A New Generation of Data Sharing Platform for Eddy Covariance Flux Data

Baolei Lv1, An Yan1, Yuqi Bai1,*, Bing Xu2

1Ministry of Education Key Laboratory for Earth System Modeling, Center for Earth System Science, Tsinghua University
2School of Environment, Tsinghua University
Beijing, China

lvbl12@mails.tsinghua.edu.cn, yan-a12@mails.tsinghua.edu.cn, yuqibai@tsinghua.edu.cn, bingxu@tsinghua.edu.cn
*Corresponding author (Tel/Fax: +8610 62795269)

Abstract—This paper introduces a case study of establishing a new generation of data sharing platform for eddy covariance flux data for the US-China Carbon Consortium. This platform manages the information about the observation sites, principle investigators, research projects, related publications, and the processed flux data. It supports data query and on-line data presentation based on these centralized managed metadata information. It introduced a relational database scheme for flux data storage. The system architecture, information model, metadata standards, data sharing and citation policies and dedicated flux data interchange format are introduced. Two advantages of this case study are discussed at the end: a dedicated flux data interchange format, and an ISO 19115-based metadata standard profile dedicated for the eddy covariance flux data.

Keywords—Flux data; data sharing platform; ISO 19115; relational database

I. INTRODUCTION

CO2 and water vapor flux between terrestrial ecosystems and atmosphere measured by eddy covariance method has enabled the understandings of regional source-sink pattern of carbon, the estimations of net ecological exchange, and the predictions of carbon balance in the future. Eddy covariance flux data is the key to quantify and interpret dynamics of water vapor and carbon dioxide balances and to understand their responses to biotic and abiotic variables [1][2][3].

During the last several decades, a large number of observation sites have been established worldwide, and therefore, a large volume of eddy covariance flux data has been produced independently and separately in different sites [4]. Under many circumstances, researchers need to collect and analyze flux data from multiple sites when performing ecology studies [2][5]. Therefore efficient and reliable data sharing mechanisms are needed to enable these scientific activities. Establishing a dedicated flux data network among data owners have been proved a best practice in promoting data sharing and exchange. FluxNet, AmeriFlux [6] and ChinaFlux have been successfully established to promote the data sharing activities among ecology research communities.

Although eddy covariance flux data could be shared through these flux data networks, there are still many pre-processing issues that researchers need to manually resolve when performing ecology studies. In particular, formats and contents of flux data files produced in different sites may not be the same due to the different environment settings and processing methods in these individual sites. Also, there is still no centralized and standardized flux data management solution.

In this paper, a new generation of flux data sharing platform will be presented. It introduces a relational database for eddy flux data, a standardized flux data interchange format, and a Web-based data query and presentation interface. The structure of the paper goes as follows: introduction of the scenario, data attributes analysis, system architecture, work progress and conclusion.

II. SCENARIO

US-China Carbon Consortium (USCCC) was founded in 2004 by several ecologists who maintained flux towers in the U.S. and China respectively. It aims to better understanding environmental factors that influence water cycling, magnitude and rate of carbon sequestration across diverse climates and ecosystems. So far, USCCC maintains more than 50 flux sites [7]. The current data sharing practices in USCCC involves many activities that need to be performed manually, such as data delivery and integrated data analysis. Therefore, a more open, standardized, and integrated solution is highly expected to facilitate the exchange of and the integrated analysis of the flux data in the USCCC community.

A. USCCC flux and meteorology data

In an observation site, carbon and water flux is continuously measured using eddy covariance method by sensors installed on the same observation tower, with micro-meteorological conditions being measured simultaneously. Once captured, the flux data need to be pre-processed, which usually includes several steps: flux calculation, quality control (QC), interpolation and decomposition [8].

1) Flux Data

10 HZ measurement data are finally processed into half-hourly flux data using EddyPro software [9][10][11]. Output files contain the following items: original filename, time stamp,
concentration and statistical parameters. Filename item indicates the name of pre-processed data file which is directly produced by the instruments. Time stamp consists of date, time and DOY (Day of a Year) information. Statistical parameters include mean value, standard deviation, skewness and kurtosis of carbon dioxide, water vapor and dew point. The files are in the CSV format.

2) Meteorological Data
Meteorological data is also processed into half-hourly data. It mainly contains data items of time stamp and meteorological variables. Noting that one meteorological variable might be measured at different heights simultaneously, it needs special attention when designing a relational database for flux and meteorological data. The files are in the CSV format as well.

B. Problems of Interchanging and Integrating Flux and Meteorology data in USCCC

There is no standardization way to organize, interchange and integrate the flux data and meteorological data in USCCC. Many flux networks, such as AmeriFlux and AsiaFlux, have strict standards on data management [4][6]. These standards are essential to guarantee the quality of the data and to protect investigators’ intellectual properties.

Furthermore, a centralized database for flux data and meteorological data, plus a Web-based data portal is still missing in USCCC to allow integrated data query and analysis.

III. PROPOSED SOLUTION

A. System Architecture

![System Architecture](image)

We adopted a three-tier architecture in system developing: presentation tier, business-logic tier and database tier.

1) Presentation tier
Presentation tier provides a Web-based data query and presentation interface for researchers. On the back-end, it talks to the business-logic tier to fulfill their queries.

2) Business-logic tier
Business tier handles user queries forwarded from the presentation tier. It interprets the queries, searches the database, generates the query results, and sends them back to the presentation tier.

3) Database tier
Database tier manages a centralized relational database for flux data and the meteorological data. It enables data manipulation operations, such as insert, update, and delete. It also supports exporting the data from the database to be in the CSV format.

B. Database Design

In this study, relational database was selected to maintain the data storage for flux data and meteorological data respectively. As shown in Figure 2, the designed database scheme allows a fine-grained storage solution at variable level, thus enabling users to fetch partial of the flux data and meteorological data for selected variables, rather than having to download the entire data. More importantly, maintaining flux data and meteorology data at variable level makes flux data online analysis possible.

One significant issue was dealing with meteorology variables measured at several different heights simultaneously. For example, there are two RH (relative humidity) values, RH_3m_avg and RH_7m_avg. The first one was measured at the height of three meters, while the second at seven meters. To explicitly main two columns in the database for these two variables respectively is not a general-purpose solution: it brings a big problem when dealing with meteorology data that could be captured at different height levels.

To address this problem, a novel database design was introduced: instead of maintaining meteorology variables in one table, height-dependent variables are stored separately in a table named “height_climate”, while all other values are in “climate” table, as shown in Figure 2. In the “height_climate” table, values of the variables and the corresponding height values are explicitly maintained. This table, and the “climate” table jointly provide a complete solution for storing meteorology data. When meteorological data are imported into database, those data items of meteorological variables with different heights will be stored in a separated table named climate_height while other meteorological data are stored in table named climate. Same variables of different height have same ID but stored in different data columns which can be seen from entity relationship diagram above.
C. Data interchange format

Data interchange format is defined to promote data exchange, sharing and integration within the USCCC community. This format mainly specifies the data content, data encoding format, organizing method and data file format.

- Data content: the data items, orders of these items, and exact name of each items;
- Data format: format of attribute values, including decimal digits;
- Organizing method: length of each variable, naming methods of files, procedures to submit data
- Data file format: the CSV format was adopted as the interchange format.

D. Data sharing policy

To boost data sharing activities, policies on data storing, requesting, sharing, publishing and submitting activities have been formalized within the USCCC community. To protect Principle Investigator (PI)'s intellectual property, researchers need to fill in an application form to request for the data of interest. Data sharing can now be conducted online and authorship should be in consensus among related PIs before results published.

IV. Resulting system

A. Data

Flux and meteorological data measured in 2011 in the site of Zhangjiangkou, which is located in Fujian province coast area, have been ingested into the database for demonstration purpose.

B. Database

MySQL database was selected to build the centralized flux data and meteorological database in this new data sharing platform. This database holds four tables, namely climate, height_climate, siteinfo and eddyflux. These four tables store data of meteorological variables, meteorological variables at different height levels, site-related information and eddy covariance flux data respectively.

C. User interface

User interface mainly serves to collect user queries and display corresponding result data. The user interface is shown below in Figure 3. From top to bottom, this interface consists of three panels, with the last one, online analysis, under development.

Figure 3. Interface of flux data sharing platform

1) Query condition panel

As shown in Figure 3, query condition panel allows users to search data by specifying conditions of flux site and flux and meteorological variables. Users can specify site conditions, including PIs, site name, temporal and spatial scope and land cover type, to find the sites of interest. Spatial coverage conditions can be specified with a Google map based image viewer by defining the interested region on the map or identifying longitude and latitude values of the upper-left and lower-right corners. Also, users can determine data items of interest by checking specific variables of flux and meteorology.

2) Matched query results panel

Matched query data will be shown in the result panel. Note that, only the selected attributes are presented. Users can further download these results as CSV files for further analysis.

V. Conclusions

The paper presents a new generation of data sharing platform for eddy flux data. The research scenario, system architecture, technological solutions and resulting system are presented in detail. Successful utilization of this platform in supporting the USCCC community shows that this platform is promising in enabling a more open, standard and integrated ecology data sharing and analysis environment.
REFERENCES


