Getting Started with CZNet Hub
Interim Data Submission Guidance

The CZ Hub Team
October 27, 2021
Our Team

- **Kerstin Lehnert**: Lamont Doherty Earth Observatory, Columbia University
- **Jeff Horsburgh**: Civil and Environmental Engineering and Utah Water Research Laboratory, Utah State University
- **Chris Calloway**: Renaissance Computing Institute, University of North Carolina Chapel Hill
- **David Tarboton**: Civil and Environmental Engineering and Utah Water Research Laboratory, Utah State University
Lucy Profeta: Lamont-Doherty Earth Observatory, Columbia University

Shannon Syrstad: Utah Water Research Laboratory, Utah State University

Clara Cogswell: Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI)

User Support
Introductions

• Name
• Institution
• Thematic cluster affiliation
• Are you ready to start sharing data?
Our Objective

Provide a robust cyberinfrastructure for **Findable, Accessible, Interoperable, and Reusable** (FAIR) data from the CZ Network Thematic Clusters

Our Approach

• Link existing data facilities and services, including:
  • HydroShare
  • EarthChem
  • System for Earth Sample Registration (SESAR)
  • OpenTopography
  • Other systems, as needed

• Develop a central CZ Hub that provides
  • Services for easy data submission
  • Integrated data discovery and access
  • Computational resources for data analysis and visualization in support of CZ
    synthesis efforts
Our Approach

CZ Thematic Cluster Network
- Data collection
- Data aggregation
- Local data management
- Quality assurance/quality control
- Metadata creation

Data Submission Portal
- User support specialist
- Data curation support
- Documentation

Data Submission
- Metadata Templates
- Data Format Standards
- Controlled Vocabularies
- Data Upload Templates
- Sample Registration
- Unique Identifier Management

Repositories for Data and Research Products
- Permanent data archival and publication
- Access control for embargoed data
- Open access for public datasets
- Citable data

API
- EarthChem
  - Samples
  - Geochemistry
- SESAR
- HYDROSHARE
  - Time series
  - Geospatial
  - NetCDF
  - Models
  - Generic

Computational Resources
- Jupyter Notebook support
- Computation and modeling
- Synthesis studies
- Reproducible analyses

Catalog Services
- Cross-repository view of CZ data and research products
- Discovery based on authors, keywords, geographic area, time, Cluster
- Schema.org metadata implementation

Other Repositories
Thematic Cluster Network

• Thematic Cluster activities
  • Data collection
  • Data aggregation
  • Data QA/QC
  • Metadata creation

• Cluster data management plans
  • Flexibility for local data management

We aim to:
• Build a CZ Data Management Community
• Promote best practices for consistency
Data Submission Portal

- New, web based development to support network
- Data curation support
- Getting data to the right repository
- Submission directly through portal
- Maintain a list of submitted datasets/products

Empower you to curate data products within appropriate repositories with support from our team

Data Submission
- Metadata templates
- Data format standards
- Controlled vocabularies
- Data upload templates
- Sample registration
- Unique Identifier Management
Repositories for Data and Research Products

• No single repository will work for all CZ Data
• Operate and partner with existing repositories
  • Promote the use of FAIR principles
  • Permanent data archival and publication
  • Access control for embargoed data
  • Open access for public datasets
  • Citable data
  • Leverage existing NSF investment in CI
• User support specialists as Hub-level data managers

EarthChem
• Samples
• Geochemistry, mineralogy, XRD, etc.

HYDROSHARE
• Time Series
• Geospatial
• NetCDF
• Models
• Generic

OpenTopography
• LIDAR
• High resolution topography
• Elevation products

Other Repositories
Interim Data Submission/Sharing Guidance

• “Interim” because we are working on a data submission portal

• https://criticalzone.org/hub/interim-guidance-for-data-publication
How to Contact Us

• How to contact us
  • Join the CZNet Slack Workspace:
    • Email cznet@cuahsi.org and request to join and we will send you an invitation
    • “discuss-data” channel
    • Other channels as needed
  
  • Contact us via Email: cznet@cuahsi.org
    • Multiple people will see this email and get your question(s) routed to the right person
  
• Zoom meetings as needed – but let’s use the “Data Help Sessions”
Some General Guidance (1)

• Start assembling data now
  • What are the products your cluster will be creating/sharing?
  • Assemble and document as early as possible

• Tag your data with appropriate metadata
  • Title, abstract, keywords, spatial and temporal coverage, etc.
  • Start looking at the repository data submission forms to see what you are going to need
  • We’ll cover submission to HydroShare, EarthChem and Zenodo in upcoming webinars
Some General Guidance (2)

• Start thinking about the data/file formats you will use
  • Use open and accepted data formats
  • Where we have commonality of data across clusters, let’s standardize
  • We can use the ”Data Help Sessions” to discuss if needed

• Use a README file
  • More detailed information than what will fit in regular metadata

• The interim guidance website has links to some potentially useful data management resources

• Workshop 2 will hit on some highlights
Which Repository? (1)

• Data Submission Portal will support:
  • **HydroShare**: General purpose repository for water-science related datasets and models  
    [https://www.hydroshare.org](https://www.hydroshare.org)
  • **EarthChem**: Repository for submitting data derived from material samples such as soil, sediment, pore water, or rock specimens; cores; and other physical objects.  
    [https://www.earthchem.org/](https://www.earthchem.org/)
  • **Zenodo**: A catch-all repository that may be an appropriate place to upload content that is not appropriate for the other repositories  
    [https://www.zenodo.org/](https://www.zenodo.org/)
Which Repository? (2)

• Registration of samples with SESAR
  https://www.geosamples.org/

• High resolution topography data to OpenTopography
  https://opentopography.org/
HydroShare Data and Model Repository

• Operated by CUAHSI as a community resource
• Manage data (and other products) throughout the research life cycle
• Sharing and formal publication with citable digital object identifiers (DOIs)
• Much of the legacy CZO data has been moved to HydroShare

http://www.hydroshare.org
EarthChem

SESAR

• Preservation, discovery, access, and advanced mining of geochemical, petrological, geochronological, & mineralogical data and samples

EarthChem Library recommended by publishers as trusted data repository

• DOI assignment; long-term archiving; links to publications, funding awards, & samples

• Already hosts geochemical data from CZOs

EarthChem Synthesis: extract & integrate data across thousands of datasets

SESAR: register samples with IGSN; share sample metadata

Best Practices for geochemical & sample data (e.g. OneGeochemistry)
Which Repository? (3)

• Not all of the products your cluster creates may fit in those repositories

• Look for repositories that provide
  • A public landing page for your dataset/product
  • A unique identifier/DOI and URL for accessing the landing page
  • A formal citation for your dataset/product

• We will provide guidance on how to tell us about these datasets
Specific Guidance for Submissions (1)

• Regardless of which repository you submit to:
  • Tag your dataset/product with a subject keyword of “CZNet”
  • Ensure that you enter funding agency information,
    • Funding agency,
    • Award number
    • Award title
  • Make sure the dataset is publicly available

The Data Submission Portal will handle these things, but not all repositories will be supported by the Portal and/or you may submit data before the portal is ready
Specific Guidance for Submissions (2)

• Communities and Groups Functionality
  • **HydroShare**: Share your dataset/product with your thematic cluster Group. It will automatically become part of the CZNet Community
  • **EarthChem**: Associate your dataset/product with the “Critical Zone” community within EarthChem [https://earthchem.org/communities/cznet/](https://earthchem.org/communities/cznet/)
  • **Zenodo**: Associate your dataset/product with the “Critical Zone Data and Research Products” community [https://www.zenodo.org/communities/czdata/](https://www.zenodo.org/communities/czdata/)

• Register physical samples with SESAR [https://www.geosamples.org](https://www.geosamples.org)
Specific Guidance for Submissions (3)

• For datasets/products submitted to any other repositories:
  • Keep a list of submitted products
  • Identifiers/URLs/Citations
  • When the Data Submission Portal is complete, we will provide functionality for registering these datasets with the Portal
Upcoming Topics

• Every Wednesday at 4:00 PM Eastern
  • November 3: Data Help Session
  • November 10: Simple and Effective Methods for Managing and Sharing Scientific data
  • November 17: Data Help Session
  • November 24: Sharing Data/Research Products in HydroShare
  • December 1: Data Help Session
  • December 8: Submitting Data to EarthChem
  • January 5: Registering Samples with EarthChem
  • . . .
Questions?

Contact us:

cznet@cuahsi.org
Catalog Services

• Cross-repository view of CZ data and research products
• Discovery based on authors, geographic area, time, cluster
• Schema.org metadata
• Communities and Groups in HydroShare

A coordinated view and data discovery service(s) for all the data produced within the collaborative network to ensure that data are **Findable** and **Accessible**.
Catalog Services

• Cross-repository view of CZ data and research products
• Discovery based on authors, geographic area, time, cluster
• Schema.org metadata
• Communities and Groups in HydroShare

A coordinated view and data discovery service(s) for all the data produced within the collaborative network to ensure that data are Findable and Accessible.

CZO “Community” in HydroShare with individual “Groups” for each observatory
Computational Resources

- Repository linkages
- Jupyter Notebook support
- Computation and modeling
- Synthesis studies
- Reproducible analyses – sharing data, results, and the executable code used to generate/recreate them
- Enhanced trust in research through transparency, replicability and reproducibility

HydroShare is a web based gateway to computing