Climate
of the Past

Technical note: The Linked Paleo Data framework – a common tongue for paleoclimatology

Nicholas P. McKay[1] **and Julien Emile-Geay**[2]

[1]School of Earth Sciences and Environmental Sustainability, Northern Arizona University, Flagstaff, AZ 86011, USA
[2]Department of Earth Sciences & Center for Applied Mathematical Sciences, University of Southern California,
Los Angeles, CA 90089, USA

Correspondence to: Nicholas P. McKay (nicholas.mckay@nau.edu)

Abstract. Paleoclimatology is a highly collaborative scientific endeavor, increasingly reliant on online databases for data sharing. Yet there is currently no universal way to describe, store and share paleoclimate data: in other words, no standard. Data standards are often regarded by scientists as mere technicalities, though they underlie much scientific and technological innovation, as well as facilitating collaborations between research groups. In this article, we propose a preliminary data standard for paleoclimate data, general enough to accommodate all the archive and measurement types encountered in a large international collaboration (PAGES 2k). We also introduce a vehicle for such structured data (Linked Paleo Data, or LiPD), leveraging recent advances in knowledge representation (Linked Open Data).

The LiPD framework enables quick querying and extraction, and we expect that it will facilitate the writing of open-source community codes to access, analyze, model and visualize paleoclimate observations. We welcome community feedback on this standard, and encourage paleoclimatologists to experiment with the format for their own purposes.

1 Introduction

Science is entering a data-intensive era, where insight is increasingly gained by extracting information from large volumes of data (Hey, 2012). This is particularly critical in paleoclimatology, as understanding past changes in the climate system requires observations across large spatial and temporal scales. Paleoclimatic observations are typically limited to small geographic domains, so investigating large scales requires integrating many disparate studies and data sets. Observational work in paleoclimatology exemplifies the "long-tail" approach to data collection (Heidorn, 2008): the majority of observations are gathered by independent scientists with no formal language for describing their data and metadata to each other – or to machines – in a standardized fashion. This results in a "digital Tower of Babel", making the curation, access, re-use and valorization of paleoclimate data far more difficult than it should be, hindering scientific progress.

Recognizing the need for data sharing, paleoclimate investigators have made a major effort over the past decade to make their data available to the broader community, largely through online archiving systems like the http://www.ncdc.noaa.gov/paleo/wdc-paleo.html and http://www.pangaea.de/. Nonetheless, the lack of consistent formatting and metadata standards (i.e., a common tongue) has made the re-use of such data needlessly labor-intensive by preventing computers from participating in the task of making connections across data sets. As the number of records in these archives has grown, making connections manually has become more and more challenging, hampering integrative efforts at the very time they should be flourishing. Paleoclimatologists thus need a common tongue to describe their data sets to each other and to machines. Achieving this goal requires addressing two major hurdles: (1) the lack of a common language used to describe our data sets (a data standard), and (2) the lack of an accepted data format – a "rule book" that describes how the data are encoded, and that allows programmatic access to the data.

These two issues are clearly related, but somewhat distinct in practice. The data format must be universally readable, a condition satisfied by, for instance, netCDF files, which have been used for paleoclimate syntheses (Wahl et al., 2010). However, such files only allow for fixed schemas and require identical fields for all proxies. In reality, each paleoclimatic data set may have a unique set of data and metadata properties. Moreover, the netCDF format is designed for large gridded data sets, and is justifiably popular in the atmosphere and ocean science communities. However, it is unfamiliar to most paleogeoscientists, because it was not designed with the peculiarities of paleogeoscientific data in mind, and would only accommodate them with extreme effort. Further, to enhance the relevance of paleoclimate data to other fields, one would like this data container to be compatible with the linked-data paradigm (Bizer et al., 2009), which allows for data-driven discovery between data sets that would otherwise be unlikely or impossible. For the broadest applicability, we require a more flexible format.

Elaborating a data standard is an even greater challenge. It requires that the community of paleogeoscientists agree on the meaning of, and relationship between, the terms they use every day, often informally, in different contexts, and with different cultural norms. For instance, some scientists use the term "proxy" to liberally describe any paleoclimatic variable, whereas others restrict its use to relationships that have been rigorously quantified. Developing a consistent standard, "a common tongue", is critical to the community moving forward, but will be an iterative community process.

In this technical note, we present a solution to both problems, and present LiPD (Linked Paleo Data) a new, flexible linked-data format designed for paleoclimate data. Such a data container is a necessary first step towards a "semantic web of paleoclimatology" (Emile-Geay and Eshleman, 2013), and provides a straightforward framework in which communities and researchers can explicitly describe their data and metadata in common terms that the community, and computers, can understand. In the process, we introduce a preliminary data standard for paleoclimatology. Indeed, such a standard is essential to structuring the metadata, though the container is flexible enough to accommodate many revisions and updates to this standard. As discussed above, an accepted standard needs to evolve out of community-wide discussions and the establishment of a consensus, which has yet to take place in our field. One goal of the present work is to spark such a discussion by giving the worldwide paleoclimate community a blueprint to improve upon.

2　A flexible container for paleoclimate data

Paleoclimate observations come in many varieties; standardizing the data and providing the framework to encode meaning to the parameters and metadata requires a flexible, and extensible format. The linked-data variety of JavaScript Object Notation (http://json-ld.org/) provides a lightweight and human-readable solution to this problem. JSON-LD may be unfamiliar to most paleogeoscientists, but JSON is becoming a leading format for data exchange on the web, has a rich set of existing tools to interact with it, and has a robust user community. JSON-LD augments JSON by defining each property via a Web-defined schema and is being used by Google, the BBC, and Microsoft, among many other institutions. More importantly for the paleogeosciences, it is almost infinitely customizable, meaning that it can adapt to fit any data set and evolve with emerging data standards in the community. Here we present the structure of the Linked Paleo Data (LiPD) data format, which utilizes JSON-LD and provides a structure that is common to the overwhelming majority (if not all) of paleoclimate observational data sets.

Despite their variety, all paleoclimate data sets share the same major features.

1. Some basic metadata about the data set, such as

    (a) identifiers (data set name, version number, data set DOI, investigators),

    (b) archive[1] type.

2. Geographic metadata, such as

    (a) latitude, longitude, elevation above or depth below sea level;

    (b) site name.

3. Publication metadata, such as

    (a) DOI (which resolves the following information);

    (b) authors, title, journal, publication date.

4. Funding metadata, such as

    (a) funding agency,

    (b) funding grant number.

5. Proxy data and metadata, including

    (a) one or more tables of measurements, and their metadata;

    (b) variable names, units, standards, and interpretations (including forward models).

6. Geochronological data and metadata, which can include

    (a) table(s) of radiometric dating measurements and associated metadata,

    (b) age model ensembles,

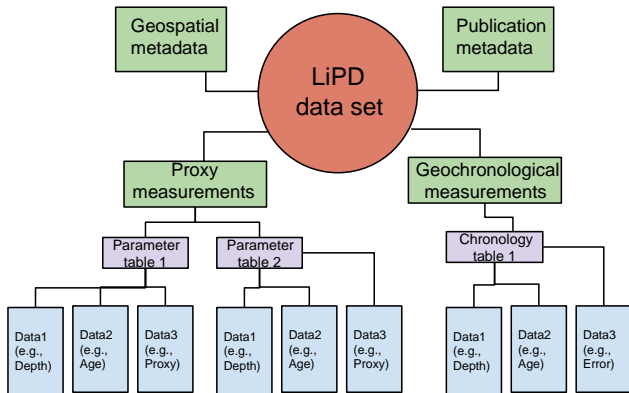   (c) author interpretation and methodological choices.

Figure 1. Schematic data model example of Linked Paleo Data.

LiPD encodes these data and metadata into a structured hierarchy that allows explicit description of any aspect of the data set at any level of the data (Fig. 1). LiPD serializes this hierarchy using JSON-LD, using nests of lists and key-value pairs. LiPD adopts the GeoJSON standard [2] to describe the geospatial metadata of a given site like this:

```
"geo": {
        "type": "Feature",
        "geometry": {
                "type": "Point",
                "coordinates": [-17.82, 62.08,
                        -1938]
        },
        "properties": {
                "siteName": "RAPiD-12-1K, South
                        Iceland Rise, northeast North
                        Atlantic"
        }
},
```

The GeoJSON standard defaults to the WGS84 ellipsoid, and units of decimal degrees for latitude and longitude and meters above sea level for elevation. This standard readily accommodates polygonal and multipoint geographic features and additional location metadata that allow for a much richer suite of geographic metadata than are typically recorded with paleoclimate data sets.

LiPD adopts the Linked Data extension of BibJSON (Johnson, 2013) to describe publication metadata, for example:

```
"pub": {
        "author": [
        {"name" : "Thornalley, D.J.R"},
        {"name" : "Elderfield, H."},
        {"name" : "McCave, N"}
        ]
        "type" : "article"
        "identifier" : [
                {"type": "doi",
                "id": "10.1038/nature07717",
                "url": "http://dx.doi.org
                        /10.1038/nature07717"}
                ],


        "year": 2009
},
```

For the paleoData and chronData components of LiPD, which include tabular data, LiPD does not store the actual tabular data in the JSON-LD format, as this becomes increasingly verbose and inefficient with large data tables. Rather, the tabular data are stored separately. Theoretically, these data could be stored in any format, so long as the information needed to read the file is included in the JSON-LD file. Presently, these data are stored in headerless comma-separated value (CSV) files that are referenced and described by the JSON-LD file, using the W3C's CSV on the Web working groups recommendations[3]; however, alternative storage models, especially cloud-based methods, may become preferable in the near future. The description of the paleoData and chronData in the JSON-LD is structured as

---

[1]The archive is the medium in which the paleoclimatic signal is imprinted: e.g., coral aragonite, ice core, sediment core.

[2]http://geojson.org

[3]https://w3c.github.io/csvw/

```
"paleoData": [{
      "paleoDataTableName": "data",
      "filename": "
          Atlantic0220Thornalley2009.csv",
      "columns": [{
            "number": 1,
            "variableName": "depth",
            "variableType": "measured",
            "description": "depth below
                ocean floor",
            "units": "cm",
            "datatype": "csvw:NumericFormat
                ",
            "notes": "depth refers to top of
                sample"
      },
      {
            "number": 2,
            "variableName": "year",
            "variableType": "inferred",
            "description": "calendar year AD
                ",
            "units": "AD",
            "datatype": "csvw:NumericFormat
                ",
            "method": "linear interpolation"
      },
      {
            "number": 3,
            "variableName": "temperature",
            "variableType": "inferred",
            "description": "sea-surface
                temperature inferred from Mg/
                Ca ratios",
            "datatype": "csvw:NumericFormat
                ",
            "material": "foraminifera
                carbonate",
            "calibration": {
                  "equation": "BAR2005: Mg/
                      Ca=0.794*exp(0.10*SST)
                      ",
                  "reference": "Barker et al
                      ., (2005), Thornalley
                      et al., (2009)",
                  "uncertainty": 1.3
            },
            "units": "deg C",
            "proxy": "Mg/Ca",
            "climateInterpretation": {
                  "variable": "T",
                  "variableDetail": "
                      seaSurface",
                  "seasonality": "MJJ",
                  "interpDirection": "
                      positive",
                  "basis": "Mg/Ca
                      calibration to SST"
            }
      }]
}]
```

Describing the columns in the data table in the LiPD framework allows explicit encoding of key metadata that are

commonly lost or misunderstood in current data structures. For example, the "climateInterpretation" section above allows the scientist to explicitly describe the details of how the variable "senses" climate. When encoded as above and explicitly defined and linked, the knowledge that this record is interpreted to record May through July sea surface temperature, and that those temperature estimates were derived from the Mg / Ca calibration equations of Barker et al. (2005) and Thornalley et al. (2009), becomes built into the data set and readable to both people and computers. It is queryable, linked to other data sets, and transparent when data sets are used in ways that are outside the published interpretations.

An advantage of using JSON as the default container for this information is that it is an extremely common vehicle for all manners of data, and can be parsed by nearly all modern programming languages. As each LiPD data set is comprised of a JSON-LD file and one or more CSV files, each data set is packaged using BagIt[4], which provides a simple format for collecting and validating files for distribution, and can be readily serialized into a compressed file for exchange between users.

3 A preliminary data standard for paleoclimatology

The flexible container described in Sect. 2 can serialize any set of paleoenvironmental data with rich metadata. However this framework only becomes useful when a common vocabulary with explicit meanings is applied to the data. Developing this vocabulary requires buy-in from experts across the disparate domains of the paleogeosciences, and will be a gradual process of evolving standards. To begin this conversation, here we outline a preliminary metadata standards for required metadata, based on phase 2 of the Past Global Changes (PAGES) past two thousand years (2k) project[5]. The following are the minimal metadata for every data set in the network. Many records include additional desirable data and metadata that are readily accommodated in LiPD. It is illustrative to look at a simple but realistic example to examine how a data set is structured in LiPD using this preliminary data standard. We use the data set of Thornalley et al. (2009) as an example in the following and in Fig. 2.

1. Base metadata

    dataSetName – name of the data set; that is, an alphanumeric string that uniquely characterizes this record in the database, often based on site, authors, year and ancillary information; example: *RAPiD-12-1K.Thornalley.2009*

    archiveType – example: *marine sediments*

    investigator – example: *David Thornalley*

2. Geospatial metadata (in GeoJSON)

---

[4]https://tools.ietf.org/html/draft-kunze-bagit-13
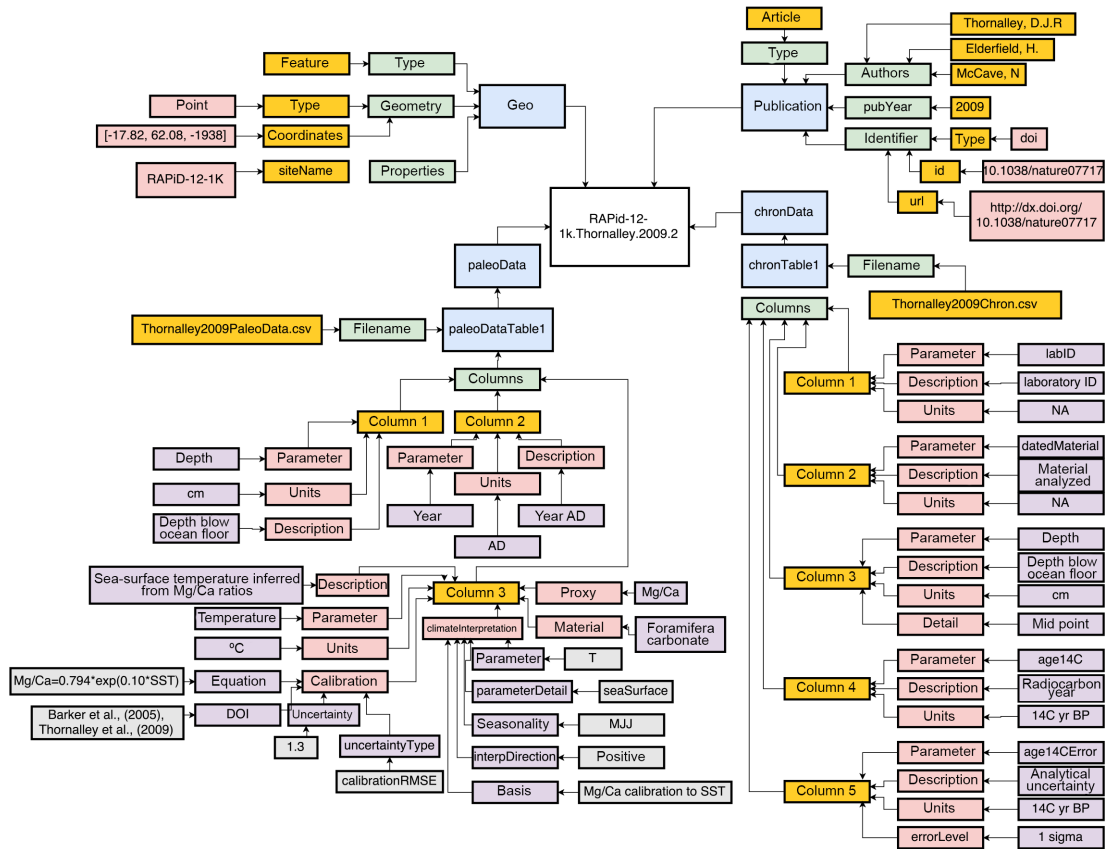[5]http://www.pages-igbp.org/workinggroups/2k-network/intro

Figure 2. Data model for RAPiD-12-1k example used in the technical note.

coordinates – longitude, latitude, and elevation, in units of decimal degrees and meters above sea level; example: *"coordinates"*: [−17.82, 62.08, −1938]

type – geographic feature type (typically "point" or "polygon"; example: *point*)

siteName – example: *RAPiD-12-1K*

3. Publication metadata (in bibJSON)

DOI – publication digital object identifier; example: *10.1038/nature07717*

citation – long publication string if DOI is unavailable; example: *Thornalley, David JR, Harry Elderfield, and Nick McCave. "Holocene Oscillations in Temperature and Salinity of the Surface Subpolar North Atlantic." Nature 457, no. 7230 (2009): 711–14.*

pubString – short text citation; example: *Thornalley et al., 2009*

4. Paleodata metadata

paleoDataTableName – short name of the paleoData table; example: *data*

filename – measurement table filename (.csv); example: *Atlantic0220Thornalley2009.csv*

chronology – short name of chronology table used in this measurement table; example: *chronology*

paleoData table columns – required in measurement table, will vary by archive and proxy type

depth – depth of sample/measurement

age/year – best estimate age/year of sample/measurements

climate-sensitive variable – measurement interpreted in terms of past climate or environmental change; example: *Mg / Ca*

paleoData column metadata – metadata characterizing each column of the paleoDataTable

column – column number; example: *3*

variableName – short variable name; example: *SST*

description – longer description of the variable; example: *sea-surface temperature inferred from Mg / Ca ratios*

units – example: *deg C*

climateInterpretation – five properties that allow for a concise description of how the climate-sensitive variable is related to climate. This is required for

at least one column in the PAGES 2k database, but may not be appropriate for all data sets.

variable – what aspect(s) of climate are recorded in this archive; example: *temperature*

variableDetail – provides detail on "climateInterpretationvariable"; example: *sea surface*

seasonality – example: *May, June, July*

interpretationDirection – do the values have a positive or negative relation to the inferred variable? Example: *positive*.

basis – quote from paper or other argument that justifies the interpretation; example: *regional core top calibration equation (Barker et al., 2005)*

5. Chronological metadata

chronDataTableName – short name of the chronData table; example: *chronology*

filename – chronology table filename (.csv); example: *Atlantic0220Thornalley2009Chronology.csv*

chronology – short name of chronology table used in this measurement table; example: *chronology*

chronData table columns – columns required in chronology table; will vary by archive and geochronological methodology type; example: ^{14}C *age*

depth depth of sample/measurement

age – radiocarbon age

ageUncertainty – analytical radiocarbon age uncertainty

datedMaterial – what was dated? (e.g., bulk sediment, terrestrial macrofossil, etc.)

chronology column metadata – metadata characterizing each column of the chronDataTable

column column number; example: *2*

variableName – short variable name; example: *14CAge*

description – longer description of the variable; example: *uncalibrated radiocarbon age*

units – example: *14C yr BP*

errorLevel – error level for uncertainty columns; example: *14C yr BP*

This is an intentionally minimal example that does not include all possible information. For example, the chronological metadata can describe any type of chronology, whether it is primarily based on tie-point constraints or layer counts. Additionally, metadata about how the ages and their uncertainties are modeled for undated layers is also readily stored, including the details needed to reproduce the analyses, and even large ensembles of simulated age–depth relations. Indeed, the need to store and share these data and metadata is a primary motivation of this effort.

4   Connectivity and compatibility

This technical note is focused on a technical description of the structure of a new data format (LiPD) and a preliminary data standard that can be used with it. Most paleogeoscientists will never want to, or need to, interact with LiPD on this level. The goal of any machine-readable data format is to reduce or eliminate the need for users to micromanage the minutiae of a data set; ideally, we would teach computers to do this for us and let us focus on the science. To achieve this goal, a rich and diverse set of utilities that can read and write LiPD to enable data exploration, analysis and visualization is needed. On the output side, we have begun developing open-source utilities for the analytical platforms that are most commonly used by paleogeoscientists (Matlab, Python, and R). On the input side, we have developed interactivity with Google spreadsheets – a free, cloud-based alternative to Microsoft Excel, recognizing that spreadsheets are the bread and butter of lab scientists, and recognizing the need for distributed editing of data/metadata (which Google spreadsheets' version control makes possible and reversible). Additionally, we have developed utilities that convert data sets formatted for the World Data Center for Paleoclimatology in Microsoft Excel into LiPD files, so that users who format their data for the former can instantly turn them into LiPD. Conversely, a partnership with WDC Paleo will ensure that LiPD-formatted data sets are easily archived on their site.

These utilities are in various stages of development, and are available as a public GitHub repository[6]. They are all designed to plug into the workflow of paleogeoscientists. Our hope is that as paleogeoscientists discover and explore the utility of this framework, the community of contributors will continue to expand; for example, LiPD integration with Neotoma[7] and the Neotoma R package (Goring et al., 2015) is planned for 2016. Finally, LiPD is the backbone of the LinkedEarth project[8], which will enable users to edit data sets via an intuitive wiki platform, leveraging the flexibility of LiPD while eliminating its complexity from the user experience.

5   Discussion

The data container and preliminary data standard described here are extremely flexible, and can accommodate any paleoclimatic or paleoenvironmental data that are based on any expansion of dependent/independent variable pairs. This encompasses all paleoclimate and paleoenvironmental data sets that we can imagine. A primary challenge for developing a sufficiently broad paleodata framework has long been defining and agreeing on all of the relevant terms for such a diverse community. The framework presented here addresses

---

[6]https://github.com/nickmckay/LiPD-utilities

[7]http://www.neotomadb.org/

[8]http://www.linked.earth

the first challenge by accommodating the complexity and inevitable proliferation of terms, variables and interpretations inherent to the interdisciplinary field of paleoclimatology, and by assigning explicit meaning to the terms through the Linked Open Data framework. Implementation of these semantics will be an evolving, community-driven process. This is critical for two reasons: first, defining an ontology[9] a priori has proven impossible to date; second, even if it were possible, such an ontology would be meaningless if it were not used. We will thus rely on usage and community discussion to reach agreement on terminology, and the community has clearly demonstrated its desire and willingness to contribute to these discussions.

Indeed, LiPD and the preliminary data standard discussed in this technical are the outcome of considerable community input and development. The concepts formalized here have emerged from half a decade of formal and informal development with hundreds of paleogeoscientists. The earliest formal development of these concepts arose from the clear recognition of the 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. Feedback on early versions of the LiPD framework and the preliminary data set was cultivated through the PAGES International Program Office, who reached out to the large community (> 5000) of paleoscientists involved with PAGES to solicit input and feedback on these ideas.

For the most part, we gathered input through the online platform Authorea, which allows online publishing, editing and feedback on manuscripts[10], to share the information on this format and receive feedback. Through this process we received excellent feedback from the community (acknowledged below) that greatly contributed to the framework. We view LiPD as community product that evolved prior to submission, continued to evolve through the discussion process in this journal, and will continue to evolve as more and more scientists use and critique it.

A second long-standing challenge has been managing the appropriate level of detail (lumping versus splitting) in the terminology. The LiPD framework readily accommodates this by adopting a hierarchical structure that starts with more general terminology and allows further detail to be assigned deeper in the structure. Consider the example of two δ^{18}O data sets, one measured on a coral archive and the other derived from foraminifera extracted from a marine sediment core. On the one hand, these two records are measuring the

---

[9]A formal definition of all the concepts used by the data model, and the relationships between these concepts.

[10]https://www.authorea.com/users/17200/articles/19163/_show_article

same variable, and there are times when researchers might be interested in investigating all δ^{18}O regardless of the details of the archive on which they were measured. On the other hand, there are some important differences between the two measurements that users would like to include in the data repository. If we were to describe each variable in a single term, we would have to decide whether to call them both "δ^{18}O", or to call one "δ^{18}O-skeletal aragonite" and the other "δ^{18}O-foraminifera $> 120\,\mu$m size class". By taking advantage of JSON's capacity to build hierarchical metadata structures, we can encode an entire set of metadata at the appropriate level in the data set as

```
{
    "variableName": "d18O",
    "description": "d18O measured on skeletal
        aragonite",
    "units": "permil",
    "standard": "VSMOW",
    "material": "skeletal aragonite",
    "instrument": "Micromass Optima gas
        source triple-collector mass
        spectrometer"
},
```
and
```
{
    "variableName": "d18O",
    "description": "d18O measured on G.
        bulloides > 120 microns",
    "units": "permil",
    "standard": "VSMOW",
    "material": "foraminifera calcite",
    "instrument": "Micromass Optima gas
        source triple-collector mass
        spectrometer"
    "species":"Globigerina bulloides"
},
```

This makes the commonalities and differences between the data sets explicit. Moreover, additional levels of metadata may be introduced into the descriptor to accommodate climate interpretation, calibration procedures or forward models as described above. The power of the hierarchical structure is that it allows the metadata to be placed at the appropriate level, avoiding logical contradictions in lumping and splitting that become necessary when trying to incorporate information from several levels into a single term – or when several users describe the same data set in slightly different ways.

An important consideration for re-use and provenance tracking is versioning: each version of a LiPD record, or collection of LiPD records, should be associated with a unique identifier, which is crucial to reproducibility. We propose the following versioning scheme:

Individual records – a number of the form $I_1.I_2.I_3$, where I_1 is an integer associated with a publication (e.g., Thornalley et al., 2009), I_2 is a counter updated every time a modification is made to the data and I_3 is another

counter updated whenever a modification is made only to the metadata.

Data compilations – a number of the form $C_1.C_2.C_3$, where C_1 is an integer associated with a publication (e.g., PAGES2k Consortium, 2013), C_2 is a counter updated every time a record is added or removed, and C_3 is a counter updated every time a modification is made to the data or metadata in an individual record.

We are presently implementing a large-scale test of the LiPD framework by using it as the primary data archive for phase 2 of the PAGES2k global temperature database (PAGES2K Consortium, 2016). Consequently, the described framework for describing the proxy data is fairly mature and field-tested. It also means that a large (> 600 data sets), robust collection of LiPD files will soon be available publicly. LiPD has evolved to meet the needs of these diverse data; however, it may not be universal, and we welcome suggestions for increased generality.

The standards for reporting and storing geochronological data are much less tested and will require far more community input. For instance, there seems to be no universal way of reporting radiocarbon, U / Th, or ^{210}Pb dates. Ideally, coordination between geochronologists would yield a universal standard for all radiometric age models; however, if there is to be any standard, it is more likely to first emerge within each sub-community. JSON-LD is flexible enough to encompass any possibility, but doing so in a way that allows research algorithms to easily read those chronologies and generate age models from them will likely require more work.

Finally, it is important to realize that the JSON-LD implementation described here is just one implementation to represent the underlying data model. One of the many features of linked open data is that the same data model could be serialized into other representations, such as XML or Turtle, without any loss of information. This makes this framework incredibly flexible and allows the community to move forward with implementing these concepts without trying to predict community needs and the evolution of technology. LiPD is not a rigid container that one must force paleoclimate data into, but rather a flexible system designed to wrap around a data set. We are committed to the continued development and expansion of LiPD and look forward to evolving this preliminary data standard with input from the community.

The Supplement related to this article is available online at doi:10.5194/cp-12-1093-2016-supplement.

References

Barker, S., Cacho, I., Benway, H., and Tachikawa, K.: Planktonic foraminiferal Mg/Ca as a proxy for past oceanic temperatures: a methodological overview and data compilation for the Last Glacial Maximum, Quaternary Sci. Rev., 24, 821–834, doi:10.1016/j.quascirev.2004.07.016, 2005.

Bizer, C., Heath, T., and Berners-Lee, T.: Linked Data – The Story So Far, Int. J. Semant. Web Inf., 5, 1–22, doi:10.4018/jswis.2009081901, 2009.

Emile-Geay, J. and Eshleman, J. A.: Toward a semantic web of paleoclimatology, Geochem. Geophy. Geosy., 14, 457–469, doi:10.1002/ggge.20067, 2013.

Goring, S., Dawson, A., Simpson, G., Ram, K., Graham, R., Grimm, E., and Williams, J.: Neotoma: A programmatic interface to the Neotoma Paleoecological Database, Open Quaternary, 1, 1–17, 2015.

Heidorn, P. B.: Shedding Light on the Dark Data in the Long Tail of Science, Library Trends, 57, 280–299, doi:10.1353/lib.0.0036, 2008.

Hey, T.: The Fourth Paradigm – Data-Intensive Scientific Discovery, in: E-Science and Information Management, 1–1, Springer Science Business Media, doi:10.1007/978-3-642-33299-9_1, 2012.

Johnson, T.: Indexing linked bibliographic data with JSON-LD, BibJSON and Elasticsearch, Code4lib Journal, 19, 1–11, 2013.

PAGES2k Consortium: Continental-scale temperature variability during the past two millennia, Nat. Geosci., 6, 339–346, doi:10.1038/ngeo1797, 2013.

PAGES2K Consortium: A global multiproxy database for temperature reconstructions of the Common Era, Scientific Data, in prep., 2016.

Thornalley, D. J. R., Elderfield, H., and McCave, I. N.: Holocene oscillations in temperature and salinity of the surface subpolar North Atlantic, Nature, 457, 711–714, doi:10.1038/nature07717, 2009.

Wahl, E. R., Anderson, D. M., Bauer, B. A., Buckner, R., Gille, E. P., Gross, W. S., Hartman, M., and Shah, A.: An archive of high-resolution temperature reconstructions over the past 2+ millennia, Geochem. Geophy. Geosy. 11, Q01001, doi:10.1029/2009gc002817, 2010.