Geosci. Model Dev. Discuss., 8, C2520–C2522, 2015
www.geosci-model-dev-discuss.net/8/C2520/2015/

Geoscientific Model Development Discussions
Open Access

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

Anonymous Referee #3

Received and published: 17 October 2015

General comments:

In this well written, good to read paper the authors present Carbon Data Explorer (CDE), a tool for inspection of space/time datacubes. On datacubes previously ingested into the system various processing functionality is available, including data extraction, processing, fusion, and visualization in a Web browser. With this, this paper lies in the current trend towards multi-dimensional data management and analysis.

The architecture is client/server, with the client being a thin Web browser. On server side, queries are evaluated against spatio-temporal x/y/t data previously ingested into a particular data management component. The tool offers functionality generally considered as useful and important for a scientist's workbench.

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

Specific comments:

Authors claim CDE is offering distributed visualization, however, this is not substantiated in detail; from what I can infer data are always loaded from (server) local files. Visualization is addressing 3D x/y/t cubes plus multi-variables, but it remains essentially 2D plus "movie", no 3D portrayal is mentioned.

Also wrt data management and processing important questions remain open: what is the exact storage scheme for the datacube in MongoDB? what part of the analysis is pushed into MongoDB, and what is solved in the middleware? In how far do the authors consider this architecture scalable? To this end, at least a few performance figures would have been helpful, even better so a comprehensive evaluation: what are response times in general? where in the architecture is time spent, eg: how much of the query response time goes into MongoDB, and how much into the JavaScript middleware? How does this compare to, eg, C/C++ implementations?

Which gets me to the final point: while the paper mentions some tools and one standard (WMS) in the field it lacks a solid comparison against immediately "competitors". Hadoop, Array Databases, as well as virtual globes like NASA WorldWind come to my mind.

As it stands, we more or less only learn that CDE exists, but cannot carry home insights which advance the state of the art. Which is a pity as there may be interesting concepts behind which we would like to learn about.

Here some detailed thoughts, on which the above evaluation is based, among others: - abstract: "any time-varying, spatially explicit dataset": I do not see this promise fulfilled. From my understanding, the tool supports spatially and temporally regularly gridded data. Hence, neither the various irregular grid types nor meshes appear supported. - OPeNDAP strictly speaking is not a standard (such as ISO standards), but a widely used tool; another aspect is that a standard could not reference an implementation (as OPeNDAP is), but would typically describe interfaces to which various

GMDD

8, C2520–C2522, 2015

implementations can, and in case of a successful standard will, exist. - may consider OGC WCS as an interface standard specifically designed for multi-dimensional geo datacubes - 2.2: data ingestion, ie: massaging heterogeneous incoming data to a suitable service structure, typically is an involved task. Section 2.2 does not detail on this, which would be interesting to know: what challenges had to be met? Any innovative approach taken? - 2.3: massive binary data encoded in text form seems like a big impediment for transfer and processing. MongoDB querying certainly does not offer competitive performance on datacubes, and only limited functionality. Unfortunately, the paper remains superficial here and does not explain the detailed storage schema. - Example in 2.4: the result to me, following the query logic, should be a 3D cube extending allong the full spatial footprint and temporally reduced to the start and end point indicated in the query. However, authors call the result a "timeseries" which earlier has been introduced as being 1-D. This might be clarified. - authors characterize retrieval from MongoDB as "very fast" but without indicating measurements, and no comparison to tools offering the same functionality.

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