Reply to reviewer 1

We thank the reviewer for bringing up several concerns with the data format and community use and support for Linked PaleoData. All of these concerns are easily addressed - indeed most have come up before in previous iterations of the discussion and feedback with the community. In this technical note we focused on the technical details of the format and the standard, without addressing the larger context of how LiPD would fit into the workflow of individual users, or the extent to which the community has been involved in its development. The review makes the valid point that we did not adequately make these points clear in the manuscript, and will attempt to do so here and as well as in a revised version:

1 The manuscript presented by McKay and Emile-Geay is putting forward a proposal for definition of a palaeoclimate file format for data exchange. This is done on the basis that a) it is arguably difficult to exchange large amounts of palaeoclimate data in a consistent way b) this prevent the automatic construction and parsing of large (proxy) databases c) the underlying format for the data and meta-data is researcher dependent, creating a large number of possibilities. The proposed format to be used is the JSON-LD format (a JavaScript originated avatar) for the metadata overlying a CSV (Comma Separated Value format) data file for the actual proxy data.

2 In the following I will refer to "data" without further mention as a shortcut for "Paleoclimatological Proxy Data", covering potentially the proxy data itself, the age models etc.

3 Several questions need to be stated when thinking about data exchange. The first one is obviously to define the communities (data gatherers, data users, data compilers, data providers) and their work methods to check the adequation of the proposed format with the present routine work with minimal adaptation. Second is the adequation of the format proposed to the work it should be used for. Third is the community support that is gathered around that proposal.

4 Taking up the first community type: data gatherers. The generation of proxy data is generally ending up in the production of non standard excel sheets (or other equivalent non-proprietary format, but the latter is very rare), that are produced on different operating systems. The proxy data gatherers (or producers) are themselves not using any other format to my knowledge for their daily work. Hence, creating all their work in a format that is not natively supported by an office production suite (without reference to a proprietary system) is not likely to be largely adopted. The underlying format is hence necessary to be compatible with this.

We recognize that data producers predominantly rely on Excel or spreadsheet software to report and manage data. The expectation is not that data producers will be entering the data *directly* into LiPD; indeed, we have developed utilities to import Excel records into LiPD, and this was the primary way that the PAGES2k were translated to LiPD. However, we also emphasize that innovation may require the use of new technology. This

should clearly be a gradual process, and we suggest that the local formats that data producers use to create datasets are excellent for record development, but should not be used as the format to archive data or to pass published data between scientists.

We have developed web-based ways to translate spreadsheets into LiPD. Ultimately, we envision open-source desktop tools that will enable users to do so. But first we need to publish and iterate on this format.

5 Furthermore, the format for actual data (measurement) exchange requires an excellent portability of the underlying format to any system or language (by the latter I infer the common computer "localization" problem). It thus requires a format that do not include any structure that could be misinterpreted in that regard.

6 Data users can be other data gatherers, data compilers or other communities (e.g. modellers). For other data gatherers, the exchange format is non-critical since the common use of the excel-type format is very comparable in all communities and therefore do not hamper the data exchange. For data compilers, the issue is more complex: it requires to use a common format indeed (as pointed by McKay and Emile-Geay) for the actual measurements and for the metadata. As an example, the largely reknown PANGAEA database is using HTML metadata and TSV (Tabulated Separated Values) text files for the actual data. For the other communities (in particular the modellers or the model and data intercomparison group), the expectation is to have a format that is robust (as opposed to error-prone) and easily transferrable.

7 The JSON-LD format proposed for the metadata is a very peculiar choice. Though is known for being relatively lightweight, it is very seldom used to gather databases. In the instance of proposing a common language in the view of creating large databases, it is a very contradicting approach. The only long-lasting advantage I can see in this choice is the capability of being easily read when opened in a simple text editor (e.g. Notepad or emacs).

This is an odd objection. JSON is a leading format for data exchange on the web, and extensible databases are becoming increasingly common for datasets that are not well suited for relational databases. Paleoclimate data clearly fall into this category. For extendable databases JSON and XML are essentially the two choices, and the simplicity and human-readability of JSON, as well as the rapidly evolving community support for the format makes it preferable to the far more verbose XML choice. Besides, the volume of paleoclimate data is not ever expected to be something so large that JSON would not be able to handle it. Certainly, it is equally as capable as XML on this front.

8 On the over hand, thinking about the daily work of a data gatherer, this format is largely at odd with the tool used. There is no simple way to create a template or better input-ready form that can be used in an office suite. I have made the simple test to open the provided jsond file in an

office suite software and the result is not something you will want to work with on a daily basis (I encourage the editor and other reviewers to perform that simple and convincing test).

Again, our intention, which we didn't make sufficiently clear in the manuscript, is that data producers will not be editing the json-ld file directly. The standard is meant to be explicit and self-describing, so the meaning is built into the file in a uniform way. This necessarily makes it not ideal for manual development, which is why we are continuing to develop tools for input and output from various platforms. We did not expand on these tools in the paper as we wanted to focus on the format itself, but if the reviewer think it is appropriate, we would be happy to discuss these tools in the paper.

9 The choice of CSV for the underlying data is simply disastrous. There is no formal description of what the CSV format is internationally and any software can decide the method, separators, text identifiers etc. that is to be interpreted. The method to be used by a software to open a CSV file depends on the localization characteristics of the medium of support!!. To put that in plain language, depending on the support of origin of your file (usb key, data hardrive, http etc.) the software will interpret its localization differently. If I open a CSV text on a french hardrive that contains http UK transfered data files, themselves generated in a lab in Danemark (eurocentric arbitrary choice of countries) the definition of the localization for the file is likely to be undefined, or defined by the last medium! This is not acceptable as an international exchange data format.

We have three objections to this comment:
1) Many of the localization problems are due to how software encodes their CSV files, not the CSV files themselves.
2) Whereas CSV doesn't have explicit standards, or we are following the conventions of the W3C's CSV on the Web working group.
3) Lastly, the localization options of the CSV are readily described in the json file that provides "instructions" on how to parse the data. The choice of delimiter is readily handled there. Again, it is important to recognize that we don't expect most users to interact directly with the data; rather, they will access the actual data via their platform of choice (e.g., Excel, R, Matlab, Python), from files created by input/output utilities. Translations to/from Matlab and R are presently operational, the others are work in progress.

10 A common example for the localization problems are the commas. On an english system, the comma might define the field (in CSV, Comma Separated Values) or the marker for thousands. On a french system, the comma defines the decimal. It naturally follows that a CSV file generated on a french system is not CSV purely since it is filed separated by a semicolon. This is most impractical and will lead to many, many errors.

See comment above. The semantic value of the Linked Data framework means that the data are shipped with 'instructions' on how to be properly read. However, we should have been more explicit about our intent to make other user-friendly files available.

11 Community support. I am extremely surprised to see that the proposal is presented by only two individuals and acknowledged, apparently, by the PAGES2K group. If that manuscript is the result of a group effort, why is this not a group co-authorship? If it is a individual proposal, how shall we expect to gather the group support and momentum that is required for the definition of something that could become a standard in the long-run? This (apparent) absence of community support and the statements made by the authors that " One goal of the present work is to spark such a discussion by giving the worldwide paleoclimate community a strawman to improve upon." makes this manuscript very inconsistent. A strawman is defined as being an "informal fallacy". Shall we, as a community accept that fallacy? Sparking a discussion is a good idea, but ideally this should not come as a formal published definition. Publishing formats first with the fallacious idea of suggesting a discussion will only lead to: a) a few groups will adopt it (like PAGES2k) b) some will not and either keep the usual format or adopt another one c) this will add layers of confusion, noise in the system and even more components to the "Digital Tower of Babel [sic]".

This is a major misunderstanding, that should have been made explicit in the manuscript. The described data format, as well as the manuscript itself, is the result of multiple iterations of community development. Certainly, some description of this should be added to the manuscript, and we're happy to do so; but let us be explicit about the community development that has supported the development of LiPD to its current point.

The reviewer acknowledges the recognized need for a standardized format for paleoclimate data; and the earliest development of these concepts arose from the clear recognition of such a need through two large community projects organized through Past Global Changes (PAGES), the PAGES 2k project, and the PAGES Arctic Holocene Transitions project. The call for standardization from the community working on these projects was clear, and PAGES has made the development of formats and standards a priority as part of its "Data Stewardship" integrated activities effort. To this end, the authors have worked with the PAGES International Program Office to reach out the large community of paleoscientists involved with PAGES, to solicit input and feedback on these ideas.

Explicitly, through PAGES we reached in multiple ways:

- 2k Mailing lists mail to >600 subscribers, 23 April
- e-news went to 5400 PAGES subscribers. It was e-news 3-2015
- Tweeted on May 8, https://twitter.com/PAGES_IPO/status/596604952363012096
- FB entry on May 8, https://www.facebook.com/PastGlobalChanges/timeline/
- "Latest news" entry on PAGES website, May 8:
http://www.pages-igbp.org/news/all-news-items/9-latest-news/1162-give-your-feedback

For the most part, we relied on the online platform Authorea, which allows online publishing, editing and feedback on manuscripts, https://www.authorea.com/users/17200/articles/19163/_show_article to share the information on this format and receive feedback. Through this process we received excellent feedback from the community that great contributed to the framework. We view this as community product that has been evolving, and that we expect will continue to evolve through discussion and feedback here on Climate of the Past Discussions.

A couple more notes:

1. **The term strawman was used in the sense of *strawman plan* (https://en.wikipedia.org/wiki/Straw_man_proposal), as in a first cut to be refined (or rebuilt altogether) based on subsequent input. Community feedback has been overwhelmingly positive, so we do not believe that the reviewer's perception is universal.**
2. **As for authorship, many community products evolve much more efficiently when a fully fledged proposal is available to evaluate, rather than by trying to build consensus from the earliest stages. The authors developed this framework, and invested many research hours into it, and it is reasonable to publish it as their idea. It is a community product, that was greatly helped by extensive helpful advice from many community members, who acknowledged in the manuscript, as well as PAGES who engaged the community on our behalf. There are no clear guidelines about what constitutes authorship on such a product; nor would a long list of authors change the extent to which this is a community-developed product.**

12 Overall, I find that the process and the definition outlined in the present manuscript are very much ill-fitting the purpose and should not be accepted. I urge the authors to seek some real community discussion and community support before engaging into formalizing it. A common group approach is to open a WiKi space, discuss over a few, tryout the definition, refine the norm definition and finally propose it for publication as a large group effort.

Again, extensive community support was involved in the development of this data structure. Climate of the Past Discussions is an excellent format for continuing this refinement, and we appreciate additional comments from the broader community here - all of which are much more positive than this reviewer. The beauty of our framework is that it can evolve organically based on community input and we look forward to taking even broader feedback to continue to evolve it to fit the needs of the community.

13 A better text delimited file format. Most of the databases (NOAA, PANGAEA) are used Tab Separated file formats, or TSV. Though being not better defined than CSV in principle, it has the massive advantage of using a simple character "tabulation" that is not used in common writing language for unique separation of fields. From a computer language perspective, it is far easier

to use. Adopting in within a utf-8 coding norm is sufficient and universal enough for the underlying data.

As written above, the CSV "problem" is a red herring, so there is no need to fix it. Although we are not convinced that TSV is fundamentally better than CSV given the discussion above (and the reviewer's own admission that tabs are not universally parsed, either), this could be simply updated if its superiority can be unequivocally proven - an example of how the structure can evolve if need be.

14 A better metadata format. The most common format used for metadata description is XML. This format has the massive advantage of being readily compatible and translatable in HTML format (an enormous advantage for user-parsing of databases), is supported by office suite software for creating standard forms and templates, both as input and output format. It is already used in many large databases in the world and hence its wide support is not to be proved. It would solves many of the problems outlined here.

XML is of course another option -- indeed, one of the authors (Emile-Geay) originally thought it would be suitable (see Emile-Geay & Eshleman, G^3, 2013). However, the simplicity and human-readability of JSON are causing it to become a leading format for data exchange on the web, with extensive community support for tools on all software platforms. The importance of XML's support in HTML and Office is overstated, as users are not expected to be editing the documents directly. Moreover, editing such documents in Office would not be an appealing or practical option, in addition to the fact that Office is not open source, and that its attendant formats keep changing (e.g. xls vs xlsx), preventing back-compatibility.

15 On the linked data approach. I have to admit I did not understand the point that the authors are trying to make on that aspect. Though I perfectly understand the advantages of linked data (like in HTML hyperlinking or XML relationships) I do not get what Figure 2 is trying to show us. The manuscript is particularly not-explicit in that regard. It is meant to "illustrate the standard more concretely". Apparently it just confuses some readers.

Figure 2 is not meant to illustrate the linked data concept, rather to show the hierarchical structure of the metadata for a real-world example. We will make this more clear. For more information about Linked Open Data, we recommend the following resources:

http://linkeddata.org/
http://www.mkbergman.com/447/what-is-linked-data/
https://vimeo.com/36752317
http://www.ted.com/talks/tim_berners_lee_on_the_next_web?language=en

16 Following all the deficiencies of the approach listed above, I can only but recommend straight rejection of the manuscript. The problem stated is very important though and should be taken up by the community, but as a community-wide approach.

Again this is a misunderstanding, that we need to make more clear in the manuscript. The Authorea comments provide examples of real community feedback. We also suggest that progress and innovation is not always most efficient in the largest groups; and large committees are often far better at rejecting ideas than developing solutions. We suggest that this is a case where a proposal needs to be put out to the community. Indeed, the extant literature on the diffusion of innovations shows that innovations are proposed by individuals or small groups, and are either adopted or rejected by the community. We are hereby putting this field-tested (cf PAGES2k) innovation to our community members, and are developing tools so they can try it in their own workflow. This is "a community-wide approach", although we acknowledge that it is (and argue that it needs to be) a different community-wide approach than the reviewer suggests.

Reply to interactive comment by Jack Williams

We thank Jack for taking the time to provide his input and support for this effort.

Reviewer 1 raises important points - can LiPD be a common format if it doesn't yet have buy-in from the community? There is a risk of proliferation of standards (see this cartoon: https://xkcd.com/927/) Of course, the paleoclimate community is dispersed across countries and proxies, so getting actual full community buy-in is impossible. And our community doesn't have any groups charged with setting community standards for paleoclimatic data - although NOAA Paleoclimatology, PANGAEA, and Neotoma (for paleoecological data) come close.

That said, I think this paper takes an important step forward. There is a need for standard data formats that can be used to pass paleoclimatic data among databases, to gather data into paleoclimatic syntheses, and to export to analytical environments such as R. This paper is an appropriate place to propose such a format. Whether the format gains widespread adoption and/or modification is still to be determined.

Some of Reviewer 1's concerns might be addressed if the purpose of LiPD was clarified. I agree with Reviewer 1 that LiPD would not be a very good data entry format. The concerns about csv format are important.

We strongly agree that we need to be explicit about the purpose of LiPD, and our vision about user's workflow and how scientist will interact with the format. We will include this in our revised manuscript.

Still, my sense is that the needs here are for textual revision (rethinking csv format and scaling back claims that this is already a community data format), rather than outright rejection. Climates of the Past seems like the natural place for this article. -Jack Williams

Reply to interactive comment by Ines Hessler

We thank Ines for her thoughtful comments and suggestions. We respond point by point below.

Overall I think the article presented by McKay and Emile-Geay reflects a great initiative that intends to facilitate research data sharing, discovery and reuse within the paleoclimate community. The authors are correct in stating that there is currently no universal way to describe, store and share paleoclimate data, which, unfortunately, also applies to most other research disciplines. However, the question is if this is due to the lack of suitable data formats, metadata standards and/or available infrastructure or due to other aspects such as competition, giving other tasks precedence or simply unawareness of how and where to share research data. As I am not a reviewer of the article by McKay and Emile-Geay, I simply would like to leave some food for thoughts.

1. In the introduction chapter the terms data standard, metadata standard, data container and format are used fairly inconsistently. At times it is hard to follow if the authors talk about data, metadata or data formats. A brief definition of the individual terms may be appropriate to ensure a consistent understanding of their meaning.

Agreed. We will be explicit about how we are defining these terms, and check for consistency in our use of them.

2. While talking about universally readable data formats the authors mention netCDF format, which is a widely applied and accepted 'self-describing, machine-independent data format that supports the creation, access, and sharing of array-oriented scientific data'. Therefore, the authors choice of the JSON-LD data format strikes me as unusual as this format, at least to my understanding, is not widely known and applied. Introducing a largely unknown/unused format to the community might result in it not being readily adopted. 3. Regarding the proposed metadata standard I am wondering about the necessity to introduce yet another standard. On the one hand netCDF offers a metadata convention for e.g. Climate and Forecast (CF) data that is easily incorporated into the netCDF files (netCDF CF 1.6) itself. On the other hand there are

several metadata standards specifically designed for geo information (e.g., ISO19115) that have several multi purpose fields that can house otherwise non assignable information.

Regarding the suitability and broad use of JSON-LD; we recognize that JSON-LD may be unfamiliar to most paleogeoscientists, however over the past decade JSON has become a leading format to describe and exchange all types data over the web due to it's simplicity, flexibility and functionality; features that are desirable for paleoclimatic data as well. The newer, linked data flavor of json allows us to leverage the semantic web, which is a particularly exciting way to integrate paleoclimate data with concepts and datasets from other fields of science, that holds the potential to lead to more efficient interdiscplinary discovery. It's unfamiliarity shouldn't be a problem for users because 1) most users won't interact with the JSON-LD directly, as we will better explain in our reply, and 2) because those interested in digging into JSON-LD component with LiPD will find it both human-readable and simple, and will be able use the incredibly rich set of utilities available for working with JSON data.

Regarding the possibility of NetCDF as a format for paleoclimate data, we recognize that it is *possible* to use it, but there several reasons why we believe it is not an ideal solution:

1.NetCDF is used extensively by climate modelers, atmospheric and ocean scientists, but it is completely opaque to most paleoclimatologists. (We once described LiPD as "NetCDF for paleo" to a renown paleoceagranographer. Her response: "What is NetCDF? I only work with Excel"). For 90% of paleoclimatologists, NetCDF is equally or even less approachable than JSON-LD.

2. NetCDF is designed for large, spatially-gridded ("array-oriented") datasets. Although it could potentially accommodate paleoclimate data, it was not designed for the highly heterogeneous datasets common to paleogeoscience, and its implementation to such datasets would be very clunky. It would make sense to force a square peg into a round hole if most holes were round and very few pegs were square. However, NetCDF is non-existent outside of the climate modeling and atmosphere/ocean sciences communities. JSON, on the other hand, is used to exchange all manner of information on the Web, making it inherently much more interoperable. Indeed, JSON modules already exist in the vast majority of modern programming languages (see http://www.json.org/), unlike NetCDF.

3. Other paleogeoscientists (e.g. paleoecologists and paleobiologists), share many of the same challenges that paleoclimatologists do. Organizations like Neotoma, Ocean Drilling Program and the Paleobiology Database have expressed interest in the LiPD framework, which they are considering adopting for their own purposes. This would make all paleogeoscientific data much more interoperable. NetCDF never had that pull with these communities.

4. While some protocols do exist to extract part of a NetCDF file via a network (.e.g OpenDAP), this information generally cannot be found unless one knows what to look for. In this sense, NetCDF files essentially create news silos (clean, organized and self-describing silos, but silos nonetheless). In contrast, JSON-LD is designed around the concept of Linked Open Data, which makes datasets instantly discoverable on the Web by broadcasting the metadata via a universal protocol (the Resource Description Format). Many organizations already archive content as Linked Open Data (e.g. the BBC, the NY Times, DBPedia), and doing so with paleoclimate data would enables it to be more easily cross-referenced with a continuously growing body of knowledge across the Web.

5. An important component of the paper is that it promotes a more structured way of storing paleoclimate data. Once this structure is achieved it is easy to translate it to other formats. If a NetCDF implementation is deemed critical, one can always write code to export from LiPD to NetCDF.

Ultimately, NetCDF is the outcome of a user group of scientists developing a data format that richly fits their need. Rather than squeezing paleoclimate data into NetCDF; the community should take the opportunity to develop the format that best fits the need of paleogeoscientists.

4. Regarding the unique identifier I would recommend looking into Digital Object identifier (DOI) that could either be associated to the individual data set or to the data collection. Most countries have central agencies, universities or research institutions that provide DOI minting services.

We agree that data DOIs would be ideal solutions, and LiPD can readily encode data DOIs as a key value pair. We however, believe that minting DOIs for datasets, rather just encoding those identifiers that were minted at institutions like the ones you described is beyond the scope of the data structure we've described here.

5. To link information from data files, metadata, authors, publications and grants I would recommend looking into the infrastructures that are already in place in universities and/or libraries. These often fairly sophisticated systems have been set up for exactly such purposes. In general it might be a good idea to involve a (local) liaison librarian, a member from the universities eResearch group (if existent) or research office, as they are often familiar with issues related to data sharing, storing, discoverability, linkage and reusability, and are potentially a good source of information. I also would recommend looking into Research Data Alliance as they intent to 'build the social and technical bridges that enable open sharing of data' across various disciplines

Thank you for this suggestion. LiPD was designed to link into existing data stores, and we will investigate the possibility of extending these data connections into LiPD.

Reply to interactive comment by Doug Fils

We thank Doug for taking the time to comment on the manuscript, and for providing his perspective on the manuscript and LiPD.

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 community 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.

We agree, please see our additional thoughts about the use of NetCDF for paleoclimate data in our response to Ines Hessler.

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 JSONLD 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 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.

Agreed, and thanks for the note. In our revised manuscript we will make the distinction between the the representation (in JSON-LD & CSV) and the model, and note that the model allows for alternate representations should that be desired, or demonstrated to be better in the future.

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.

Reply to interactive comment by editor Valerie Masson-Delmotte.

We thank the editor for her efforts to find reviewers, and for her comments on the paper.

Dear authors, Thank you for your patience! I must acknowledge that it was very difficult to find reviewers for this manuscript (only one out of 12 accepted my invitation!). Hopefully, several additional comments were provided that reflect a diversity of perspectives. Your technical paper addresses a crucial topic with respect to paleoclimate data sharing. I would like you to provide us with a revised version of the manuscript which addresses the concerns of all the comments. You may be interested to exhange with Tim Bolliet (bolliet@lsce.ipsl.fr) who has a manuscript submitted to Climate of the Past Discussion who addressed the same challenge of data sharing with a focus on stable isotope records (including web tools).

We are looking forward to reading Dr. Bolliet's manuscript when it becomes available and look forward to contacting him to discuss how we can integrate our efforts.

I note one typo (page 4322, line 7, "foraminfera" to be corrected).

Thank you, we will correct this.

In the revised version, it would be helpful to further expand on the methods to incorporate existing records into your format, and, vice-versa, tools to extract records in your format so that they can be used by office softwares.

Agreed, we will expand on this in the text. We are also working to release a first version of utilities to interact with LiPD simultaneously with our revised manuscript.

Finally, the method chosen to report the chronological metadata is not applicable to all archives, and it would be useful to adjust it so that it can include information for ice cores (e.g. reference volcanic horizons, layer counting, thinning models, tie points...). Please account for these suggestions in the revised manuscript.

The chronology component of LiPD was designed to be flexible enough to handle all chronological methods and metadata that we could conceive of, however we recognize that this may not be clear given the example and description in the manuscript. We will expand on this in our revisions.

Sincerely, Valerie Masson-Delmotte.