**Things to change about the co2 data model**

**Phase 2, project development**

**Project Tile: Web Portal for Concentration and Flux of CO2 in Global Inland Waters**

**The Problem:**

Observational data, especially *in situ* CO2 concentration and flux measurements, are essential for correctly modeling CO2 evasions from global inland waters. However, these measurements were collected and published separately by different research groups and there is a lack of a cohesive synthesis of direct measurements, hampering our ability to accurately model CO2 emissions from inland waters. We believe that a global synthesis of direct CO2 measurements would greatly enhance our understanding of the role that inland water plays in contributing CO2 to the atmosphere. A publically accessible, easy-to-use web portal for researchers to easily input, visualize and download data would be a favorable approach for the synthesis ahead.

**Application Example:**

An impressive web portal from Global Ocean Acidification Network for visualizing ocean field campaigns can be found via <http://portal.goa-on.org/Explorer>.

**Sample Data:**

We’ve compiled some 6000 individual CO2 concentration/flux measurements from the literature in MS Excel form, which can be used to jump start construction of the online web-portal. We expect, however, that most of the project time is spent on construction of the front- and backend of the web portal rather than using the data for actual analysis. We hope that size of dataset can be doubled or tripled over the next 1–2 years through involving the bigger community with a well-designed web portal.

**Specific Questions (breaking up of the project task):**

The project contains two main parts: designing the frontend web application and backend database of the web-portal.

*1. Designing frontend of the web-portal*

* ***The main page*** — the main page of the web portal includes three major functionalities for uploading, visualizing and downloading CO2 observational data in global inland waters. Included also is a map interface which displays background world map, site locations and CO2 data from a connected backend database.
* ***The map interface*** — choose an online map service which can be integrated into the main page as a background map. The map must also have capabilities to be connected to the backend database and to visualize georeferenced point data. Adjustable tools for simple visualizations of the data points (in symbol size, color ramp, etc.) will also be developed.
* ***Uploading data*** — an UPLOAD DATA button unfolds two possible data uploading options: Batch Upload and Single Upload, which lead users to web interfaces for batch and single uploading options, respectively.
* ***Batch Upload***
  + In the Batch Upload interface, users are asked to select a series (e.g., 10) of columns and units and specify the total number of observations (e.g., 15) to generate a spread table for data input. Selected columns, units and number of rows correspond to the structure and formats of user’s to-be-upload data.
  + A spread table is generated with user-specified columns, units and number of rows (15 rows × 10 columns with specified units in this case). The user can paste his/her data to the spread table. Column sequence is adjustable by giving sequence number to each column to be adjusted. A SUBMIT button sends data to the backend database.
* ***Single Upload***
  + In the Single Upload interface, users will type data directly into the data-submitting form and select units from dropdown menus. User has the option to save typed-in data at each intermediate step and submit data to the database by clicking on SUBMIT.
  + Note — considering most of users’ data are stored in spreadsheet formats (MS Excel sheet, csv file, etc.), the batch upload option which generates a spread table for users to paste and send data is probably recommended.
  + Note — allowing user to upload csv files to the database is probably not a good option for taking in user data considering very limited control on data format even with associating metadata.
* ***Viewing Data*** — on the main page, user can choose different data types (CO2 concentration, flux and other ancillary data) for visualization and viewing by checking on/off different boxes.
* ***Downloading Data*** — on the main page, for data downloading, user can filter through a series of data types and click on the DOWNLOAD DATA button (this option can be probably developed at a later stage of the web-portal development).

*2. Designing backend database of the web-portal*

* The backend database hosts data input through batch and/or single user upload and responds to user queries. The database should contain columns of the same data types and unit options as in the web interface. Columns in the frontend web interface and backend database are connected and work together to add, update and store user data and respond to user queries. The database can also perform computing (e.g., unit conversion to default) in the background.

*3. List of database/web-interface columns for user data taking-in*

* ***Part 1:*** ***Site Information*** —
  + ***Site Type***: stream, river, lake, reservoir, pond, estuary, wetland, estuary, floodplain, and others
  + ***Site Location***: longitude and latitude in decimal degree or “deg, min, sec“, and altitude in meters above sea level (masl.)
* ***Part 2:*** ***Physical Properties*** —
  + ***Sampling date:*** yyyy-mm-dd
  + ***Discharge:*** m s-1, L s-1, ft s-1
  + ***Water Temp:*** °C
  + ***Air Temp:*** °C
  + ***Width:*** m, km
  + ***Depth:*** cm, m, ft
  + ***Surface Area (for lakes, ponds, etc.):*** m2, ft2, km2
  + ***Flow Velocity:*** m s-1, ft s-1
  + ***Wind Speed:*** m s-1, ft s-1
* ***Part 3:*** ***Greenhouse Gas Concentrations*** —
  + ***CO2:*** ppm, µatm, µmol L-1, mg C L-1 mg L-1
  + ***CO2 method: …***
* ***Part 4:*** ***Greenhouse Gas Fluxes*** —
  + ***CO2 Flux:*** g C m-2 yr-1, mg C m-2 d-1, mg C m-2 hr-1, µg C m-2 d-1 g C m-2 d-1
  + ***CO2 Flux method:***
* ***Part 5:*** ***Gas Transfer Velocity*** —
  + ***k600-CO2:*** m d-1, cm s-1
  + ***k-CO2:*** m d-1, cm s-1
* ***Part 6: Other field measurements —***
  + ***DOC:*** mg L-1, µmol L-1
  + ***POC:*** mg L-1, µmol L-1
  + ***TOC:*** mg L-1, µmol L-1
  + ***pH***
  + ***Alkalinity:*** 100 mg CaCO3, µmol L-1
  + ***DO:*** mg L-1
  + ***Chl a:*** µg L-1

Note: listed selection options and units are to be added or changed.

**Existing methods**

Platform: AWS, MS AZURE, Google Cloud, Yale Server

Frontend web application: HTML, CSS, Python, etc.

Backend database: MySQL, Oracle, etc.

**Proposed methods/tools**

To be determined.

**Background reading**

Abril, G., S. Bouillon, F. Darchambeau, C. R. Teodoru, T. R. Marwick, F. Tamooh, F. Ochieng Omengo, N. Geeraert, L. Deirmendjian, and P. Polsenaere (2015), Technical Note: Large overestimation of pCO2 calculated from pH and alkalinity in acidic, organic-rich freshwaters, Biogeosciences, 12(1), 67-78.

Allen, G. H., and T. M. Pavelsky (2018), Global extent of rivers and streams, Science.

Raymond, P. A., J. Hartmann, R. Lauerwald, S. Sobek, C. McDonald, M. Hoover, D. Butman, R. Striegl, E. Mayorga, and C. Humborg (2013), Global carbon dioxide emissions from inland waters, Nature, 503(7476), 355-359.