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

Communities of Practice
Technology
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
Management System for Sample Metadata

My Home

Welcome, Sarah Ramdeen

REGISTRATION
- Register an individual sample
- Download batch registration template
- Upload my batch samples
- Update my existing samples profile

SAMPLES
- Search sample catalog
- View/Edit my samples
- View/Edit shared samples
- View/Edit my groups
- Upload files or images to samples

MY ACCOUNT
- Edit my account
- Transfer my samples to another user
- Set permissions for my user code

Pending Batch Registrations
There is a total of 2 batch registration(s) awaiting processing

- batch_1E93_083_1605543.063
  submitted on 2020-11-16 15:49:26Z
- batch_1E93_084_1605552343.66
  submitted on 2020-11-16 18:49:03Z

My Samples
You have a total of 585 registered samples in SESAR.
- 9 Grab
- 1 Hole
- 13 Core
- 1 Core Half Round
- 25 Core Piece
- 1 Core Sub-Piece
- 1 Core Core
- 497 Individual Sample

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
Catalog for Discovery and Access

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

<table>
<thead>
<tr>
<th>AVO Field</th>
<th>SESAR Field</th>
</tr>
</thead>
<tbody>
<tr>
<td>StationID</td>
<td>NA</td>
</tr>
<tr>
<td>SampleID</td>
<td>Sample Name</td>
</tr>
<tr>
<td>Latitude</td>
<td>Longitude</td>
</tr>
<tr>
<td>Longitude</td>
<td>Collector/Chief Scientist</td>
</tr>
<tr>
<td>DateVisited</td>
<td>Collection date</td>
</tr>
<tr>
<td>Volcano</td>
<td>Primary physiographic feature</td>
</tr>
<tr>
<td>Location</td>
<td>Location description</td>
</tr>
<tr>
<td>Description</td>
<td>Sample description</td>
</tr>
<tr>
<td>SampleType</td>
<td>Field name (informal classification)</td>
</tr>
<tr>
<td>AT GNC</td>
<td>Current archive</td>
</tr>
<tr>
<td>Material</td>
<td>Material</td>
</tr>
<tr>
<td>NA</td>
<td>Classification</td>
</tr>
<tr>
<td>NA</td>
<td>Country</td>
</tr>
<tr>
<td>NA</td>
<td>State/Province</td>
</tr>
<tr>
<td>NA</td>
<td>Collection method</td>
</tr>
</tbody>
</table>

**Repository Metadata Fields**

**SESAR Metadata Fields**
### Metadata mapping

<table>
<thead>
<tr>
<th>StationID</th>
<th>SampleID</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Geolatitude</th>
<th>Date/Visited</th>
<th>Volcano</th>
<th>Location Description</th>
<th>Sample Description</th>
<th>SampleType</th>
<th>AT QAC</th>
<th>Material</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSLO0001</td>
<td>OSLO0001A</td>
<td>51.9585</td>
<td>176.51265</td>
<td>Neat Christina A</td>
<td>09/26/05</td>
<td>Lilloe 8500</td>
<td>Dense clasts from the lava</td>
<td>yes</td>
<td>Thule-rock</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSLO0001</td>
<td>OSLO0001B</td>
<td>51.9585</td>
<td>176.51265</td>
<td>Neat Christina A</td>
<td>09/26/05</td>
<td>Lilloe 8500</td>
<td>Dense clasts from the lava</td>
<td>yes</td>
<td>Thule-rock</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSLO0001</td>
<td>OSLO0001C</td>
<td>51.9585</td>
<td>176.51265</td>
<td>Neat Christina A</td>
<td>09/26/05</td>
<td>Lilloe 8500</td>
<td>Dense clasts from the lava</td>
<td>yes</td>
<td>Thule-rock</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSLO0007</td>
<td>OSLO0007A</td>
<td>51.95359</td>
<td>176.64247</td>
<td>Neat Christina A</td>
<td>10/01/06</td>
<td>Lilloe 8500</td>
<td>Yoldiakava lava flow</td>
<td>yes</td>
<td>Thule-rock</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSLO0001</td>
<td>OSLO0001A</td>
<td>51.955217</td>
<td>175.518973</td>
<td>Loren Jacobs Fault</td>
<td>02/28/05</td>
<td>Lilloe 8500</td>
<td>south Rarle</td>
<td>Large dark blocky Bomb</td>
<td>yes</td>
<td>Thule-rock</td>
<td></td>
</tr>
<tr>
<td>OSLO0001</td>
<td>OSLO0001B</td>
<td>51.955217</td>
<td>175.518973</td>
<td>Loren Jacobs Fault</td>
<td>02/28/05</td>
<td>Lilloe 8500</td>
<td>south Rarle</td>
<td>Sample collected in lava</td>
<td>yes</td>
<td>Thule-rock</td>
<td></td>
</tr>
<tr>
<td>OSLO0002</td>
<td>OSLO0002</td>
<td>51.95496</td>
<td>175.5172</td>
<td>Loren Jacobs Fault</td>
<td>02/28/05</td>
<td>Lilloe 8500</td>
<td>south Rarle</td>
<td>Sample from a clear lava</td>
<td>yes</td>
<td>Thule-rock</td>
<td></td>
</tr>
<tr>
<td>OSLO0003</td>
<td>OSLO0003</td>
<td>51.95496</td>
<td>175.5144</td>
<td>Loren Jacobs Fault</td>
<td>02/28/05</td>
<td>Lilloe 8500</td>
<td>south Rarle</td>
<td>Juvenile sample from Bomb</td>
<td>yes</td>
<td>Thule-rock</td>
<td></td>
</tr>
<tr>
<td>OSLO0005</td>
<td>OSLO0005</td>
<td>51.926133</td>
<td>178.517033</td>
<td>Loren Jacobs Fault</td>
<td>02/28/05</td>
<td>Lilloe 8500</td>
<td>south Rarle</td>
<td>Juvenile clasts from the lava</td>
<td>yes</td>
<td>Thule-rock</td>
<td></td>
</tr>
<tr>
<td>OSLO0008</td>
<td>OSLO0008</td>
<td>51.926493</td>
<td>178.526443</td>
<td>Loren Jacobs Fault</td>
<td>02/28/05</td>
<td>Lilloe 8500</td>
<td>south Rarle</td>
<td>Anode lava flow from the lava</td>
<td>yes</td>
<td>Thule-rock</td>
<td></td>
</tr>
<tr>
<td>OSLO0009</td>
<td>OSLO0009</td>
<td>51.921217</td>
<td>178.528233</td>
<td>Loren Jacobs Fault</td>
<td>02/28/05</td>
<td>Lilloe 8500</td>
<td>south Rarle</td>
<td>Anode lava flow from the lava</td>
<td>yes</td>
<td>Thule-rock</td>
<td></td>
</tr>
<tr>
<td>OSLO0010</td>
<td>OSLO0010</td>
<td>51.924417</td>
<td>178.55333</td>
<td>Loren Jacobs Fault</td>
<td>02/28/05</td>
<td>Lilloe 8500</td>
<td>south Rarle beach</td>
<td>Light gray agpilite lava</td>
<td>yes</td>
<td>Thule-rock</td>
<td></td>
</tr>
</tbody>
</table>

Convert original data file to the SESAR Batch template.
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

Sample
IGSN: IEOAZ01169

Collecting Scientist
ORCID: 0000-0001-8700-9725

Publication
URN: idb:doi:28-cpus4-169981

Institution
ROR: 0f7hpc57

Core Piece
IGSN: BJO000001

Zircon Subsample
IGSN: BJO00000Q

Zircon Subsample
IGSN: BJO00000V

Same sample-different system

Parent-child relationships

14310,221
Johnson Space Center Database

15075.0
APO150750

IEABC12345
AstroMat

Parent-child relationships

BJO000001
Core Piece

BJO00000Q
Zircon Subsample

BJO00000V
Zircon Subsample

Linking to other digital objects

Supporting relationships