

Meeting the Needs of Interdisciplinary Critical Zone Scientists by Leveraging and Linking Existing Domain Repositories

Jeffery S. Horsburgh

Utah State University

Kerstin Lehnert, Chris Calloway, Jerad Bales



Critical Zone Collaborative Network

- In 2020 NSF funded the next phase of their Critical Zone research program
- Nine Thematic Cluster study areas with a wide range of geological, climatic, and land use settings working to better understand the evolution and function of the Critical Zone
- One Coordinating Hub to help coordinate activities across Clusters – including data management



BEDROCK

Expanding knowledge of the deep critical zone and its feedbacks with surface processes.



COASTAL

Investigating the processes that transform landscapes and fluxes between land and sea.



DYNAMIC WATER

Advancing the understanding of the interactions among dynamic water storage, CZ processes, and water provisioning in western U.S. montane ecosystems.



BIG DATA

Using field observations, existing data, & advanced statistical and process-based tools to investigate how the Critical Zone responds to disturbances.



DRYLANDS

Quantifying and predicting dryland carbon budgets across land-use and climatic gradients.



GEOMICROBIO

Studying how soil microbes, roots, mineral composition, and soil organic matter interact and drive Critical Zone biogeochemistry and soil formation.



CINET

Investigating the role of critical interfaces for regulating the storage & transport of material such as water, sediment, carbon, & nutrients.



DUST²

A source-to-sink investigation of the dust system in the southwestern US as a component of the critical zone.

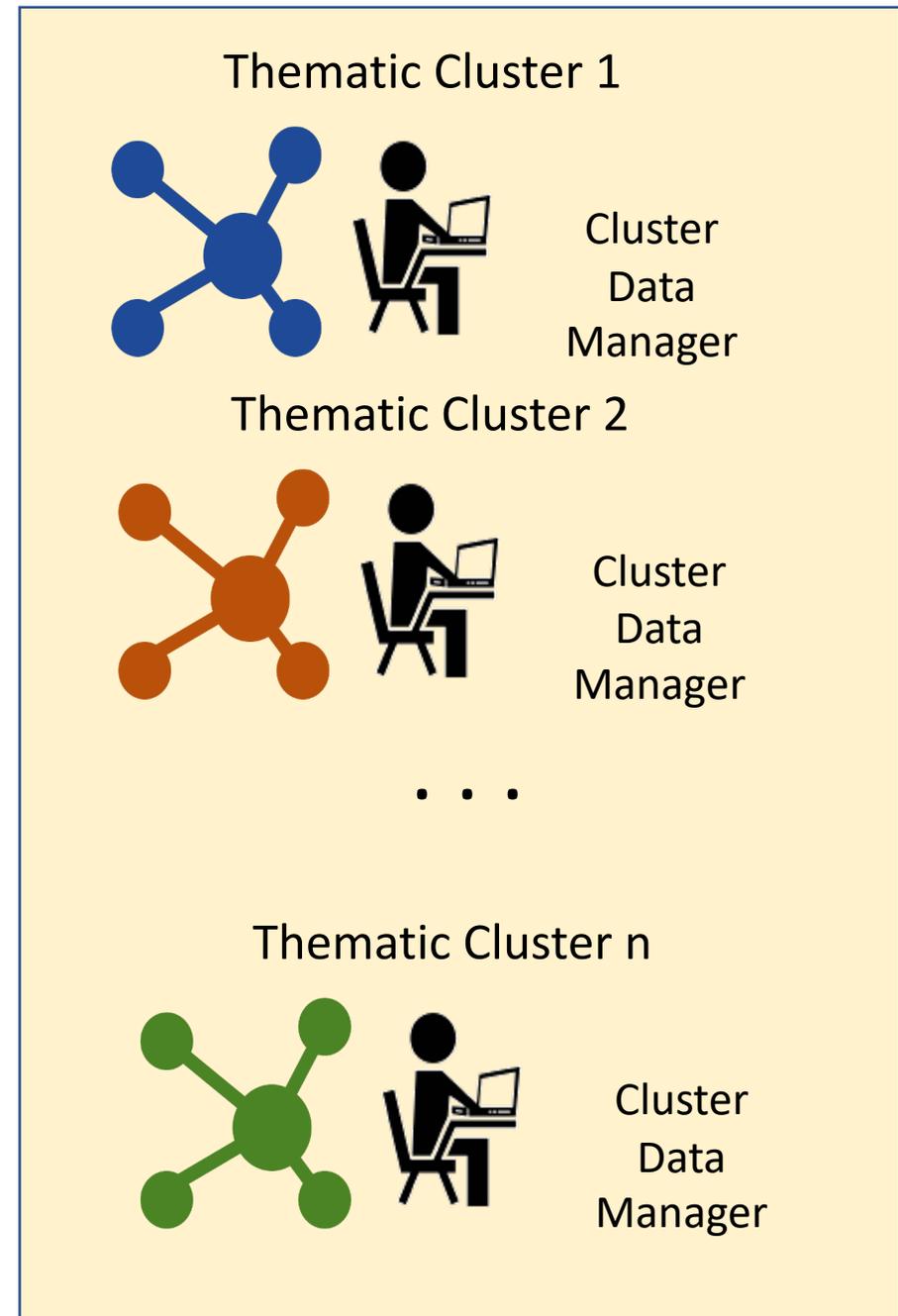


URBAN

Studying the interaction between the geologic template and the urban footprint and the effects on critical zone processes along the Eastern Seaboard.

Thematic Cluster Projects

- Activities
 - Data collection
 - Data aggregation
 - Data QA/QC
 - Defining data products
 - Metadata creation
- Data Management Plans
 - Each Thematic Cluster submitted their own Data Management Plan
 - Flexibility for local data management



Challenges

- Thematic Cluster teams and data are diverse
- Some are collecting new data, others are aggregating existing data, some are doing both
- No single data repository will meet the needs of interdisciplinary Critical Zone Scientists

Thematic Clusters	
1	Bedrock
2	Coastal
3	Dynamic Water
4	Big Data
5	Drylands
6	Geomicrobio
7	CINet
8	Dust ²
9	Urban

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

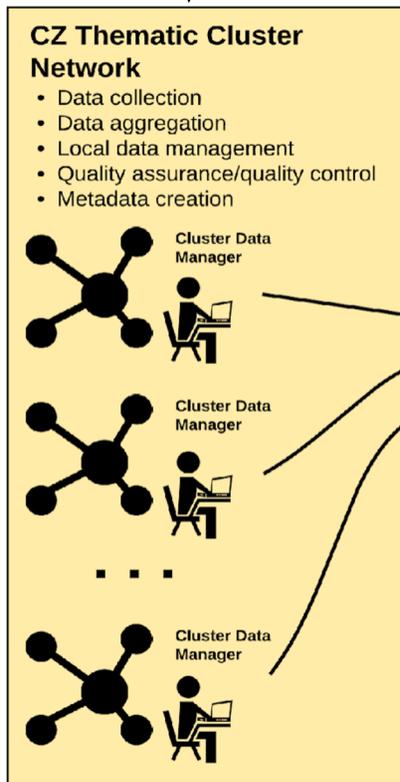
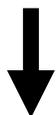
Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3:160018, <https://doi.org/10.1038/sdata.2016.18>.

CZ Hub Approach

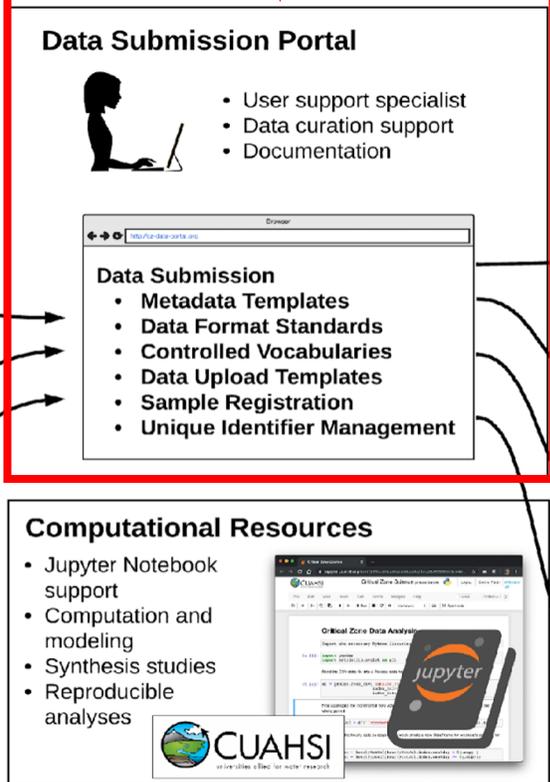
- Link existing data facilities and services, including:
 - HydroShare
 - EarthChem
 - System for Earth Sample Registration (SESAR)
 - OpenTopography
 - Other repositories, as needed
- Develop a central CZ Hub that provides
 - Services for easy data submission
 - Integrated data discovery and access

CZ Hub Approach

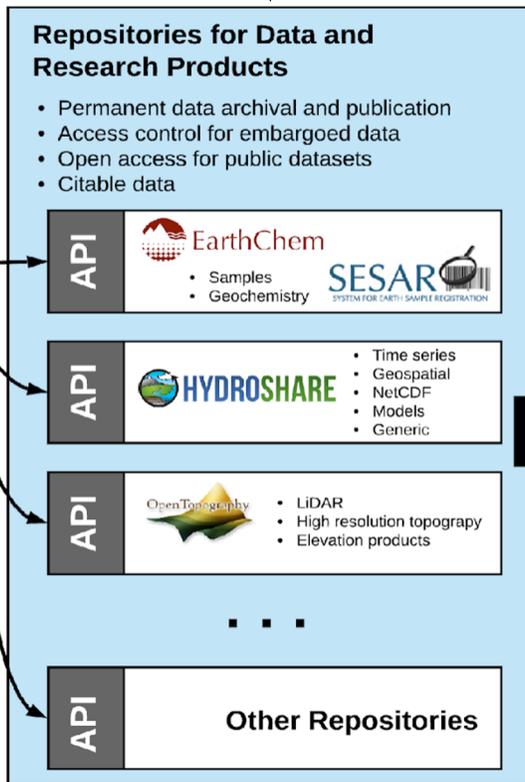
Diverse data and research products from CZ scientists



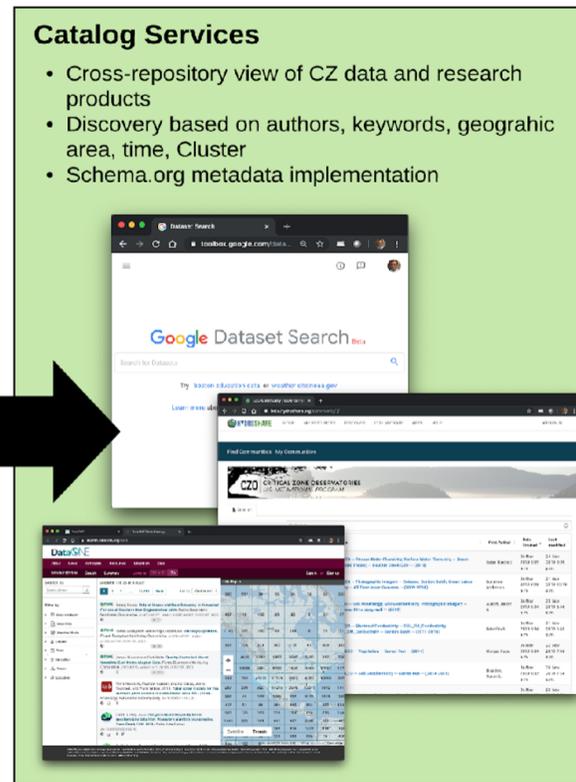
New Submission Portal promoting repository selection, metadata, templates, and formats



Using existing Earth science data repositories via automated submission



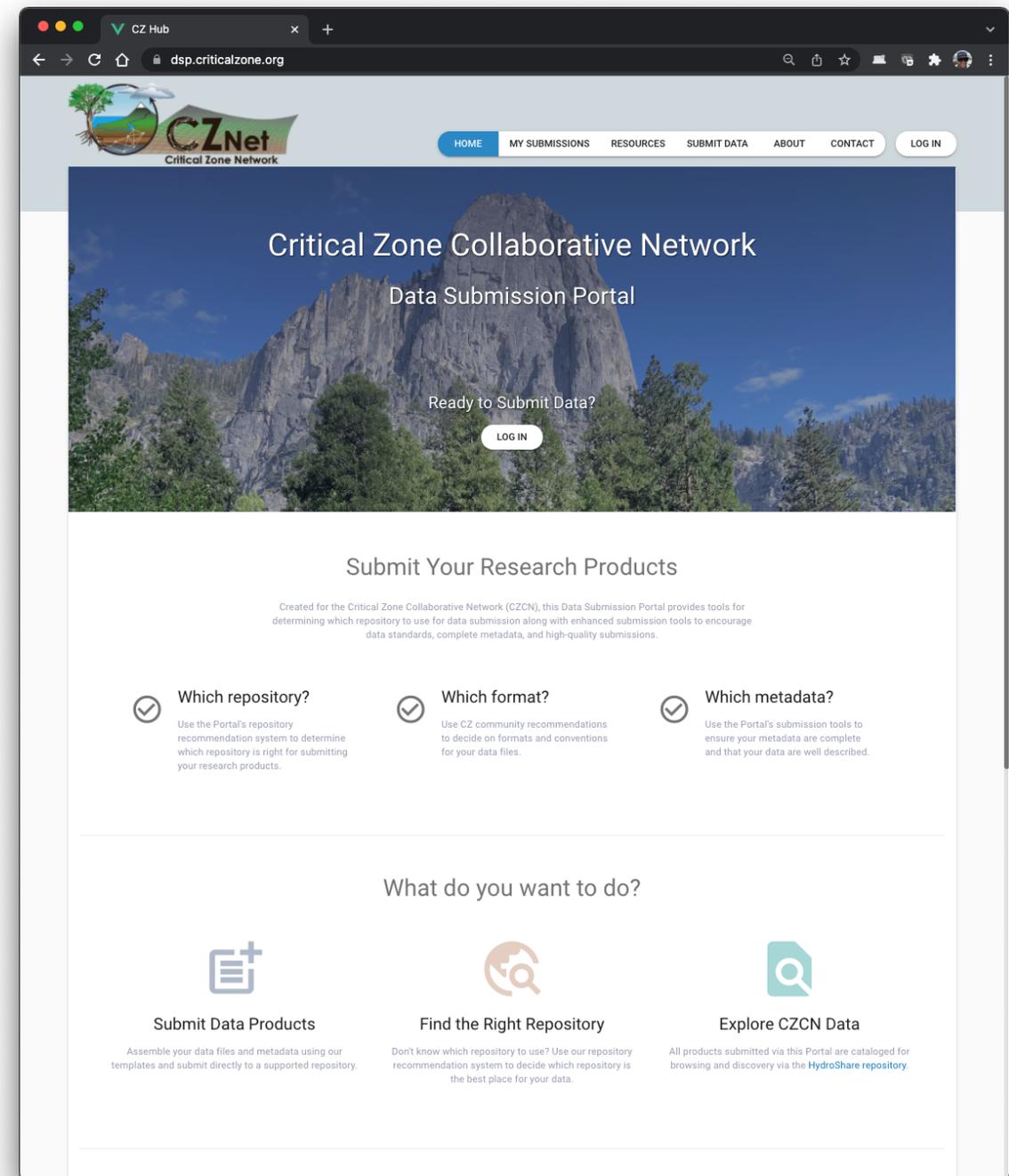
Making products Findable, Accessible, Interoperable, and Reusable (FAIR)



Data Submission Portal

- New, web application to support CZ Net
- Enables submission to multiple geoscience data repositories through one portal
- Getting data to the right repository
- Submission directly through portal

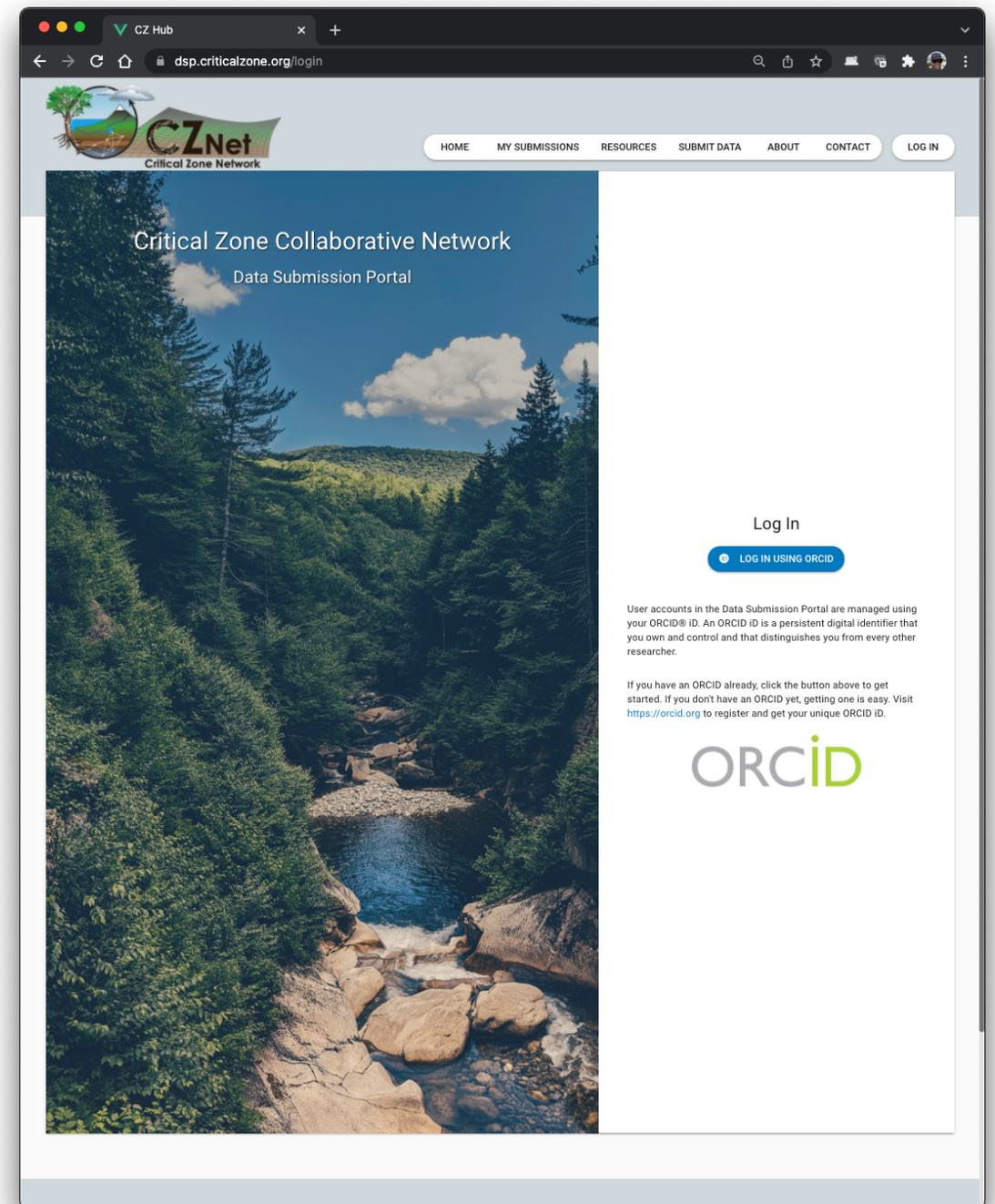
Empower data managers to curate research products within appropriate repositories with support from our team



User Account Management

- Authentication and user management via ORCID
- Using an account that most people have already
- Enables authentication across multiple repositories

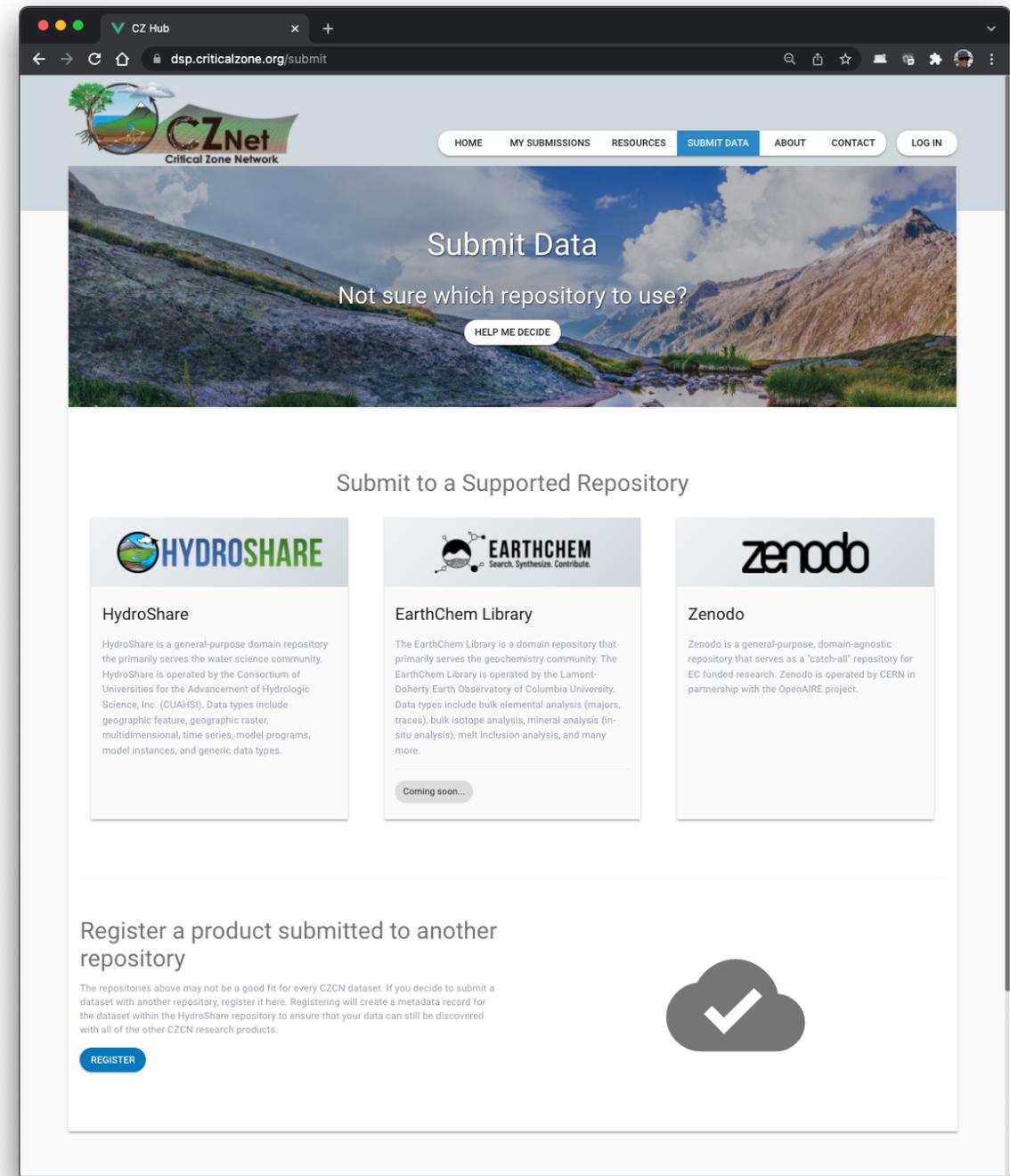
ORCID



The screenshot shows a web browser window with the URL `dsp.criticalzone.org/login`. The page features the CZNet logo (Critical Zone Network) and a navigation menu with links for HOME, MY SUBMISSIONS, RESOURCES, SUBMIT DATA, ABOUT, CONTACT, and LOG IN. The main content area has a large background image of a forest stream and the text "Critical Zone Collaborative Network Data Submission Portal". A "Log In" section contains a blue button labeled "LOG IN USING ORCID". Below this, a paragraph explains that user accounts are managed via ORCID iD and provides a link to <https://orcid.org> for registration.

Supported Repositories

- 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



The screenshot shows the 'Submit Data' page of the Critical Zone Network (CZNet) website. The page features a navigation bar with links for HOME, MY SUBMISSIONS, RESOURCES, SUBMIT DATA, ABOUT, CONTACT, and LOG IN. Below the navigation bar is a large banner image of a mountain landscape with a river, overlaid with the text 'Submit Data' and 'Not sure which repository to use?' with a 'HELP ME DECIDE' button. The main content area is titled 'Submit to a Supported Repository' and lists three repositories: HydroShare, EarthChem Library, and Zenodo. Each repository card includes a logo, name, and a brief description of its focus and data types. The EarthChem Library card has a 'Coming soon...' button. At the bottom, there is a section titled 'Register a product submitted to another repository' with a 'REGISTER' button and a cloud icon with a checkmark.

Submit Data
Not sure which repository to use?
HELP ME DECIDE

Submit to a Supported Repository

HYDROSHARE
HydroShare
HydroShare is a general-purpose domain repository that primarily serves the water science community. HydroShare is operated by the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI). Data types include geographic feature, geographic raster, multidimensional, time series, model programs, model instances, and generic data types.

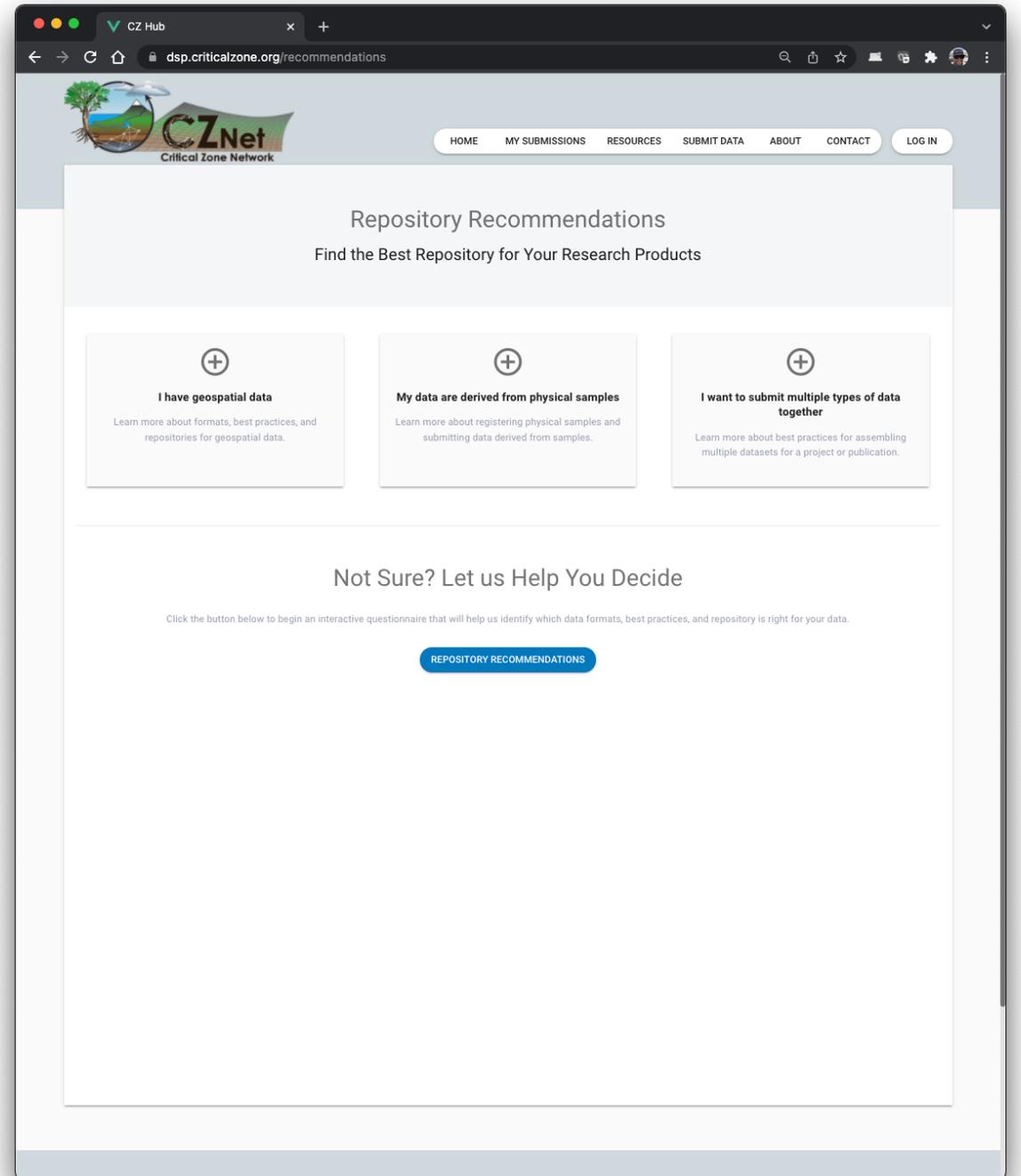
EARTHCHEM
EarthChem Library
The EarthChem Library is a domain repository that primarily serves the geochemistry community. The EarthChem Library is operated by the Lamont-Doherty Earth Observatory of Columbia University. Data types include bulk elemental analysis (majors, traces), bulk isotope analysis, mineral analysis (in-situ analysis), melt inclusion analysis, and many more.
Coming soon...

zenodo
Zenodo
Zenodo is a general-purpose, domain-agnostic repository that serves as a "catch-all" repository for EC funded research. Zenodo is operated by CERN in partnership with the OpenAIRE project.

Register a product submitted to another repository
The repositories above may not be a good fit for every CZCN dataset. If you decide to submit a dataset with another repository, register it here. Registering will create a metadata record for the dataset within the HydroShare repository to ensure that your data can still be discovered with all of the other CZCN research products.
REGISTER

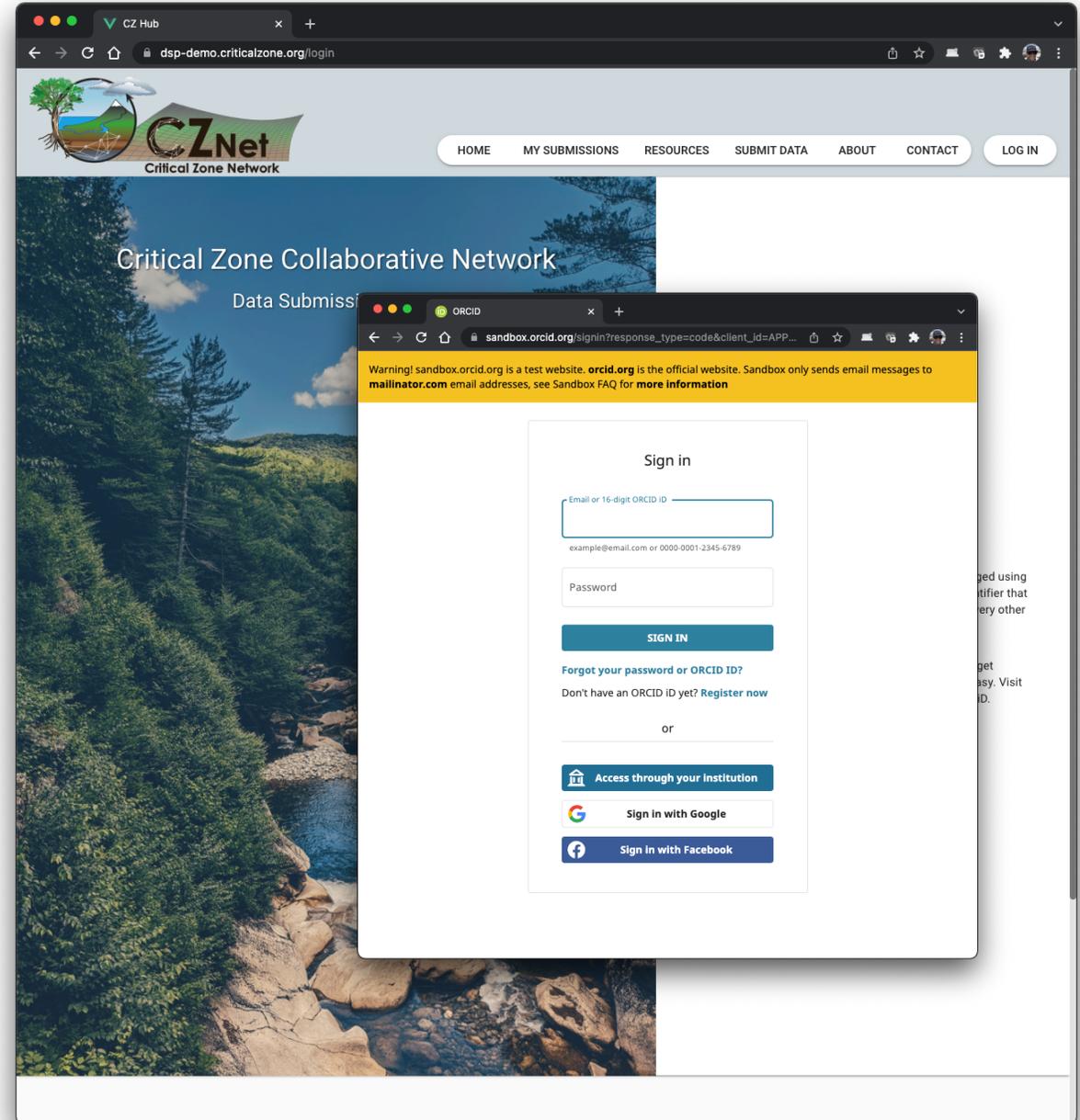
Which repository?

- Which repositories to target for different data types?
- Assist data managers in selecting an appropriate repository
 - Geospatial data
 - Data derived from physical samples
 - Hydrologic time series
 - Data and code packages
 - Models



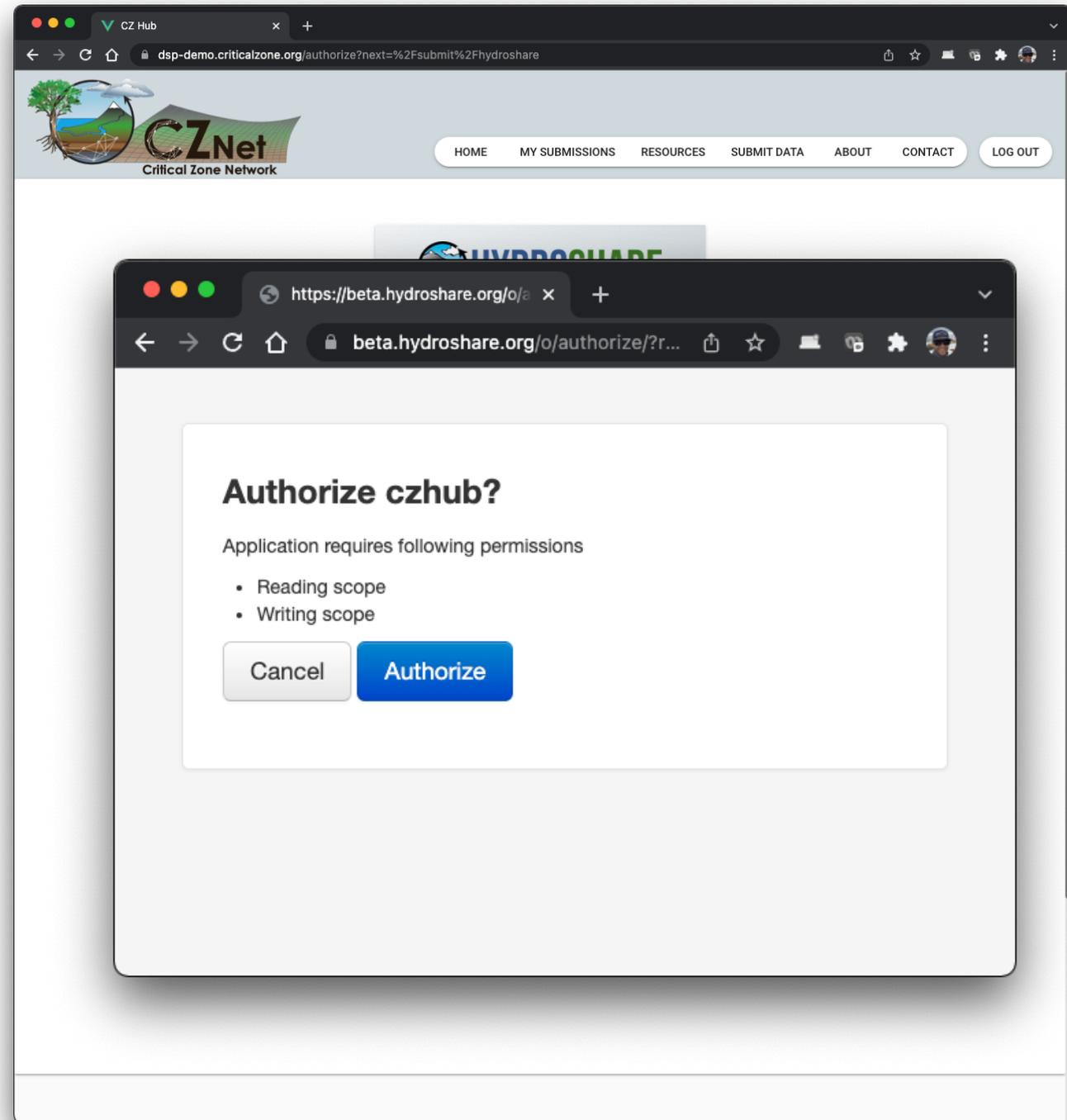
Step1: Log in

- User logs into the Portal using their ORCID
- User accounts in the Portal are associated with the ORCID



Step 2: Choose Repo

- User chooses a repository to submit to from the Submit Data page
- The user authorizes the Portal to submit to HydroShare
- The one-time authorization is stored in the user's profile



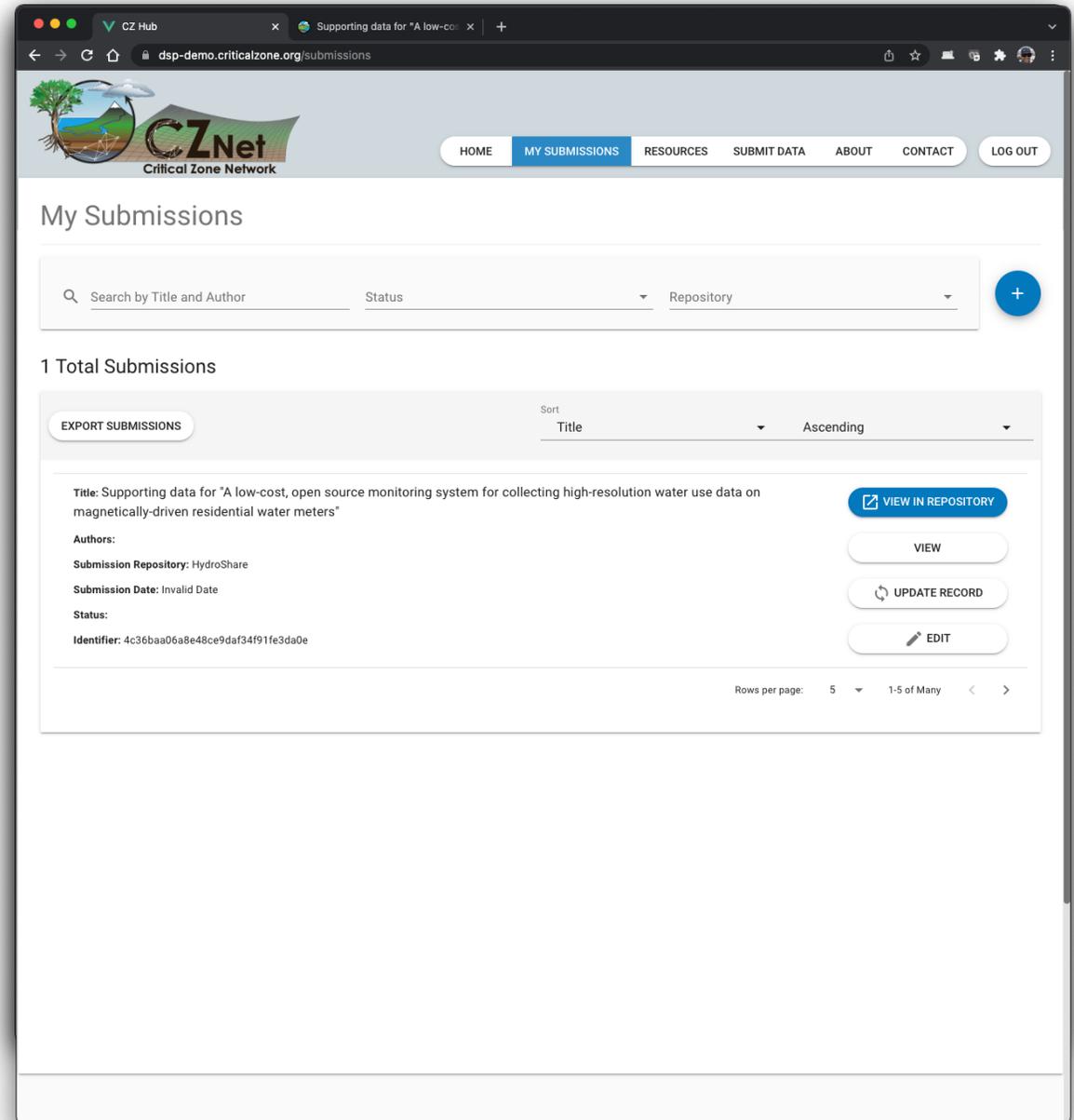
Step 3: Create Content

- User enters metadata and selects content files on the data submission form for the chosen repository
- Each repository has its own submission form
- Submission forms are built from a JSON schema that defines:
 - Required and optional metadata
 - Default values
 - Etc.

The screenshot shows a web browser window with the URL `dsp-demo.criticalzone.org/submit/hydroshare`. The page title is "New Submission" and it features the "HYDROSHARE" logo. A navigation bar at the top includes links for HOME, MY SUBMISSIONS, RESOURCES, SUBMIT DATA, ABOUT, CONTACT, and LOG OUT. The form includes a "Drop your files here or click to upload" area, followed by input fields for Title, Abstract, Subject keywords, Creators, Contributors, Sources, Related resources, and Rights Statement. The Rights Statement field is marked as required. A "UI SCHEMA" button and a "SAVE" button are located in the top right corner of the form area.

Step 4: Submit Content

- Metadata and data files are sent to the repository
- A new resource is created in the repository
- A new record is created in the user's My Submissions page
 - Return later to Edit
 - Export submissions to file



The screenshot displays the 'My Submissions' page on the CZ Hub website. The page features a navigation bar with links for HOME, MY SUBMISSIONS (highlighted), RESOURCES, SUBMIT DATA, ABOUT, CONTACT, and LOG OUT. Below the navigation bar, there is a search bar with a magnifying glass icon and a plus sign button. The search bar contains the text 'Search by Title and Author', 'Status', and 'Repository'. Below the search bar, it indicates '1 Total Submissions'. There is an 'EXPORT SUBMISSIONS' button and a 'Sort' dropdown menu set to 'Title' and 'Ascending'. The submission record is displayed with the following details:

- Title: Supporting data for "A low-cost, open source monitoring system for collecting high-resolution water use data on magnetically-driven residential water meters"
- Authors:
- Submission Repository: HydroShare
- Submission Date: Invalid Date
- Status:
- Identifier: 4c36baa06a8e48ce9daf34f91fe3da0e

On the right side of the record, there are four buttons: 'VIEW IN REPOSITORY' (with an external link icon), 'VIEW', 'UPDATE RECORD' (with a refresh icon), and 'EDIT' (with a pencil icon). At the bottom right, there is a 'Rows per page' dropdown set to '5' and a '1-5 of Many' indicator with navigation arrows.

JSON Schema-based Metadata

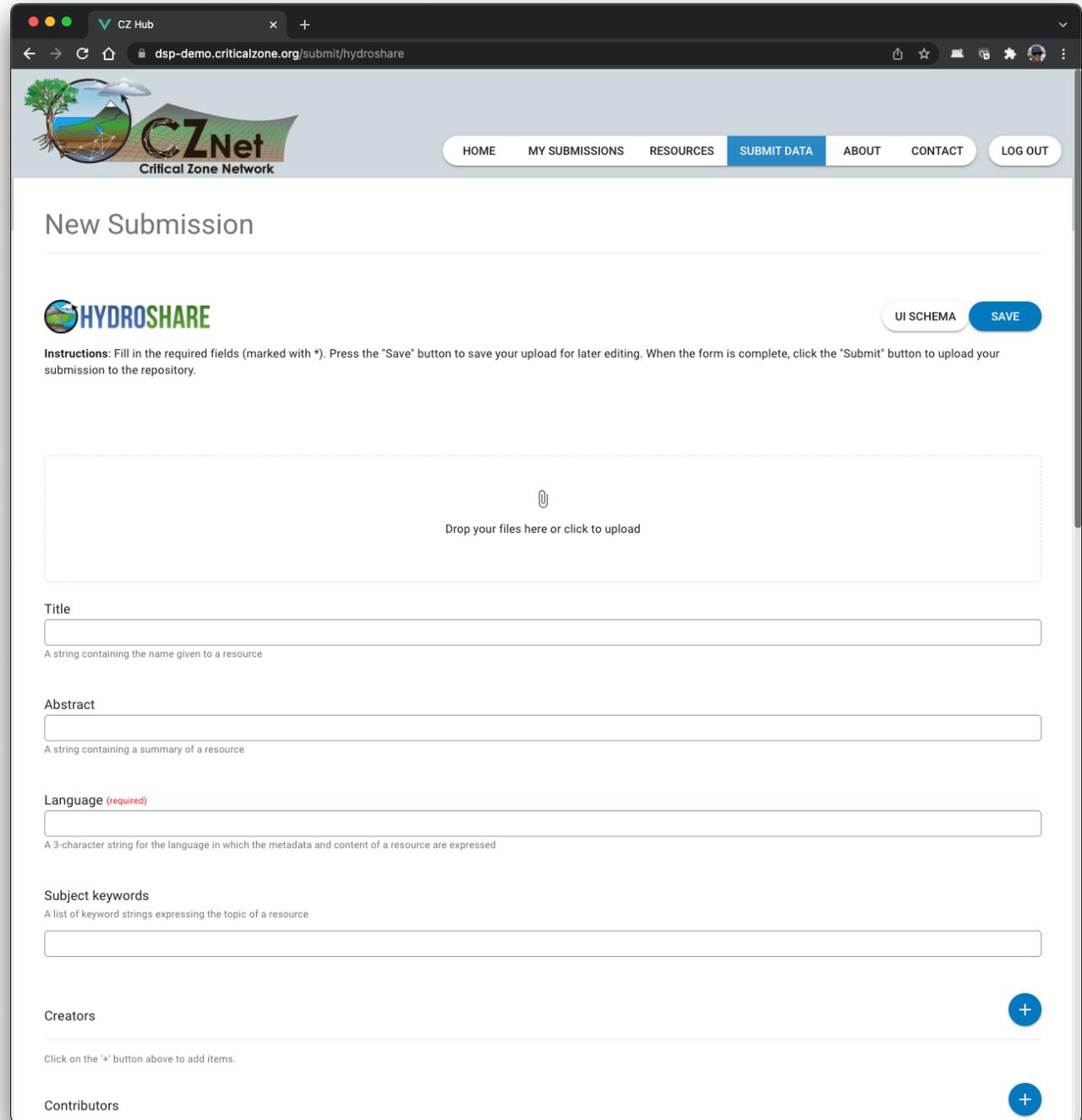
- A JSON schema defines required and optional metadata for each repository
- Submissions validated based on JSON schema
 - Data types
 - Default values
 - Required/optional
- Data submission form dynamically built from the JSON schema
- Adding a new repository to the Portal means adding a new JSON schema

```
{
  "title": "Resource Metadata",
  "description": "A class used to represent the metadata for a resource",
  "type": "object",
  "properties": {
    "title": {
      "title": "Title",
      "description": "A string containing the name given to a resource",
      "maxLength": 300,
      "type": "string"
    },
    "abstract": {
      "title": "Abstract",
      "description": "A string containing a summary of a resource",
      "type": "string"
    },
    "language": {
      "title": "Language",
      "description": "A 3-character string for the language in which the metadata and content of a resource",
      "type": "string"
    },
    "subjects": {
      "title": "Subject keywords",
      "description": "A list of keyword strings expressing the topic of a resource",
      "default": [],
      "type": "array",
      "items": {
        "type": "string"
      }
    },
    "creators": {
      "title": "Creators",
      "description": "A list of Creator objects indicating the entities responsible for creating a resource",
      "default": [],
      "type": "array",
      "items": {
        "$ref": "#/definitions/Creator"
      }
    },
    "contributors": {
      "title": "Contributors",
      "description": "A list of Contributor objects indicating the entities that contributed to a resource",
      "default": [],

```

JSON Schema-based Metadata

- Data submission form dynamically built from the JSON schema
- Files and metadata sent directly to repository
- Data Submission Portal maintains a record of submission



CZ Hub

dsp-demo.criticalzone.org/submit/hydroshare

HOME MY SUBMISSIONS RESOURCES **SUBMIT DATA** ABOUT CONTACT LOG OUT

New Submission

HYDROSHARE UI SCHEMA **SAVE**

Instructions: Fill in the required fields (marked with *). Press the "Save" button to save your upload for later editing. When the form is complete, click the "Submit" button to upload your submission to the repository.

Drop your files here or click to upload

Title
A string containing the name given to a resource

Abstract
A string containing a summary of a resource

Language (required)
A 3-character string for the language in which the metadata and content of a resource are expressed

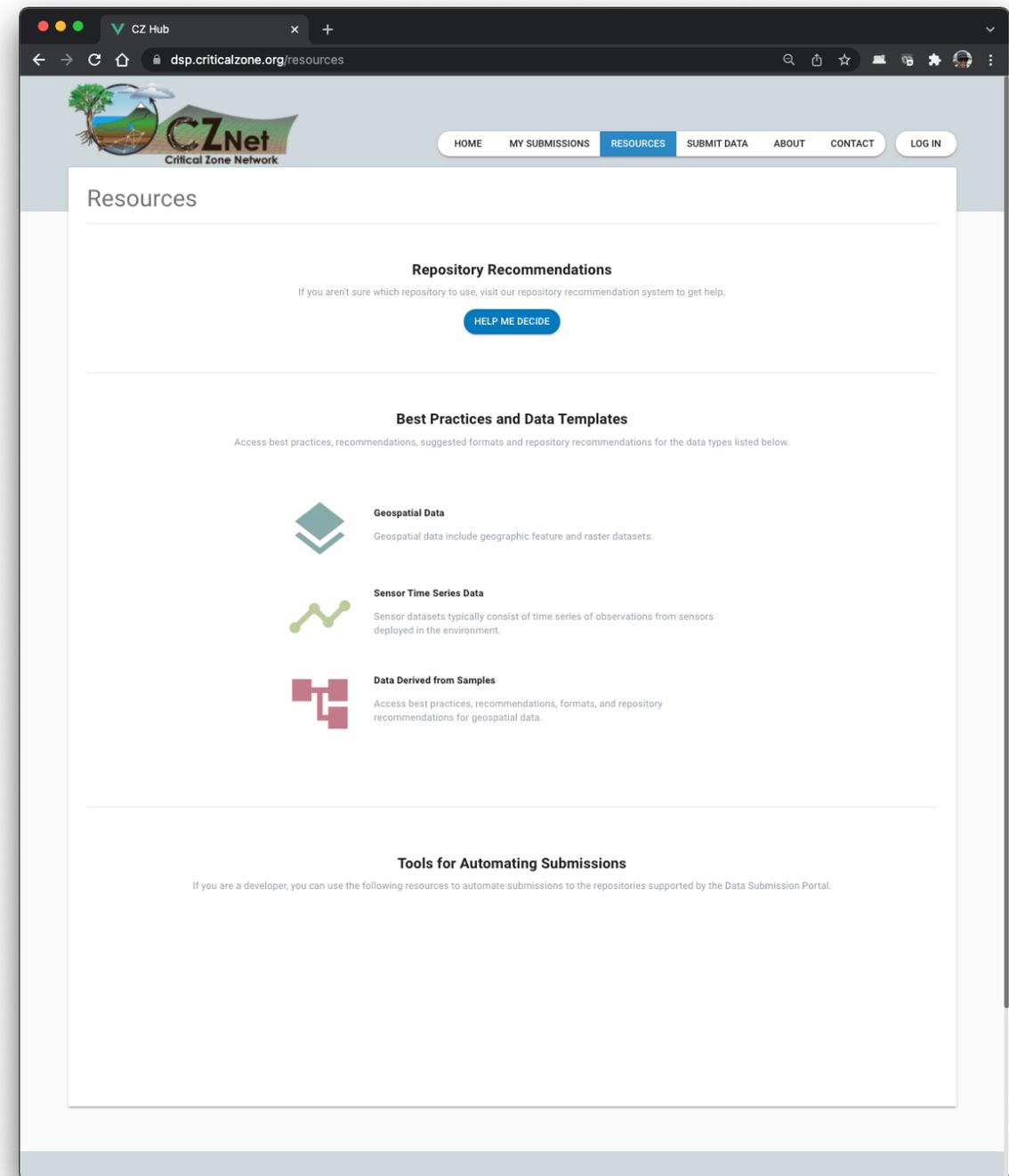
Subject keywords
A list of keyword strings expressing the topic of a resource

Creators +
Click on the "+" button above to add items.

Contributors +

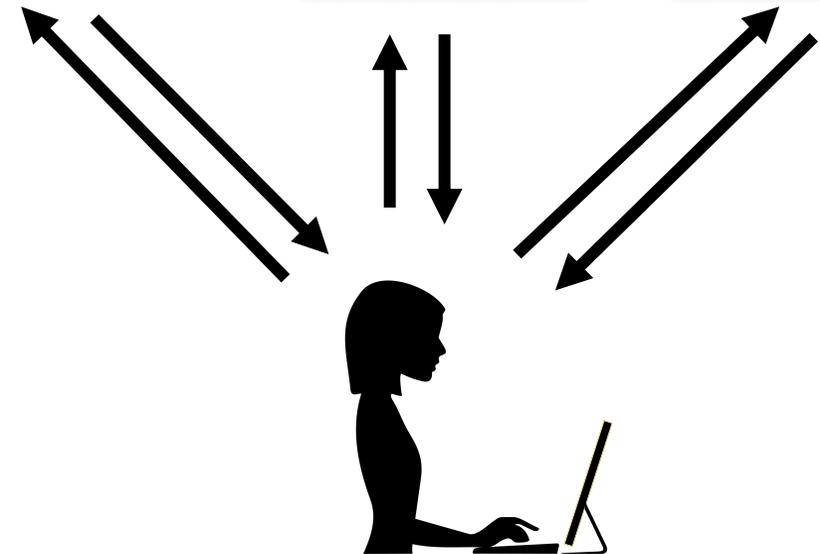
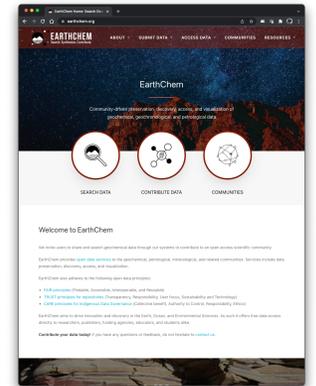
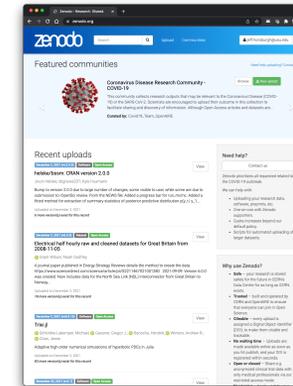
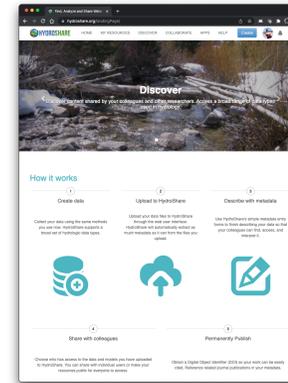
Promoting best practices

- Repository functionality is not specific to a community of users
- CZ Net may want to use:
 - Community standard formats
 - Templates
 - Best practices
- We can promote those through the Data Submission Portal



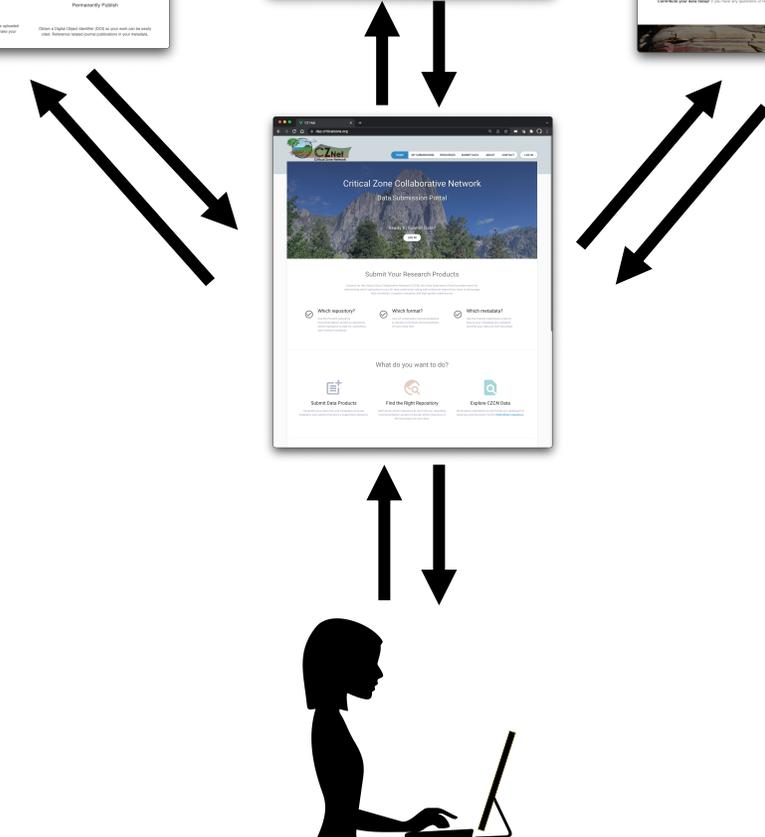
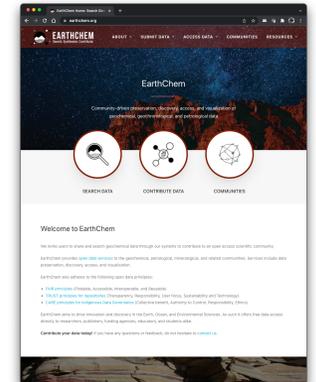
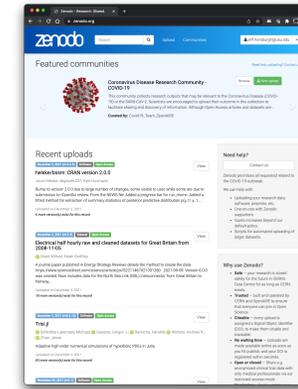
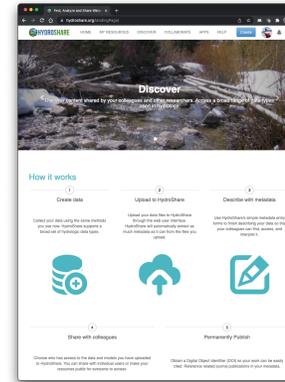
Without the DSP

- Data managers must navigate user interfaces of multiple systems
- Must keep track of what has been submitted to each one
- Difficult for CZ Hub Team to track what has been submitted



With the DSP

- Data managers need only interact through one user interface
- Submissions to all repositories tracked in one place
- Submissions automatically registered for cataloging/discovery



Data Submission Portal Advantages

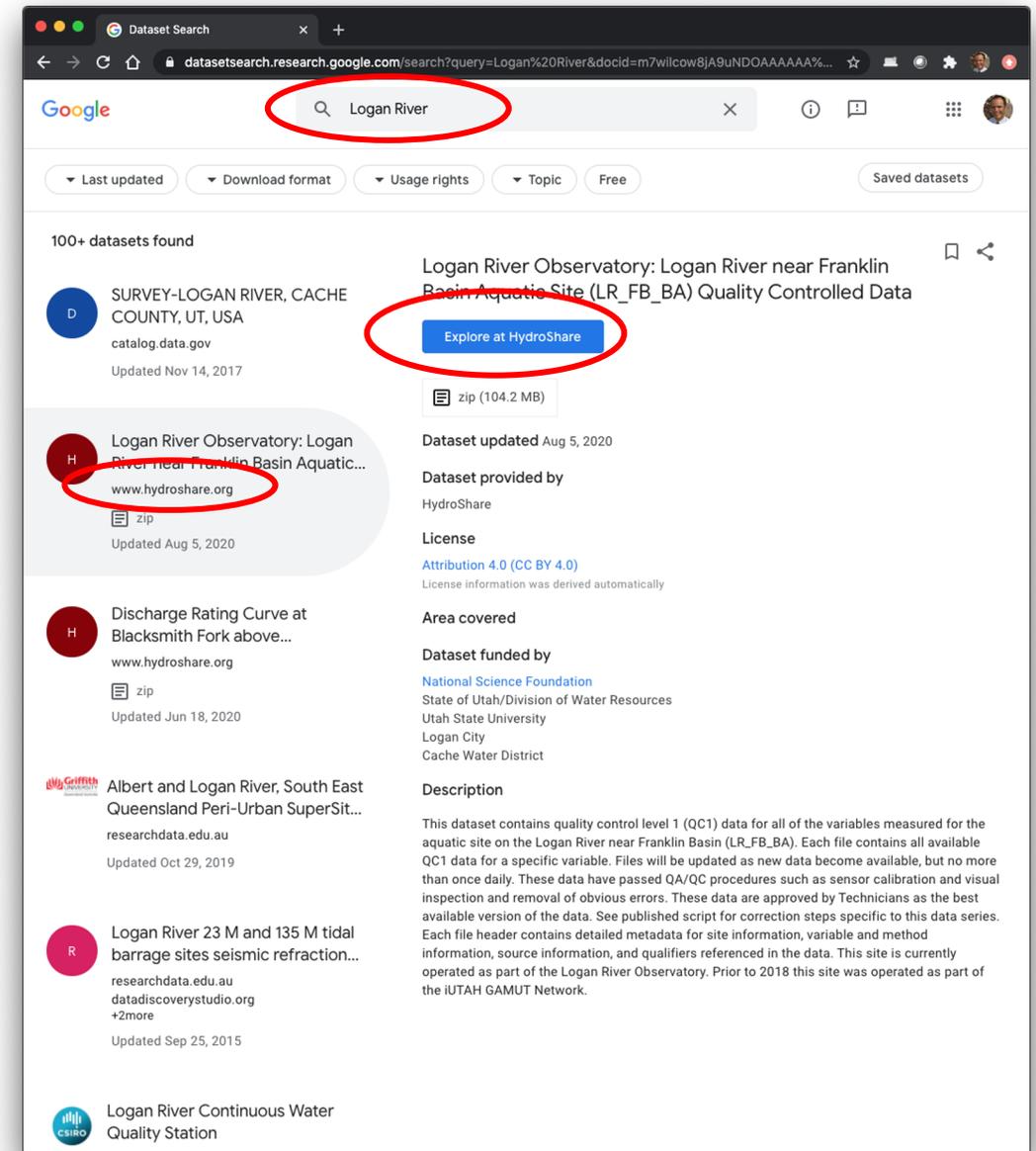
- Through validation, promote consistency in CZ Net data products across repositories
- Ensure data products end up in an appropriate, trusted repository
- Enforce minimum metadata requirements
 - Consistent keywords
 - Funding agency/grant information
- Enable use of controlled vocabularies where needed
- Promote templates, common formats, and best practices
- Enable data managers to use a single interface/tool to submit data
- Enable simple and consistent registration of CZ Net datasets with a metadata index for discovery
- Helping Thematic Clusters keep track of what has been submitted

CZ Net Catalog Services

- Cross-repository view of CZ Net data and research products
- Discovery based on authors, geographic area, time, cluster
- Schema.org metadata

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



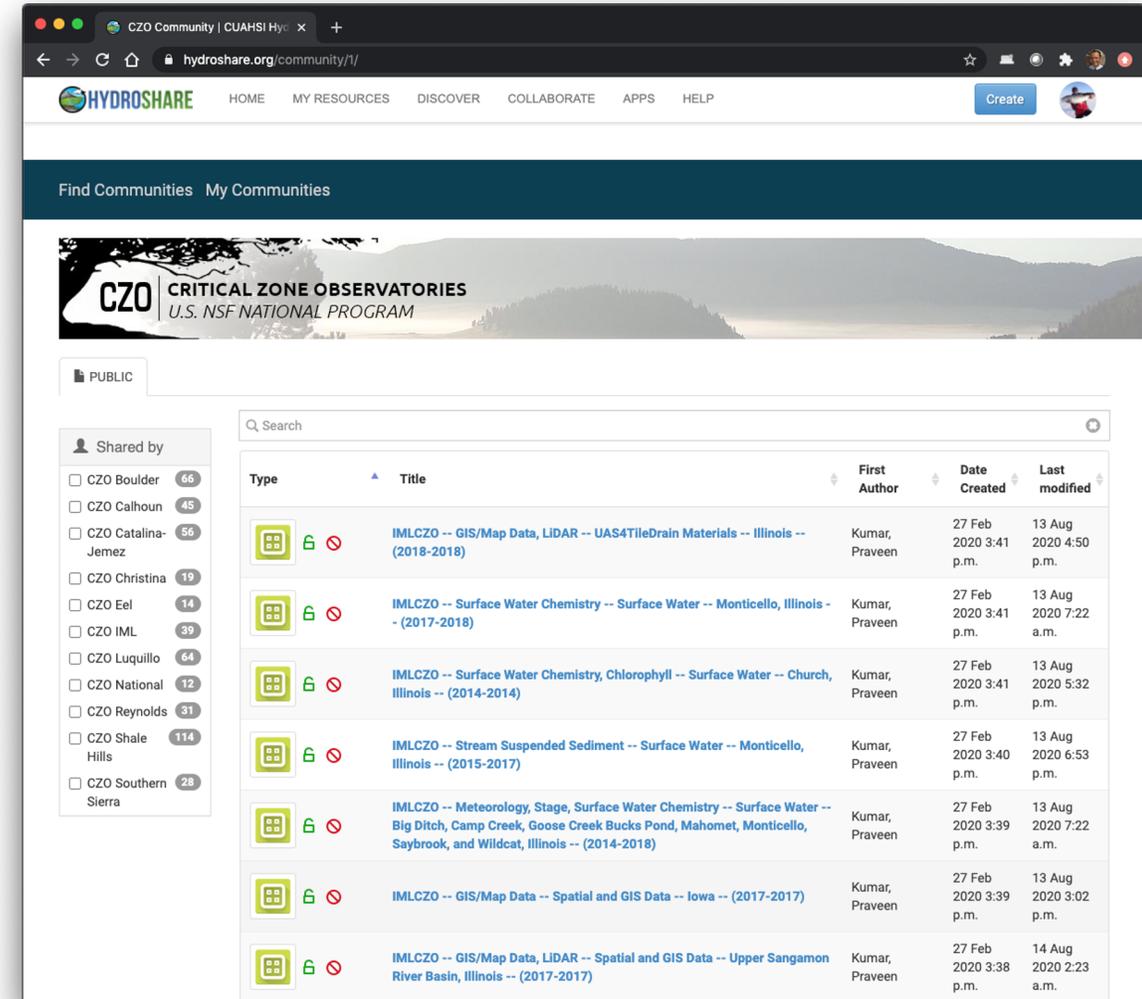
The screenshot shows a Google Dataset Search page for the query "Logan River". The search bar is circled in red. The results list several datasets, with the second one, "Logan River Observatory: Logan River near Franklin Basin Aquatic Site (LR_FB_BA) Quality Controlled Data", highlighted with a red circle. This dataset's "Explore at HydroShare" button is also circled in red. A red arrow points from the "Schema.org metadata" bullet point in the left slide to this dataset entry. 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.

CZ Net Catalog Services

- Cross-repository view of CZ Net 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



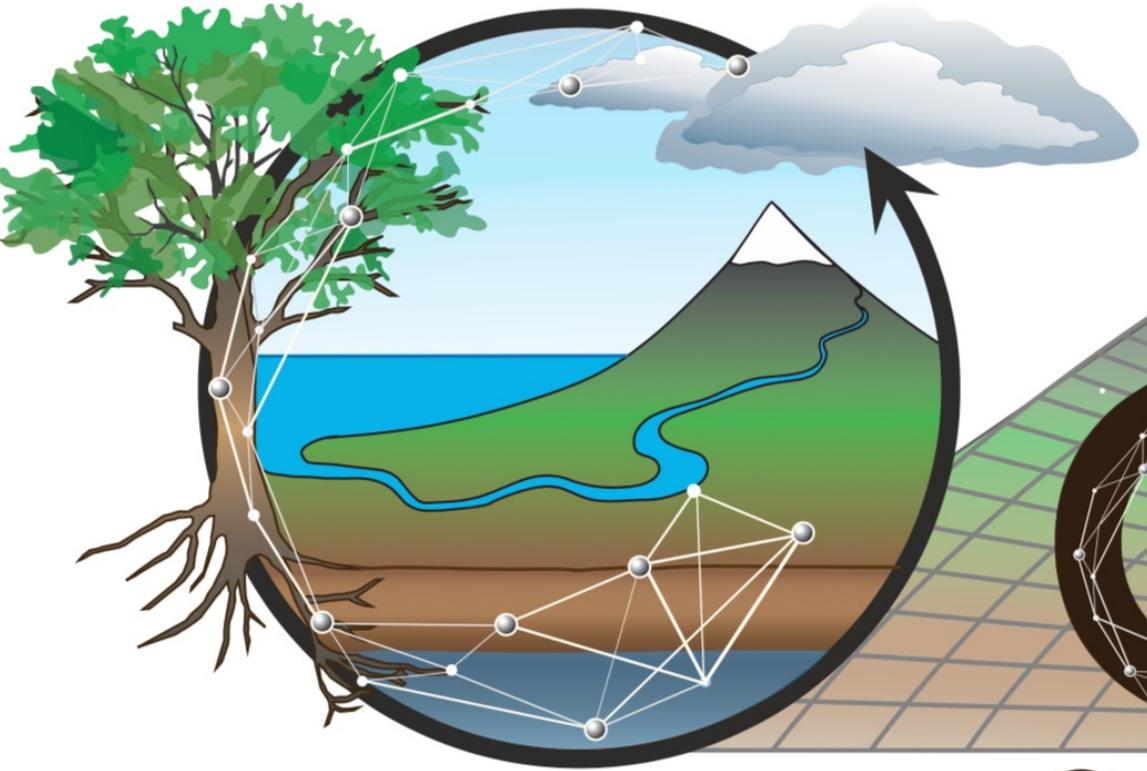
The screenshot shows the HydroShare interface for the CZO Community. The page features a navigation bar with links for HOME, MY RESOURCES, DISCOVER, COLLABORATE, APPS, and HELP. Below the navigation bar, there are tabs for "Find Communities" and "My Communities". The main content area displays a list of data items under the "PUBLIC" tab. The list is organized into columns: Type, Title, First Author, Date Created, and Last modified. A red arrow points from the text "Communities and Groups in HydroShare" to the "Shared by" sidebar on the left.

Type	Title	First Author	Date Created	Last modified
IMLCZO	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	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	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	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	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	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	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.



GitHub

The Data Submission Portal is on GitHub
<https://github.com/cznetHub>



CZNet

Critical Zone Network



Support: 2012893,
2012748, 2012593

Questions?

Jeffery S. Horsburgh
jeff.horsburgh@usu.edu