

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

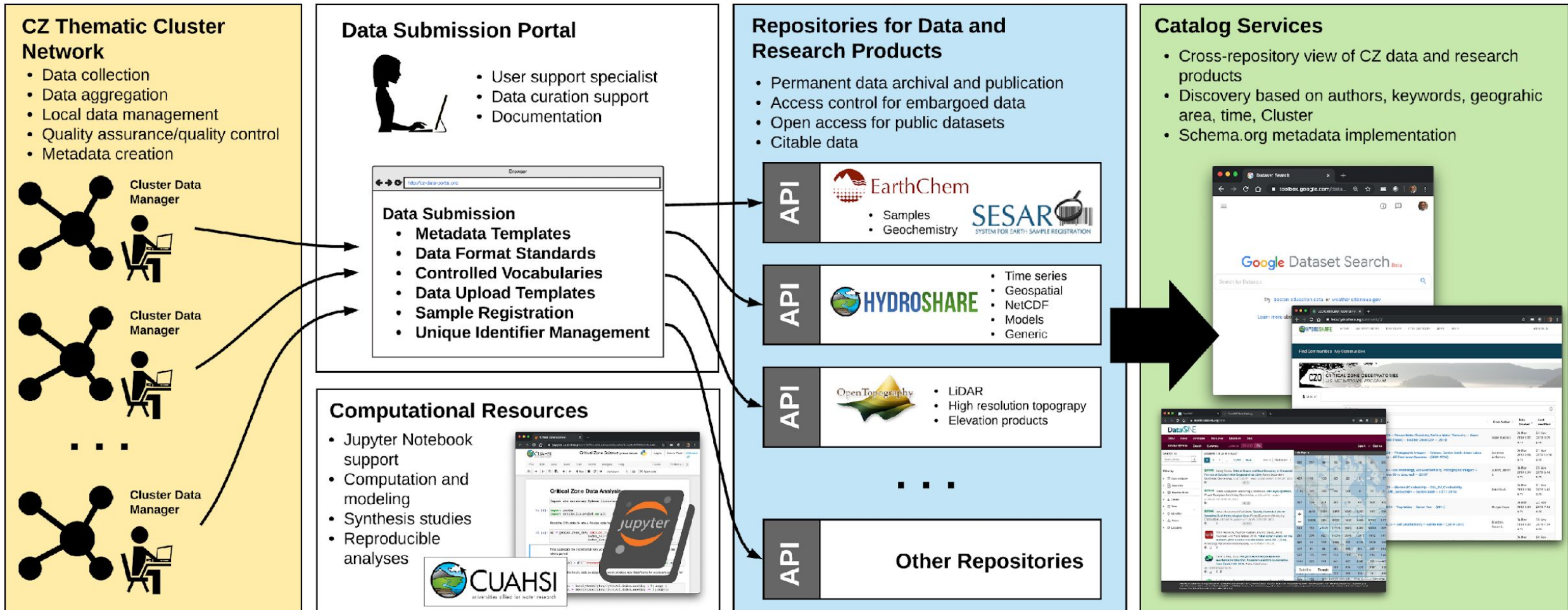
Provide a robust cyberinfrastructure for Findable, Accessible, Interoperable, and Reusable (FAIR) data from the CZ Network Thematic Clusters

Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3:160018, <https://doi.org/10.1038/sdata.2016.18>.

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

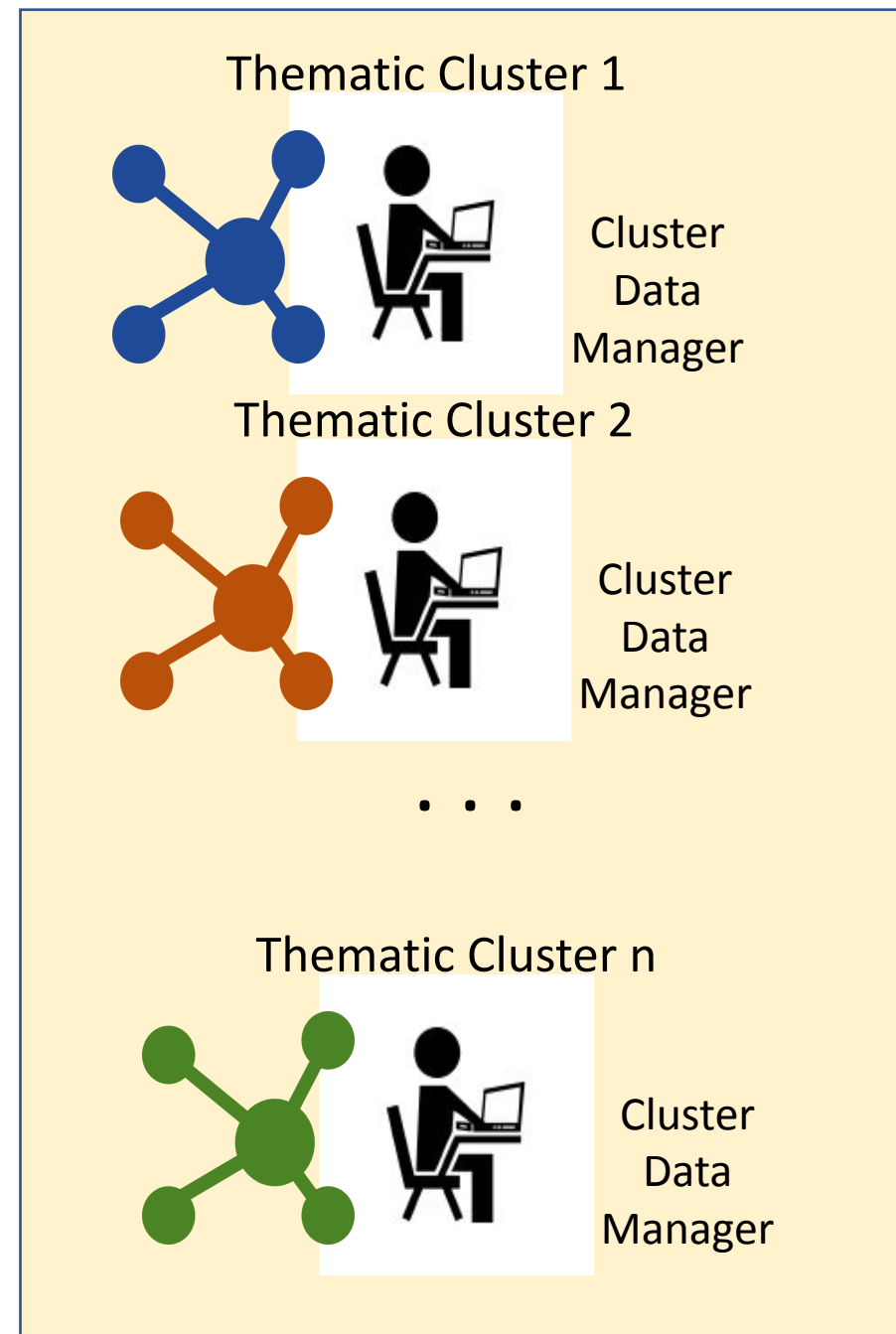


Thematic Cluster Network

- Thematic Cluster activities
 - Data collection
 - Data aggregation
 - Data QA/QC
 - Metadata creation
- Cluster data management plans
 - Flexibility for local data management

We aim to:

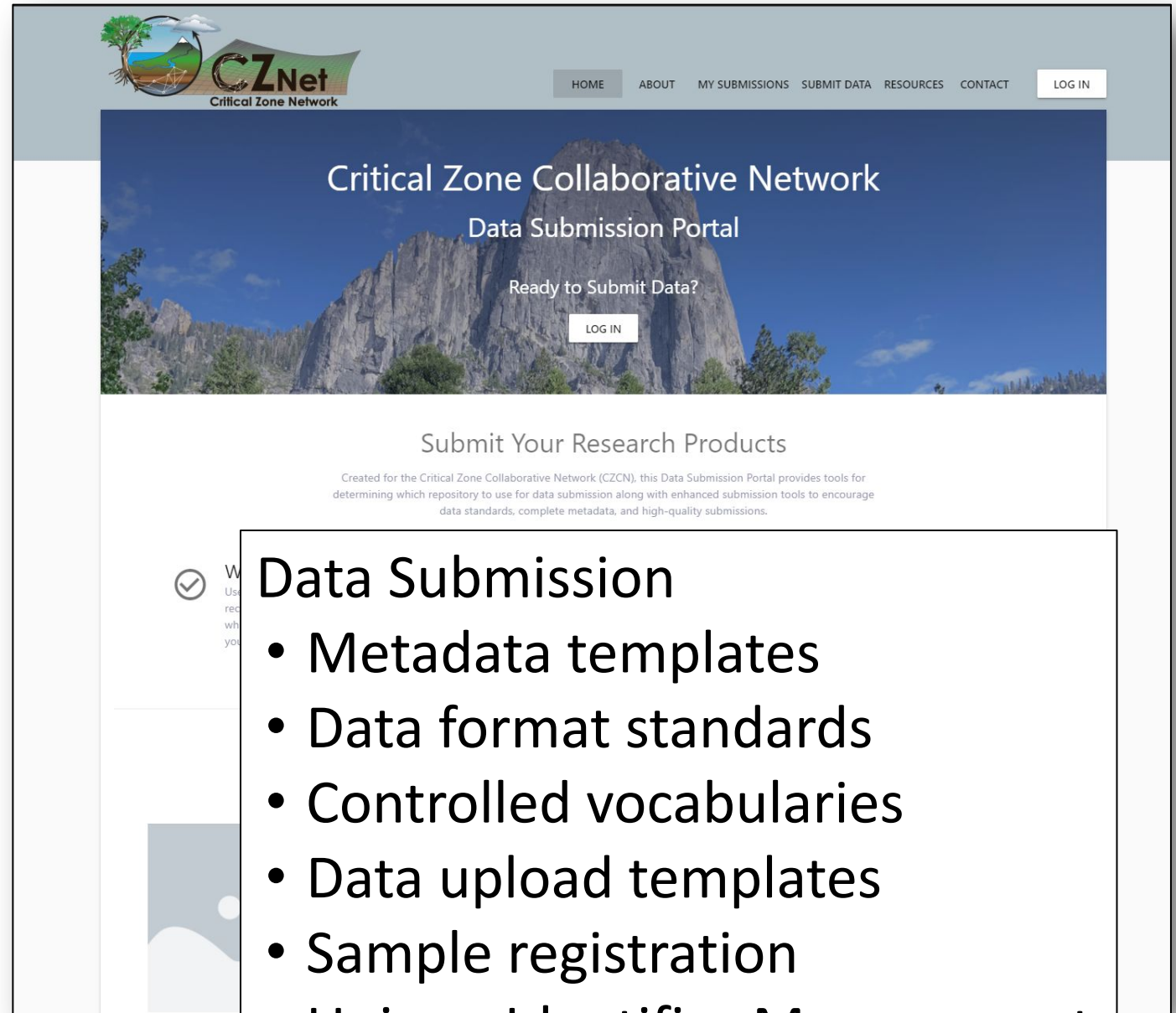
- Build a CZ Data Management Community
- Promote best practices for consistency



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

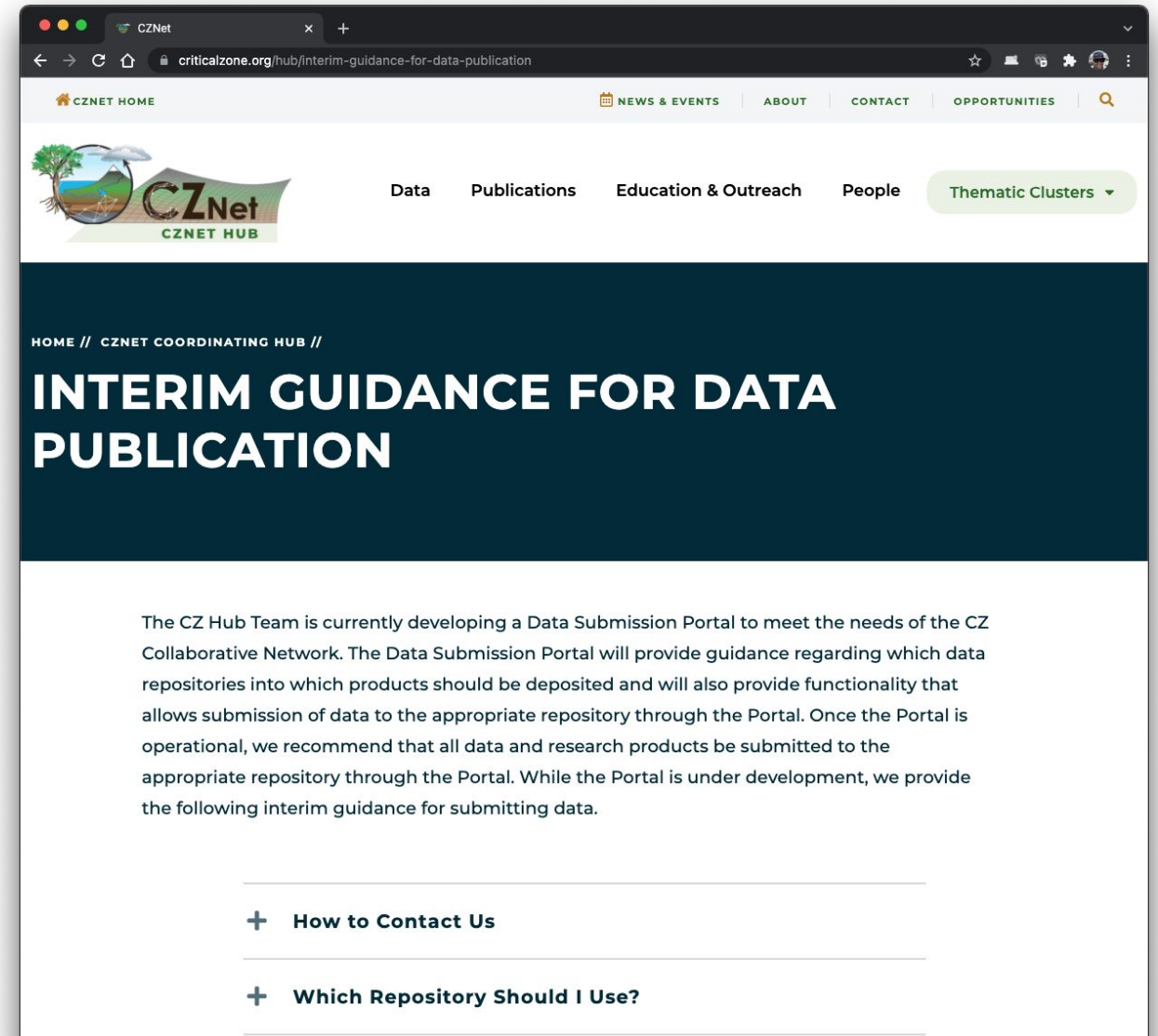
- LIDAR
- High resolution topography
- Elevation products

• • •

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

New Resource | CUAHSI Hydro

hydroshare.org/resource/60fc0d9a95884fa1b5ff096eee9e302d/

HYDROSHARE HOME MY RESOURCES DISCOVER COLLABORATE APPS HELP Create

This is the landing page for the resource you just created. Add files in the content area below and enter metadata where needed. The following metadata is required for your resource to be published or made public:

- Abstract
- Keywords

You must also add content files to your resource before it can be published, public or discoverable.

New Resource

Authors: + Jeffery S. Horsburgh ⓘ

Owners: Jeffery S. Horsburgh

Resource type: Composite Resource

Storage: The size of this resource is 0 bytes

Created: Oct 26, 2021 at 9:57 p.m.

Last updated: Oct 26, 2021 at 9:57 p.m. Jeffery S. Horsburgh

Citation: See how to cite this resource

Sharing Status: Private

Views: 0

Downloads: 0

+1 Votes: Be the first one to +1 this.

Comments: No comments (yet)

Abstract

Subject Keywords

Examples: Hydrologic_modeling, USU, land use Add

No subject keywords have been added.

Deleting all keywords will set the resource sharing status to private.

Coverage

You can set the spatial and temporal coverage manually by using the map to place a point or box or by filling in coordinates. Alternatively, you can add content files to your resource that have spatial coverage information (e.g., geographic feature, geographic raster, multidimensional, etc.) and then click the button to set the coverage from the content files.

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)

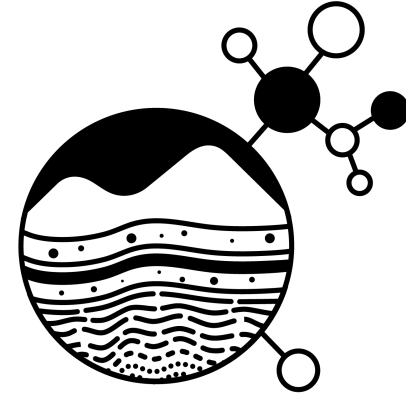
- Data Submission Portal will support:
 - **HydroShare**: General purpose repository for water-science related datasets and models
<https://www.hydroshare.org>
 - **EarthChem**: Repository for submitting data derived from material samples such as soil, sediment, pore water, or rock specimens; cores; and other physical objects.
<https://www.earthchem.org/>
 - **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/>



Which Repository? (2)

- Registration of samples with SESAR
<https://www.geosamples.org/>

- High resolution topography data to OpenTopography
<https://opentopography.org/>



SESAR

System for Earth
Sample Registration



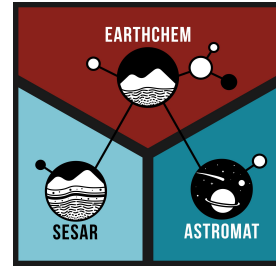
HydroShare Data and Model Repository

- 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>

The image displays two overlapping screenshots of the HydroShare website. The top screenshot shows the landing page with a navigation bar containing 'HOME', 'MY RESOURCES', 'DISCOVER', 'COLLABORATE', 'APPS', and 'HELP'. A large banner image shows a person working on a water sampling station in a forest. Below the banner, the text reads 'Collabo' and 'Upload and share a broad set of hydrologic data types including time content you st'. The bottom screenshot shows a resource page for 'Time Series Data for Dissolved Organic Matter Study in the Northwest Field Canal in Logan, UT'. The page includes metadata such as authors (Bryce Mihalevich, Jeffrey S. Horsburgh, Anthony A. Melcher), owners (Bryce A. Mihalevich, Jeffrey S. Horsburgh, JUTAH Data Manager), resource type (Composite Resource), storage (54.3 MB), creation date (Oct 18, 2017), and DOI (10.4211/hs.c1be74eeea614d65a29a185a66a7552f). It also features an abstract, subject keywords (Urban, Dissolved Organic Matter, time series, Stormwater, water quality), and a map showing the resource level coverage.

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)

EarthChem Library

← Return to ECL

Data DOI: [10.1594/IEDA/111144](https://doi.org/10.1594/IEDA/111144)

Citation: Chorover, J., Perdrial, J., McIntosh, J., Troch, P., Amistadi, M., Losleben, M., Condon, K., Pedron, S. 2018. Jemez River Basin Soil Solution Chemistry 2014 (New Mexico, USA), Version 1.0. Interdisciplinary Earth Data Alliance (IEDA). <https://doi.org/10.1594/IEDA/111144>, Accessed 2020-08-14.

Title: Jemez River Basin Soil Solution Chemistry 2014 (New Mexico, USA)

Creator(s): Chorover, J., Perdrial, J., McIntosh, J., Troch, P., Amistadi, M., Losleben, M., Condon, K., Pedron, S.

Abstract: Soil solution samples in the Jemez River Basin field sites of the Catalina-Jemez Critical Zone Observatory (CZO) are collected with the following two types of soil solution samplers: i) Prenart Super Quartz suction cups (www.prenart.dk). Prenart suction cups are optimized for all chemistry analyses and were installed without addition of Si-slurry to allow for artifact-free Si analyses. Applied suction for each Prenart is ~ 80 kPa. ii) Custom made, fiberglass wick-based passive capillary wick samplers (PCaps, Perdrial et al. 2012). PCaps are optimized for water flux determination and sampling for organic carbon, most (non-carbonate) anions and trace metals. PCap samples should however not be used for major cations (Na, Mg, Si, K, Ca) and dissolved inorganic carbon because of artifacts from the fiberglass materials (see Perdrial et al (2014) for a complete list). Passive (continuous) suction, based on the length of the hanging water column, is ~3 kPa. Soil solution samplers were installed in each of six pedons in the Mixed Conifer Zero Order Basin (MC-ZOB) and the fire impacted site (2011 Burned ZOB) at 3 (PCaps) and 4 (Prenarts) depths, respectively. Pedon locations were selected to capture differences in catchment aspect (MC-ZOB SE facing: Pit 3 and 4, NW facing: Pit 1 (MC-ZOB: hollow Pit 2 and 5, planar Pit 1 and 6, divergent Pit 3, convergent Pit 4), elevation and burn seve Burned ZOB: low Pit 1 and 2, mid Pit 3, high Pit 4 to 6). All samplers are co-located with Decagon soil moist Ridge fire) in June 2013 and was then renamed to 2013 Burned ZOB.

Keyword(s): Coverage Scope: Other
Geographic Location: Valles Caldera, Jemez River Basin, New Mexico

Related Publication(s): Perdrial, J.N., Perdrial, N., Harpold, A., Gao, X., LaSharr, K.M., Chorover, J. (2012) Impacts of sampling dis: Society of America Journal, 76: 2019-2030, doi: 10.2136/sssaj2012.0061.
Vazquez-Ortega, A., Perdrial, J., Harpold, A., Zapata-Rios, X., Rasmussen, C., McIntosh, J., Schaap, M., Pi reactive tracers of biogeochemical weathering in forested rhyolitic terrain. Chemical Geology, 391: 19-32, doi: 10.1016/j.chemgeo.2016.01.001.
Perdrial, J.N., Perdrial, N., Vazquez-Ortega, A., Porter, C., Leedy, J., and Chorover, J. (2014) Experimental Constituents. Soil Science Society of America Journal, 78(2): 486-495, doi: 10.2136/sssaj2013.07.0279.
Vazquez-Ortega, A., Huckle, D., Perdrial, J., Amistadi, M. K., Durick, M., Rasmussen, C., McIntosh, J., Chor to different hydrologic fluxes. Chemical Geology, 426: 1-18, doi: 10.1016/j.chemgeo.2016.01.001.

Publication DOI: <https://doi.org/10.2136/sssaj2012.0061>

User Contributed Keyword(s): Soil solution, Pore water, Soil Water, Cations, Anions, Metals, Water chemistry, pH, Carbon, Nitrogen, Fluoride, Sodium, Magnesium, Aluminum, Silicon, Potassium, Calcium, Titanium, Vanadium, Chromium, Manganese, Cadmium, Tin, Antimony, Barium, Lanthanum, Cerium, Praseodymium, Neodymium, Samarium, Europium, Lead, Uranium, Fluorescence, Humification index, Ultraviolet absorbance

Date Available: 2018-02-16

Date Created: ;

Resource Type:

File Format(s):

Funding source(s):

Related Information

IsReferencedBy: DOI: [10.2136/sssaj2012.0061](https://doi.org/10.2136/sssaj2012.0061)

References:

Sample ID: [IGSN: IECJCH800](#)
 Sample ID: [IGSN: IECJCH801](#)
 Sample ID: [IGSN: IECJCH802](#)
 Sample ID: [IGSN: IECJCH803](#)
 Sample ID: [IGSN: IECJCH804](#)
 Sample ID: [IGSN: IECJCH805](#)
 Sample ID: [IGSN: IECJCH806](#)
 Sample ID: [IGSN: IECJCH807](#)
 Sample ID: [IGSN: IECJCH808](#)
 Sample ID: [IGSN: IECJCH809](#)
 Sample ID: [IGSN: IECJCH810](#)
 Sample ID: [IGSN: IECJCH811](#)
 Sample ID: [IGSN: IECJCH812](#)
 Sample ID: [IGSN: IECJCH813](#)
 Sample ID: [IGSN: IECJCH814](#)
 Sample ID: [IGSN: IECJCH815](#)

IGSN: IECJCH800

IGSN: IECJCH800
Sample Name: JEMEZ_G918-Burn_Pit1-10cm
Other Name(s):
Sample Type: Individual Sample>Liquid
Parent IGSN: IECJCG008

Description

Material: Liquid>aqueous
Classification: Not Provided
Field Name: 2011 Burned ZOB Pedon 1
Description: Soil solution
Age (min): Not Provided
Age (max): Not Provided
Collection Method: Wick
Collection Method Description: Custom made, fiberglass wick-based passive capillary wick samplers (PCaps, Perdrial et al. 2012)
Size: 525 mL
Geological Age: Not Provided
Geological Unit: Not Provided
Comment: Analysis Notes: TL; Sampling Notes: evacuated until could no longer draw pressure; did not measure full volume;
Purpose: Not Provided

Geolocation

Latitude (WGS84): 35.850556
Longitude (WGS84): -106.450015
Northing (m) (UTM NAD83): 3968343
Easting (m) (UTM NAD83): 369064
Zone: 13N
Vertical Datum: NAVD88
Elevation Start: 2707.1 meters
Elevation End: Not Provided
Nav Type: GPS
Physiographic Feature: Resurgent dome
Name Of Physiographic Feature: Valles Caldera
Location Description: Not Provided
Locality: Rabbit Mountain
Locality Description: Not Provided
Country: United States
State/Province: New Mexico
County: Sandoval

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/>
 - **Zenodo**: Associate your dataset/product with the “Critical Zone Data and Research Products” community <https://www.zenodo.org/communities/czdata/>
- Register physical samples with SESAR <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.**

HydroShare datasets discoverable via Google Dataset Search

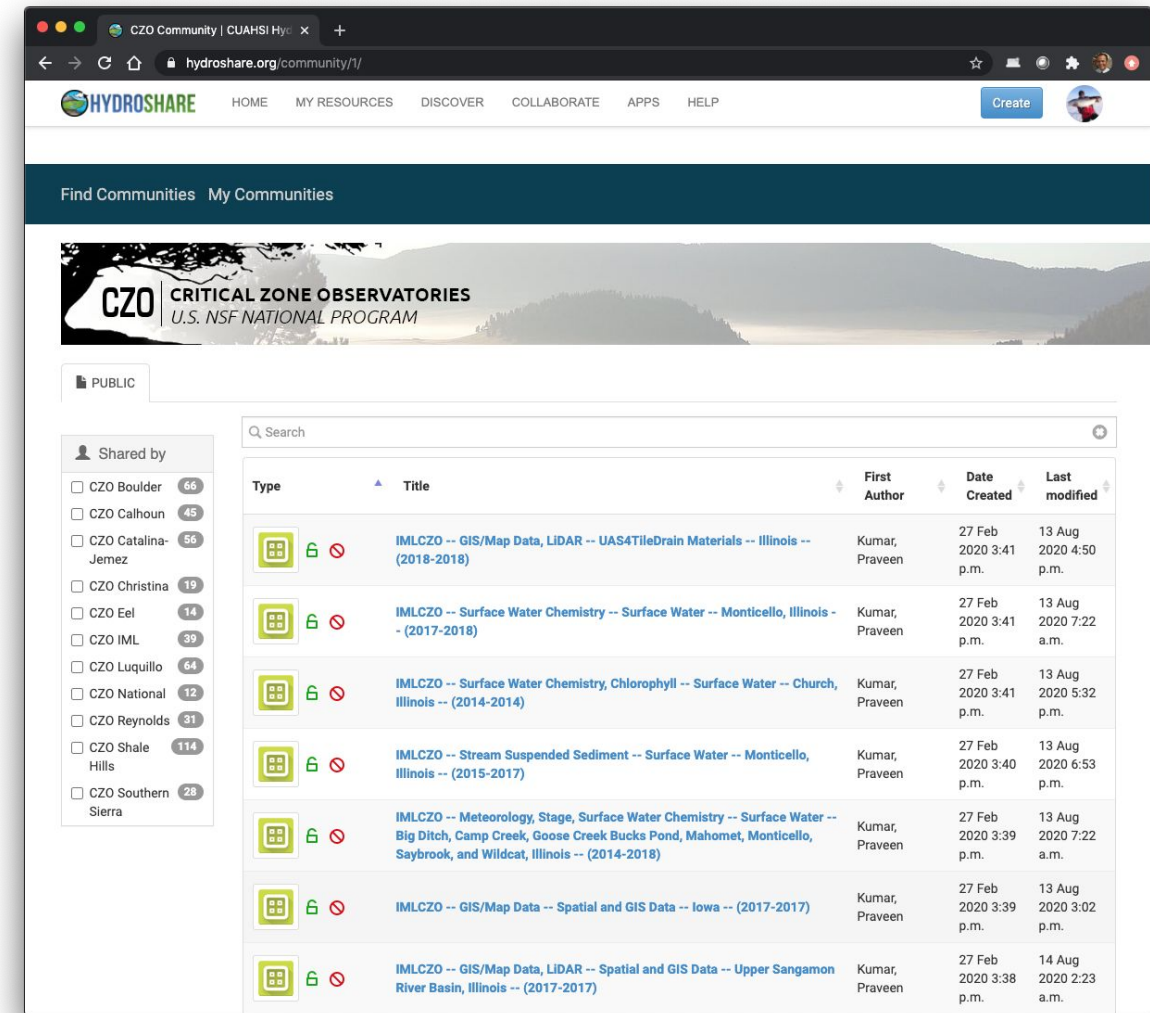
The screenshot shows a Google Dataset Search page for the query 'Logan River'. The search bar is circled in red. The results list several datasets, with the second one, 'Logan River Observatory: Logan River near Franklin Basin Aquatic Site (LR_FB_BA) Quality Controlled Data', highlighted with a red circle. A red arrow points from the 'Schema.org metadata' bullet point in the left slide to this dataset. The 'Explore at HydroShare' button for this dataset is also circled in red. The dataset details on the right include: 'Dataset updated Aug 5, 2020', 'Dataset provided by HydroShare', 'License Attribution 4.0 (CC BY 4.0)', 'Area covered', 'Dataset funded by National Science Foundation, State of Utah/Division of Water Resources, Utah State University, Logan City, Cache Water District', and a 'Description' section.

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



The screenshot shows the HydroShare interface for the CZO Community. The page features a navigation bar with links for HOME, MY RESOURCES, DISCOVER, COLLABORATE, APPS, and HELP. Below the navigation bar, there is a search bar and a list of data items. A red arrow points from the text 'Communities and Groups in HydroShare' to the 'Shared by' list on the left side of the page.

Type	Title	First Author	Date Created	Last modified
	IMLCZO -- GIS/Map Data, LIDAR -- UAS4TileDrain Materials -- Illinois -- (2018-2018)	Kumar, Praveen	27 Feb 2020 3:41 p.m.	13 Aug 2020 4:50 p.m.
	IMLCZO -- Surface Water Chemistry -- Surface Water -- Monticello, Illinois - (2017-2018)	Kumar, Praveen	27 Feb 2020 3:41 p.m.	13 Aug 2020 7:22 a.m.
	IMLCZO -- Surface Water Chemistry, Chlorophyll -- Surface Water -- Church, Illinois -- (2014-2014)	Kumar, Praveen	27 Feb 2020 3:41 p.m.	13 Aug 2020 5:32 p.m.
	IMLCZO -- Stream Suspended Sediment -- Surface Water -- Monticello, Illinois -- (2015-2017)	Kumar, Praveen	27 Feb 2020 3:40 p.m.	13 Aug 2020 6:53 p.m.
	IMLCZO -- Meteorology, Stage, Surface Water Chemistry -- Surface Water -- Big Ditch, Camp Creek, Goose Creek Bucks Pond, Mahomet, Monticello, Saybrook, and Wildcat, Illinois -- (2014-2018)	Kumar, Praveen	27 Feb 2020 3:39 p.m.	13 Aug 2020 7:22 a.m.
	IMLCZO -- GIS/Map Data -- Spatial and GIS Data -- Iowa -- (2017-2017)	Kumar, Praveen	27 Feb 2020 3:39 p.m.	13 Aug 2020 3:02 p.m.
	IMLCZO -- GIS/Map Data, LIDAR -- Spatial and GIS Data -- Upper Sangamon River Basin, Illinois -- (2017-2017)	Kumar, Praveen	27 Feb 2020 3:38 p.m.	14 Aug 2020 2:23 a.m.

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

The screenshot shows the HydroShare interface for a resource titled "Introduction to TauDEM". The page includes metadata such as authors (David Tarboton), owners, storage size (54.2 MB), and creation date (Dec 08, 2019). A red box highlights the "Open with..." dropdown menu, which lists several applications: HydroShare GIS, CUAHSI JupyterHub, OPeNDAP, CyberGIS-Jupyter for Water, and MATLAB Online.

The screenshot shows the Jupyter web interface's file browser. The file "TauDEM.ipynb" is highlighted with a red box. The interface shows the file is currently "Running" and was last modified 13 minutes ago. Other files like "logan.tif" and "logan.vrt" are also visible in the file list.

The screenshot shows the Jupyter notebook interface for the "TauDEM" notebook. The title is "Hydrologic Terrain Analysis Using TauDEM". The introductory text states: "The purpose of this notebook is to introduce **Terrain Analysis Using Digital Elevation Models (TauDEM)** software for Hydrologic Terrain Analysis in Jupyter. TauDEM is a free and open source set of Digital Elevation Model (DEM) tools for the extraction and analysis of hydrologic information from topography as represented by DEM. This software is developed at Utah State University (USU) for hydrologic digital elevation model analysis and watershed delineation."