

Interactive comment on “Technical Note: The Linked Paleo Data framework – a common tongue for paleoclimatology” by N. P. McKay and J. Emile-Geay

D. Fils

dfils@oceanleadership.org

Received and published: 19 November 2015

I have been following the evolution of LiPD for a while and I have given my input previously to the authors of this paper.

I feel compelled to chime in as a data manager working on scientific drilling data from both the Integrated Ocean Discovery Program (<http://iodp.org/>) and Continental Scientific Drilling Coordinating Office (<http://www.csdco.umn.edu/>) facilities. The following are my personal observations and opinions.

CSV as an underlying data file format is pragmatic on many fronts. The paleo com-

C2397

munity is strongly invested in CSV based work flows and is unlikely to easily switch to another. netCDF would require installing/using new tooling, training and changes to existing workflows. All this would make adoption problematic. Additionally, LiPD's use of CSV for the web and associated metadata approach allow for translation to structured formats like JSON or RDF.

In terms of structure for CSV it would be good for this community to push the recognition of IETF effort on CSV (<https://tools.ietf.org/html/rfc4180>) that address issues of delineators and structure. W3C efforts around CSV (used by these authors) build on 4180 and are moving toward efforts around a W3C standards track for this work (<http://w3c.github.io/csvw/syntax/>). 4180 aware CSV libraries exist for (or are already native to) many languages and largely removes issues of delimiter selection of placeholder characters. Additionally support for UTF-8 unicode is also well in place for many CSV libraries. This is an area of active work by the W3C which we as a community would be wise to engage.

I was thrilled at the author's selection of JSON-LD. JSPON-LD is a light weight linked data format that allows me to represent my RDF models in JSON. This means I can connect efforts in semantics from other vocabularies into JSON-LD via the context element along with spatial information via GeoJSON. Again, the authors have shown this in the paper as well. The potential for this is large in my opinion. It means I can begin to make connections to other semantic efforts in this and other communities. LiPD's connection to metadata, spatial and bibliographic elements in the package demonstrates the ease with which other extensions could be added for special cases while not adversely impacting the simple core of LiPD.

More explicit, in terms of Linked Open Data (LOD) efforts, the selection of JSON-LD and the semantic connection this brings allows me to explore connecting LiPD described data to GeoLink (<http://geolink.org/>), part of the NSF EarthCube effort. Also, connections to efforts in the iSamples Research Coordination Network (RCN), another EarthCube project, allows me to make connections between samples and data and

C2398

express this in the LiPD metadata. For the first time I can begin to see the blocks coming into place to allow myself and other data providers a set of tools to build a web of data around scientific communities and their data. LiPD is one part of this larger mosaic.

It's good to keep in mind that JSON-LD is representing an underlying data model (RDF or quads). It is simply one representation of that model with many benefits. However, there is the XML version (RDF/XML) and more human friendly versions such as turtle (<http://www.w3.org/TR/turtle/>). Converting to display formats like HTML or even visual SVG is well established in the community as well. This means many existing personal and machine tool chains are supported. I've been working LiPD into my existing metadata code base and the process leverages well off the existing approaches. It is worth noting here that we must be alert not to conflate the representation with the model.

There has also been discussion among other community data providers on incorporating LiPD. A rich base of paleo, linked data aware, metadata in a format that fits well with existing semantic development efforts is a win for the entire community and I look forward to evolving this effort with the authors and other users.

Interactive comment on Clim. Past Discuss., 11, 4309, 2015.