJL005	05LSJL005	51.926133	178.517033	Larsen Jessica Faust	09/29/05	Little Sitkin	south flanks	juvenile clast from fr	Lapilli	yes	Whole-rock
05LSJL008	05LSJL008	51.9225	178.526483	Larsen Jessica Faust	09/29/05	Little Sitkin	south flanks	Andesite lava flow. M	Lava	yes	Whole-rock
05LSJL009	05LSJL009	51.921217	178.528233	Larsen Jessica Faust	09/29/05	Little Sitkin	south flanks	Andesite lava flow sa	Lava	yes	Whole-rock
05LSJL013	05LSJL013	51.904117	178.5313	Larsen Jessica Faust	09/29/05	Little Sitkin	south flanks beach	Light gray sugary and	Lava	yes	Whole-rock

Convert original data file to the SESAR Batch template

Object Type:	Site	User Code:	IEAVO					
Sample Name	IGSN	Parent IGSN	Release Date	Other name(s)	Latitude	Longitude	nary physiographic feat	of physiographic fe	
05AMJL001	IEAVO0002			249	51.40488	179.28092	volcano		Amchitka NW
K-11H	IEAVO0003			33797	58.0272	-155.6642	volcano		No latitude/lon Latitude/longit volcano.
K-12B	IEAVO0004			33798	58.0272	-155.6642	volcano		No latitude/lon Latitude/longit volcano.
K-12C	IEAVO0005			33799	58.0272	-155.6642	volcano		No latitude/lon Latitude/longit volcano.

Discussion and Questions

- FAQs:
  - SESAR supports pre-registration of samples (before collection)
  - You can register a sample that has been or will be destroyed
  - Samples can be kept private for up to 2 years after registration
- Metadata mapping
  - Consider which metadata is important for search and discovery (re-use)
  - Should there be
    - CZ community specific metadata?
    - CZ Cluster specific metadata?

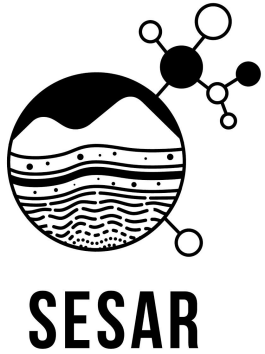
We are available for one on one meetings to discuss individual needs!

Thank you!

Explore the link below for more information about using SESAR

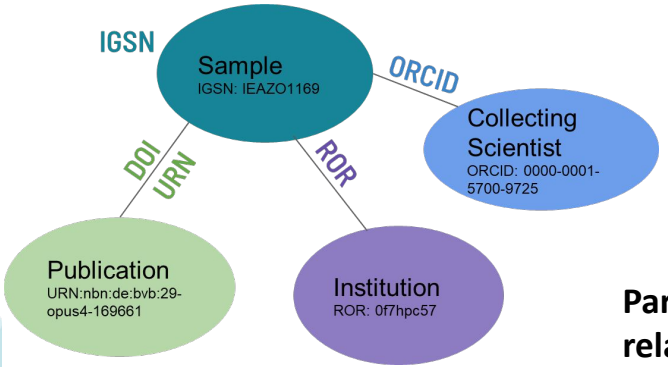
https://bit.ly/3kVJeQ6

Questions? Contact us at: info@geosamples.org

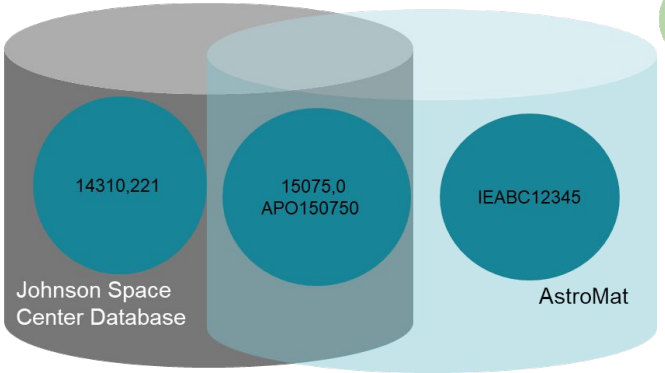


Supporting relationships

Linking to other digital objects



Same sample-different system



Parent-child relationships