CDI Monthly Meeting 20210310

March 10, 2021: CDI FY20 Projects - Waterbody Rapid Assessment Tool (WaterRAT), FAIR practices for eDNA data in the USGS Nonindigenous Aquatic Species database, and a real-time coastal salinity index

The Community for Data Integration (CDI) meetings are held the 2nd Wednesday of each month from 11:00 a.m. to 12:30 p.m. Eastern Time.

Meeting Recording and Slides

Recordings and slides are available to CDI Members approximately 24 hours after the completion of the meeting.

These are the publicly available materials. Log in to view all the meeting resources. If you would like to become a member of CDI, join at https://listserv.usgs.gov/mailman/listinfo/cdi-all.

During the call, you can ask and up-vote questions at slido.com, event code #CDIMAR.

Agenda (in Eastern time)

11:00 am Welcome and Opening Announcements

11:15 am Working Group Announcements

11:25 am **Waterbody Rapid Assessment Tool (WaterRAT): 3-dimensional Visualization of High-Resolution Spatial Data,** Andrea Medenblik, USGS

11:45 am **Implementing FAIR practices: Storing and displaying eDNA data in the USGS Nonindigenous Aquatic Species database**, Jason Ferrante, USGS

12:05 pm **Real-time Coastal Salinity Index for monitoring coastal drought and ecological response to changing salinity values**, Matt Petkewich, USGS

12:30 pm *Adjourn*

Abstracts

The presentations this month are all from FY2020 CDI-supported projects.

Waterbody Rapid Assessment Tool (WaterRAT): 3-dimensional Visualization of High-Resolution Spatial Data

Autonomous Underwater Vehicles (AUVs) are instruments that collect water-quality, depth, and other data in waterbodies. They produce complex and massive datasets. There is currently no standard method to store, organize, process, quality-check, analyze, or visualize this data. The Waterbody Rapid Assessment Tool (WaterRAT) is aPython application that processes and displays water-quality data with interactive two-dimensional and three-dimensional figures, but it runs offline with few capabilities and for just one study site. This project will transition WaterRAT to an online application that the public can easily use to view all AUV data. A database of all AUV datasets will be developed to improve accessibility, organization, consistency of publication procedures, and publication time. Stakeholders and the public will be able to quickly access and visualize water-quality data, which is especially critical for hazardous incident response. This project will enhance our knowledge of complex water-quality issues and enhance management of natural resources.

Implementing FAIR practices: Storing and displaying eDNA data in the USGS Nonindigenous Aquatic Species database

We are working to incorporate environmental DNA (eDNA) data into the Nonindigenous Aquatic Species (NAS) database, which houses over 570,000 records of nonindigenous species nationally, and already is used by a broad user-base of managers and researchers regularly for invasive species monitoring. eDNA studies have allowed for the identification and biosurveillance of numerous invasive and threatened species in managed ecosystems. Managers need such information for their decision-making efforts, and therefore require that such data be produced and reported in a standardized fashion to improve confidence in the results. As we work to gain community consensus on such standards, we are finalizing the process for submitting such data to the NAS database. We are seeking support to expand the NAS database to store and present this newly Findable source of eDNA data so that it is Accessible to the public, Interoperable for use with new tools, and ultimately Resuable without limitation.

Real-time Coastal Salinity Index for monitoring coastal drought and ecological response to changing salinity values

Many coastal areas are experiencing departures from normal conditions due to changing land use and climate patterns, including increased frequency, severity, or duration of floods and droughts, in some cases combinations of the two. To address these issues, the U.S. Geological Survey developed the Coastal Salinity Index (CSI) to identify and communicate fluctuating salinity conditions due to such disturbance events through quantitative analyses of long-term salinity records. This project aims to make the CSI broadly useful as a monitoring, forecasting, and decision-making tool, extending the platform to enable real-time reporting of disturbance events as they unfold and covering a larger user base than what existing resources allow. The framework that supports this work addresses the Community for Data Integration theme of producing Findable, Accessible, Interoperable, and Reusable data by acquiring existing real-time salinity data, integrating into an accessible database, computing gage CSI statistics, and creating and displaying web-based visualization products.

Highlights

1. Reminder that the CDI workshop will be held virtually from May 25-28.
2. Look out for WaterRAT Online, coming soon. Metadata and data templates for AUV data are available from Andrea Medenblik or Bradley Huffman.
3. eDNA team has worked to integrate data into the Nonindigenous Aquatic Species (NAS) database through an Aquatic Invasive Species (AIS) viewer.
4. A real-time Coastal Salinity Index is currently being reviewed for release to the the public; in the meantime, there is a code repository available.

Notes

Welcome and Opening Announcements

1. For more frequent announcements, join our Microsoft Teams or wiki forum.
2. Kevin Gallagher remarks
    a. **Reminder that the CDI workshop will be held virtually from May 25-28.**
    b. The program team is currently going through session proposals from the CDI community.

Working Group Announcements

1. For more on all collaboration areas, see slides or wiki.
2. Geomorphology
    a. **Next event 4/27**: specific gage analysis
3. Metadata Reviewers
    a. **Next event: 4/5**
4. Semantic Web
    a. **Next event: 3/11**, continuing discussion on semantic web 101 session at CDI Workshop and potential ESIP lab project
5. Data Management Working Group
    a. **Next event: 4/12**, Update on QMS implementation and the role of data management
6. Usability
    a. **Next events: 4/7 and 4/21**, user research overview and demo of a user research technique
7. eDNA
    a. **New publication** from members: https://onlinelibrary.wiley.com/doi/10.1111/mec.15811
    b. **Logo contest** to represent the eDNA community of practice
8. Model Catalog
    a. **Next event: 3/24**, update on the Model Catalog - ongoing development and resources to support scientific model best practices
9. Imagery
    a. **New collaboration area;** planning session at CDI workshop
10. TechStack
    a. **Next event: 3/11**, earth data extraction, exploration and visualization using the AppEARS Platform
11. DevOps
    a. **Next event: 4/6**, USGS NGTOC - dynamic mapping

Waterbody Rapid Assessment Tool (WaterRAT): 3-dimensional Visualization of High-Resolution Spatial Data, Andrea Medenblik, USGS

1. Background
    a. **WaterRAT was designed for data from AUV's (autonomous underwater vehicles)**
    b. AUVs are underwater vehicles that swim around collecting data like conductivity, pH, dissolved oxygen, etc.
    c. They can also estimate unknown water quality
2. Problem
    a. **AUVs collect high-resolution 3D data which can be difficult to process, organize**

    b. Usually required proprietary software like Matlab
    c. AUV data are also not very accessible
        i. Spread across ScienceBase, difficult to compare/combine datasets
        ii. No standard method to organize these large datasets
3. Solution
    a. **Created an AUV database with all data releases from across USGS**
        i. Used python to import datasets into database
    b. Created a metadata and data template to standardize AUV data organization
4. WaterRAT Online
    a. Can select different parameters to view them within the AUV path/location
    b. Can use color bar sliders to edit how visualizations appear
    c. **Prototype was previously available as offline Python software; this new iteration is easily accessible online without Python knowledge, interactive, infinite options for data viz**
5. Connecting WaterRAT with database
    a. Challenges
        i. Data integration
            1. **Compiling 8 sites and 727,000 data points**
                a. Looking at previous versions of AUVs and newer ones
                b. Split up metadata to conform to best design practices of database management
        ii. Data processing
            1. **Defining AUV mission center lines**
                a. In the past, have manually created the center line
                b. This time, wanted to define center line programmatically
                c. Developed scripts which worked well for some center lines, but not all
                d. Created a custom interface where users can draw the center line
        iii. Getting WaterRAT online
            1. Needed a web server to securely expose WaterRAT to public internet traffic
                a. Used Apache and Mod_WSGI to work with WaterRAT on NatWeb
                b. Caching data retrieved from the MySQL database and output from computationally expensive tasks
6. How can AUV operators use WaterRAT?
    a. Code is under review, not publicly available yet, but look out for it soon
        i. Incorporate WaterRAT into your work plan
        ii. **Use the metadata template and data template (contact Andrea Medenblik or Bradley Huffman)**
7. How can scientists and the public use WaterRAT?
    a. Explore available datasets
    b. Interact with data - understand variations in water quality in water bodies
    c. Share with cooperators - online interface is intuitive and accessible
8. Follow-up questions and future
    a. **How to integrate with EarthMAP?**
    b. Integrate other 3D data (natural hazard response, tree canopy structure, fish abundances and habitat structure)
    c. Add built-in quality assurance prior to database entry to ward against errors and duplications
    d. Automate database - so AUV operators can publish directly to database (depends on quality assurance)
    e. Expand analytical capabilities

Implementing FAIR practices: Storing and displaying eDNA data in the USGS Nonindigenous Aquatic Species database, Jason Ferrante, USGS

1. eDNA is **environmental DNA that is shed by an organism into its environment**
    a. eDNA allows scientists to collect water and soil samples particularly on cryptic species
    b. Allows for **biosurveillance of numerous invasive and threatened species**
2. The Nonindigenous Aquatic Species (NAS) database
    a. Houses over 570,000 visual sightings nationally - generally pictures or publications with field sampling
    b. Regularly used by a broad user-base of managers and researchers working with invasive species
3. Problem
    a. Managers need data and metadata produced and reported in an accurate and standardized fashion to inform decision making efforts and ensure confidence
4. **Developed robust standards for reporting accurate data to the NAS database that are FAIR (findable, accessible, interoperable, reusable)**
    a. Common question: why not more open source, why use NAS?
        i. NAS is target species focused
        ii. NAS gave a structure to build into
5. A fair, open, centralized database
    a. eDNA data was scattered among manuscripts and reports, not easily retrievable, none dedicated to aquatic invasive species
    b. **AIS viewer allowed data accessibility, vetting and integration of data from multiple sources**, and improves coordination of research an management activities
6. Process for consensus derived standards to submit and simplify eDNA data in NAS
    a. Started internally with an advisory panel, put together proposed documentation, and did outreach to the community for peer review and feedback
7. **How it's FAIR**
    a. Findable: worked with in an existing database; putting data in front of people looking at similar data
    b. Accessible: existing mechanism for obtaining data form the database
    c. Interoperable: data submission template standardizes data submissions
    d. Reusable: Interest in secondary analysis (modelers, etc.)
8. NAS distribution map mock up
    a. See recording or slides
    b. Map will have educational component about eDNA caveats
9. Challenges
    a. **Community feedback**

        i. Planned for face to face meetings, but pivoted to webinars, virtual meetings, shared online documents, etc.
            1. More work to synthesize the input, but broader reach
    b. Takeaways
        i. eDNA data added to NAS creates a more complete distribution records of target species
        ii. Can help with response time to new invasions
        iii. Improves estimation of cryptic species occurrence rates, etc.
        iv. Putting together manuscript on community outreach for this project
    c. Follow-up
        i. Looking for help developing new tools which use eDNA data to inform management
        ii. Re-started eDNA community of practice

Real-time Coastal Salinity Index for monitoring coastal drought and ecological response to changing salinity values, Matt Petkewich, USGS

1. Motivation
    a. Drought early-warning system for North and South Carolina is missing components like coastal salinity
        i. Higher than normal salinity levels affect fisheries, municipal water intakes
        ii. Societal, economic, and ecological effects have not been conveyed clearly before
2. CSI is an approach similar to the Standardized Precipitation Index
    a. CD stands for coastal drought
    b. CW stands for coastal wet
3. Example graphs
    a. See slides for average salinity over time graphs
4. Existing Products and Resources
    a. **CSI R-package located in GitHub**
    b. Historic CSIs along the Gulf and SE Atlantic Coasts publication, plus three others
5. Used CDI Funding to:
    a. Develop CSI R-scripts for ecological analyses
        i. Allows comparison of anomalies in climate, hydrology, and salinity
        ii. See slides for graphs
    b. **Identify and integrate new salinity datasets**
        i. USGS (71 gage)
        ii. National Park Service (35 gage)
        iii. National Estuarine Research Reserve System (23 gages)
        iv. over 16 years of records, need to be electronically available
    c. **Enhance existing website and user interface to accommodate new CSIs**
        i. Conducted a survey and held a usability webinar to get feedback from potential users
        ii. Surveyed federal, state and academia users
        iii. Usability webinar walked through the website and obtained feedback.
    d. **CSI webpage**
        i. Draft, hopefully published in the next few weeks
        ii. See recording/slides for screenshots and tour of website
        iii. Includes interactive map, time interval options, CSI classification table, and station legend
        iv. Station pop-ups allow users to get a link to the originating agency data, can download input salinity data and access CSI drafts
        v. CSI About page gives background and team information
        vi. Additional Data tab allows users to select data pertaining to salinity, water temperature, etc., and time interval options
        vii. Graphs are also available for download from Additional Data tab
        viii. Resources page contains CSI tools and resources, publications, and more

Questions

1. AUVs
    a. Did you follow any specific standards while creating the AUV metadata? Are they ISO or CSDGM compliant?
        i. Consulted CDI folks on how best to make a metadata template. Followed practices recommended.
        ii. Jordan Wilson: The metadata are CSDGM compliant. The AUV metadata template was created using the Metadata Wizard as a launching point.
    b. Can you share a link to the metadata and data standard that you created?
    c. What viz technology was used in the online tool?
        i. Plotly, a Python package for graphics. The maps themselves are Matbox and another Python library.
        ii. Dash & Plotly information: https://plotly.com/dash/
    d. Does any one know if a similar effort is being conducted for UAV data?
    e. Is there a URL for WaterRAT?
        i. Not yet; it is under review.
    f. Does any one know if a similar effort is being conducted for UAV data?
        i. I wish we had more information. This is of interest, but we didn't have time to compile that data.
    g. How can we join the AUV working group
        i. Contact waterrat@usgs.gov
        ii. Please email Lee Bodkin at ljbodkin@usgs.gov for information on the AUV Working Group and interfacing AUV/UAV data.
    h. Have you explored ways to utilize and/or visualize data from the AUV's ADCP?
        i. Bradley Huffman: ...in the past we've explored visualizing the ADCP data but there were two hurdles. One, I've been told that the data isnt of a high quality from another AUV user which kills the motivation to attempt processing it myself. The second, is time I've got the documentation from the manufacturer on what the parameters mean in the log files produced by the ADCP but I'd need to have the time and funding to look further at correcting/processing it. We as a center dont publish the ADCP data to ScienceBase but that doesnt mean that cant change
    i. Maybe publish the AUV working group within the CDI if there is enough interest?
2. eDNA

a. Will this be ported up to GBIF?
    i. Yes, that is in the plan
b. If data are submitted to NAS, will they have a DOI for citation in a publication?
    i. I don't believe so. The data are submitted freely; some are from publications which already have a DOI. We ask in the submission form for linkages to already published works.
    ii. Matt Neilson: We do not mint a DOI, but we do provide a unique suggested citation.
c. Do you guys have a data template for eDNA data that others can use?
    i. We do have a specific template for submitting to the database. Not what you might like to use for other projects. We will provide in documentation the data fields that are required as part of the submission process, and the format data are required to be in to be interoperable. Additional part through the eDNA working group.

3. CSI
a. Were there challenges in integrating the three different sources of data?
    i. All have challenges. Gages we could use were limited because we only accepted ones with over 16 years of continuous data. Oher agency metadata can be difficult to sort through.
b. Any recommendations for expansion to other source gages? How could they be more easily aligned/integrated?
    i. The short time frame limited the depth; potentially would like to work with individuals maintaining the data at the different locations to see how it's easiest to filter metadata.
c. Can you share the link to the R pacakage?
    i. https://github.com/USGS-R/CSI
d. Could you talk about what kind of modeling users can do with this data?
    i. Hoping that modelers can pull in CSI data and correlate CSI to ecological results/information they're interested in.
e. Is salinity data collected along the west coast, AK, HI?
    i. Yes. In early times, we evaluated the data throughout all the U.S. (Washington, California, Oregon); timing and funding forced us to focus on East Coast and the Gulf; hope to expand out west.