

901915 - Improving MBARI Data Impact

Week 50

1. Diana LaScala-Gruenewald
 - a. Been bad about filling this out lately, apologies
 - b. Continuing to work primarily on Improving Impact
 - c. Still working to get STOQS data loaded into dockerized TimescaleDB. Everything is set except there's a bug around getting my load script to run automatically at container startup.

Week 46

1. Diana LaScala-Gruenewald
 - a. Have been getting familiar with STOQS, its architecture, loaders, running it locally, running through some of Mike's notebooks, extracting data
 - b. Met with Mike to review questions
 - c. At his suggestion, comparing the experience of pulling big chunks of rovs ctd data off of stoqs in parquet format with pulling the same data off of erddap in csv
 - d. As part of this, learning to use the holoviews Python package for interactive data visualization
 - e. Also spent some time looking into [CKAN](#) - a data management software system that includes cataloging and search
 - i. Useful [intro video](#), see 7:00-29:11

Week 43

1. Diana LaScala-Gruenewald
 - a. Tested serving all ROV CTD data from all vehicles/years through ERDDAP. After I figured out some details with the connection, it went pretty smoothly. Kept notes on what needed to be done manually.
 - b. Started getting familiar with TimescaleDB and STOQS
 - c. Meeting w/ Mike to get an intro to STOQS today

Week 42

1. Kevin Gomes
 - a. Looking into ERDDAP as mechanism to wrap access to timeseries DBs

Week 41

1. Diana LaScala-Gruenewald
 - a. Completed initial documentation
 - b. Reviewed ERDDAP docs/other materials to evaluate "ideal" data format, ways to serve large amounts of data, high request volumes
 - c. Began trying to serve ROV CTD data using ERDDAP - currently troubleshooting connection issues

Week 40

1. Diana LaScala-Gruenewald
 - a. Worked through adding database tables, views, and tables/views with lat, lon, time, depth to ERDDAP
 - b. Put together some documentation on BitBucket

Week 38

1. Diana LaScala-Gruenewald
 - a. Got reasonably comfortable adding data tables to test ERDDAP server
 - b. Got a test postgres db running in a docker container and linked to test ERDDAP via docker compose
 - c. Am currently troubleshooting the process of adding a db table to ERDDAP

Week 37

1. Diana LaScala-Gruenewald
 - a. Continuing to get familiar with adding data sets on ERDDAP, documenting what currently needs to be done manually during this process

Week 36

1. Diana LaScala-Gruenewald
 - a. Working on adding data tables to ERDDAP. Got a test data table up, but uncertainty about “correct” way to set up datasets.xml for different data types

Week 34

1. Diana LaScala-Gruenewald
 - a. Learning how ERDDAP is set up, how to add content
 - b. Now have an ERDDAP running in a docker container on my local machine
 - c. Working on adding tables, then connecting to database

Week 33

1. Diana LaScala-Gruenewald
 - a. Groundwork researching relevant technologies

Week 32

1. Diana LaScala-Gruenewald
 - a. Background work to learn about some of the technologies involved

Week 30

1. Kevin Gomes
 - a. Working through data lake stacks and working on matrix of cloud <-> open source <-> MBARI architectures related to data analytics
 - b. Preparing for project kickoff
2. Diana LaScala-Gruenewald
 - a. Intro to this project
 - b. Started thinking about integration with ERDDAP

Week 28

1. Duane Edgington
 - a. Project proposal
2. Kevin Gomes
 - a. Proposal writing

Week 26

1. Kevin Gomes
 - a. Starting proposal

Week 24

1. Kevin Gomes
 - a. Started drafting up docs to catalog data collection systems at MBARI

Week 23

1. Kevin Gomes
 - a. Just now getting back to this. Starting to research architectures that will meet requirements both on cloud and on-prem.
 - b. Researching AWS Data Analytics

Week 19

1. Mike McCann
 - a. Attended virtual OCG (Ocean Coordination Group)-12 workshop on Data and Metadata: Towards mutual benefit for data users

Week 10

1. Kevin Gomes
 - a. Data catalog GraphQL API work

Week 9

1. Kevin Gomes
 - a. Starting work again (finally!) on the data catalog
 - b. Moved first couple of components of catalog to Docker and new MkDocs format
 - c. Converting API to GraphQL before moving on to other components

Week 6

1. Kevin Gomes
 - a. Researching Apache NiFi for workflow management (recommended by LLNL)

Week 4

1. Kevin Gomes
 - a. Short discussion with group about goals for project

Week 3

1. Kevin Gomes
 - a. Follow on work to SamplesDB discussion (more meetings needed)
 - b. Defining first data sets for catalog work

Week 2

1. Kevin Gomes

- a. Discussions and meetings about SamplesDB and whether we can move it to a service while incorporating changes to support EDNA analysis (sub-samples)