Geosci. Model Dev. Discuss., 8, C2810–C2816, 2015
www.geosci-model-dev-discuss.net/8/C2810/2015/
© Author(s) 2015. This work is distributed under
the Creative Commons Attribute 3.0 License.

Geoscientific Model Development Discussions
Open Access

GMDD
8, C2810–C2816, 2015

Interactive Comment

Interactive comment on "Distributed visualization of gridded geophysical data: a web API for carbon flux" *by* K. A. Endsley and M. G. Billmire

K. A. Endsley and M. G. Billmire

endsley@umich.edu

Received and published: 7 November 2015

Thanks for your feedback! We'll address your comments one at a time below.

> "The paper struggles to make the case for why having the data directly in the user's browser is desirable."

Having the data available directly in the user's browser allows for seamless rescaling of the visualization, e.g., changing the stretch "on-the-fly." We have seen performance issues in comparable approaches to this problem, e.g., with Web Mapping Services, which must request the data again from the server whenever scaling changes are desired. We believe scale changes, changes in the palette, and similar changes are in the

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

purview of the client application; as they are merely changes in the application's state, they should be performed asynchronously in the client application without requiring interaction with the remote server.

> "The paper also does not put the CDE's ability to "manage, aggregate, visualize, and share datasets" in the context of the THREDDS Data Server."

It is possible that we haven't provided enough context for the THREDDS Data Server, as we're less familiar with its operation. We will address that in revising the text. Specifically, we will compare the CDE's features with those of the THREDDS Data Server.

> "The text claims the CDE and this approach results in high performance (page 4, line 22), but no performance metrics are given."

We will provide performance metrics for the API in the revised manuscript.

> "I also cannot understand how a text representation of a number used in the CDE does not massively increase the size of the original data."

I believe this comment is related to a remark in the manuscript on page 4, line 24: "This text-based representation has the added benefits of compressing the data and enabling rapid filtering and aggregation." This remark is unclear and we will revise it in the manuscript. The text-based representation is not compressed relative to the original data, rather, we have eliminated redundancies that would come from a straightforward rendition of the data as text. Specifically, the spatial structure of the data have been separated from the measurement values so they can be transmitted to the client separately and without redundancy.

GMDD

8, C2810–C2816, 2015

Interactive Comment

cc BY

“The authors need to be clear what core goals the CDE addresses that these existing tools [THREDDS and Ferret-THREDDS] do not capture. In the context of data sharing, what does CDE do that OpenDAP does not? Is the distinction that OpenDAP generally requires desktop applications, whereas the goal here is to provide web browser access? Again, the case needs to be made for why client-side data provides superior capabilities to the current server-side approach.”

We will revise the manuscript to elaborate on the differences between the CDE and Ferret-THREDDS and between the CDE and OpenDAP. As you intuited, a core difference is that the CDE can be run in a web browser on any computer without requiring the installation of dependency libraries (e.g., Java). In addition, the provision of the data directly in the web browser allows for a level of interaction and immediacy with the data we feel are not available with alternatives.

“Could the concept of JSON data slice access be integrated into the THREDDS software stack as opposed to using MongoDB? ...When datasets can be rather large, I don’t practically see modeling centers hosting two copies of their datasets, one for THREDDS (which they will do for CMIPs) and a text version for CDE.”

That is a compelling idea and one worth exploring in future work. I don’t see any reason why a module for THREDDS could not be developed to support this. Fast JSON serialization of global-scale climate data is one of the things we’ve demonstrated with CDE, as many data representations are transformed inside the database on-demand, so we’re confident that THREDDS could accomplish this “on the fly” in a similar manner without storing a redundant JSON format.

“It is not clear from the introduction that the CDE is software that is installed at a modelling data center as opposed to an online service or API.”

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

CDE could be installed anywhere but we anticipate CDE would be installed by the people who “own” the data. We will strive to make this more clear in our revisions.

> “Likewise, since the software is installed locally, should the name Carbon Data Explorer be generalized?”

Other reviewers have suggested the name could be changed as the software is not restricted to visualizing carbon science datasets. However, as the software we developed with the aim of visualizing carbon science datasets, and as its features are likely best suited to that purpose, we feel the name is appropriate. As for installation, the CDE could be used only locally (that is, “in-house” by the data providers) but they are still “exploring” their data, as far as that metaphor can carry them. It is also our hope that the serialization to ASCII Grids and GeoTIFFs would allow non-experts to access climate data through the CDE and, thus, it would truly allow scientists in other domains, not just climate experts, to “explore” such data.

> “Can large-scale meaningful analysis (beyond mapping) practically be done in real-time on the client browser with current technologies?”

We are not domain experts ourselves, so we collaborated with climate scientists in the development of this software. With the exception of visualizing covariance structure, the analytical capabilities they requested are fully implemented in the CDE, including side-by-side measurement and uncertainty visualization, for global-scale gridded data on the order of 1-degree. We will provide performance metrics for this use case. For datasets with significantly higher spatial resolution, CDE might not be useful.

> “What are the ‘rich analytical capabilities’ mentioned? Are they JavaScript math libraries? Is the text referring to temporal and spatial averages or more complex forms of analysis?”

GMDD
8, C2810–C2816, 2015

CC BY

Our use of the word "rich" may be subjective here. JavaScript math libraries enable the computations but they are, specifically, the aggregation of multiple data slices, the differencing of two data slices, the calculation and visualization of anomalies, and the computation of summary statistics over arbitrary regions. In working with our domain experts, we identified these at the key analytical capabilities they needed. In presenting early versions of the software to groups like DataONE and the OCO-2 Science Team, we did not identify any significantly different analytical workflows for these types of data (Level III gridded products).

> "Is CMIP6 the first time 'distributed analyses' will be used in a CMIP? I don't think the Meehl et al. 2014 reference support[s] this, but rather the organization of CMIP6 will be distributed (in a non-computing sense."

Thanks for this correction. We will consult this paper again and revise this section for clarity.

> "Secondly, sharing 'web-compatible scientific datasets' was accomplished with the ESGF used in CMIP5."

We did not give due credit to the prior work done by ESGF here and will revise this section accordingly.

> "How does the CDE 'lower or eliminate barriers to bringing scientific results online'? Files and metadata need to be organized prior to loading into CDE. How is the process of making data available via CDE easier than other software such as THREDDS?"

CDE provides a fairly high-level command-line interface for managing the database, including for inserting data. Files and metadata only need to be organized insofar as they either need to be formatted to match a predefined data model or a data model needs to be written to accommodate them. The latter option is not a low barrier to entry but it does require only a few lines of Python code. In addition, if a data provider consistently generates data in a given format, the data model for that format can be reused for all future datasets, so subsequent database uploads are literally "one-line" commands. Once data are in the database, they are automatically available through the API and a connected CDE front-end web application. We will revise the manuscript to include a comparison to data entry with THREDDS.

> "The application seems to handle one polygon/ROI, can data 'quickly' be aggregated over many polygons such as counties, states, or countries?"

The CDE only supports aggregation to one polygon/ROI at a time. We have experimented with offline aggregation over multiple geometries at once, for instance, in preparing a time-series animation of state-level carbon flux (http://spatial.mtri.org/flux/us-states-breathing.html), but this functionality was not identified as an important feature for the CDE.

> "Many atmospheric models and ocean models have irregularly spaced grids. Can CDE read / map those grids appropriately?"

CDE supports irregular spacing (i.e., holes) but not irregular cell sizes or non-rectangular cell shapes. To be precise, CDE supports only "structured" grids and not "unstructured" grids. Support for unstructured grids is a compelling use case as we're aware that there are geostatistical advantages to non-rectangular grid cells. It would be possible to modify CDE to support unstructured grids (e.g., hex-bins) in the future.

Regarding the web app feedback, temporal aggregation (and difference) is only enabled when summary statistics from the "Current Data Frame" is selected. These features do not work with "Population" summary statistics because it would not be useful

CC BY

to look at aggregated or differenced data scaled based on, e.g., the mean and standard deviation of the entire dataset. There may be some differences in performance and function on different browsers. Not all web browsers are equally performant with respect to JavaScript Array calculations and handling large datasets. We have found that Google Chrome was the most performant for our application during development, thus, the CDE has been tested comprehensively with Google Chrome.

Interactive comment on Geosci. Model Dev. Discuss., 8, 5741, 2015.

CC BY