PaCTS 1.0: a crowdsourced reporting standard for paleoclimate data

Khider, D., Emile-Geay, J., McKay, N. P., Gil, Y., Garijo, D., Ratnakar, V., Alonso-Garcia, M., Bertrand, S., Bothe, O., Brewer, P., Bunn, A., Chevalier, M., Comas-Bru, L., Csank, A., Dassié, E., DeLong, K., Felis, T., Francus, P., Frappier, A., Gray, W., Goring, S., Jonkers, L., Kahle, M., Kaufman, D., Kehrwald, N. M., Martrat, B., McGregor, H., Richey, J., Schmittner, A., Scroxton, N., Sutherland, E., Thirumalai, K., Allen, K., Arnaud, F., Axford, Y., Barrows, T. T., Bazin, L., Pilaar Birch, S. E., Bradley, E., Bregy, J., Capron, E., Cartapanis, O., Chiang, H. W., Cobb, K., Debret, M., Dommain, R., Du, J., Dyez, K., Emerick, S., Erb, M. P., Falster, G., Finsinger, W., Fortier, D., Gauthier, N., George, S., Grimm, E., Hertzberg, J., Hibbert, F., Hillman, A., Hobbs, W., Huber, M., Hughes, A. L. C., Jaccard, S., Ruan, J., Kienast, M., Konecky, B., Le Roux, G., Lyubchich, V., Novello, V. F., Olaka, L., Partin, J. W., Pearce, C., Phipps, S. J., Pignol, C., Piotrowska, N., Poli, M. S., Prokopenko, A., Schwanck, F., Stepanek, C., Swann, G. E. A., Telford, R., Thomas, E., Thomas, Z., Truebe, S., von Gunten, L., Waite,

PaCTS 1.0: A Crowdsourced Reporting Standard for Paleoclimate Data

D. Khider[1,2,*], J. Emile-Geay[2], N.P. McKay[3], Y. Gil[1], D. Garijo[1], V. Ratnakar[1], M. Alonso-Garcia[4], S. Bertrand[5], O. Bothe[6], P. Brewer[7], A. Bunn[8], M. Chevalier[9], L. Comas-Bru[10,11], A. Csank[12], E. Dassié[13], K. DeLong[14], T. Felis[15], P. Francus[16], A. Frappier[17], W. Gray[18], S. Goring[19], L. Jonkers[15], M. Kahle[20], D. Kaufman[3], N. M. Kehrwald[21], B. Martrat[22,23], H. McGregor[24], J. Richey[25], A. Schmittner[26], N. Scroxton[27], E. Sutherland[28], K. Thirumalai[29], K. Allen[30], F. Arnaud[31], Y. Axford[32], Timothy T. Barrows[24], L. Bazin[18], S.E. Pilaar Birch[33], E. Bradley[34], J. Bregy[35], E. Capron[36], O. Cartapanis[37], H.-W. Chiang[38], K. Cobb[39], M. Debret[40], R. Dommain[41], J. Du[26], K. Dyez[42], S. Emerick[43], M. P. Erb[3], G. Falster[44], W. Finsinger[45], D. Fortier[46], Nicolas Gauthier[47], S. George[48], E. Grimm[49], J. Hertzberg[50], F. Hibbert[51], A. Hillman[52], W. Hobbs[53], M. Huber[54], A.L.C. Hughes[55,56], S. Jaccard[37], J. Ruan[57], M. Kienast[58], B. Konecky[59], G. Le Roux[60], V. Lyubchich[61], V.F. Novello[43], L. Olaka[62], J.W. Partin[63], C. Pearce[64], S.J. Phipps[65], C. Pignol[31], N. Piotrowska[66], M.-S. Poli[67], A. Prokopenko[68], F. Schwanck[69], C. Stepanek[70], G. E. A. Swann[71], R. Telford[72], E. Thomas[73], Z. Thomas[74], S. Truebe[75], L. von Gunten[76], A. Waite[77], N. Weitzel[78], B. Wilhelm[79], J. Williams[80], J.J. Williams[81], M. Winstrup[82], N. Zhao[83], Y. Zhou[84].

[1]Information Sciences Institute, University of Southern California, Marina del Rey, California, USA, [2]Department of Earth Sciences, University of Southern California, Los Angeles, California, USA, [3]School of Earth and Sustainability, Northern Arizona University, Flagstaff, Arizona, USA, [4]Department of Geology, University of Salamanca, Salamanca, Spain, [5]Renard Centre of Marine Geology, Ghent University, Ghent, Belgium, [6]Helmholtz-Zentrum Geesthacht, Geesthacht, Germany, [7]Laboratory of Tree-Ring Research, Tuscon, Arizona, USA, [8]Western Washington University, Bellingham, Washington, USA, [9]University of Lausanne, Lausanne, Switzerland, [10]School of Earth Sciences, University of College Dublin, Belfied, Ireland, [11]School of Archaeology, Geography and Environmental Sciences, Reading University, United Kingdom, [12]University of Nevada, Reno, Nevada, USA, [13]CNRS, Bordeaux University, Bordeaux, France, [14]Louisiana State University, Baton Rouge, Louisiana, USA, [15]MARUM - Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany, [16]Institut National de la Recherche Scientifique, Québec, QC, Canada, [17]Geosiences, Skidmore College, Saratoga Springs, New York, USA, [18]Laboratoire des Sciences du Climat et de l'Environnement (LSCE/IPSL), Gif-sur-Yvette, France, [19]Univerisity of Wisconsin-Madison, Madison, Wisconsin, USA, [20]Physical Geography, University Freiburg, Freiburg, Germany, [21]US Geological Survey, Geosciences and Environmental Change Science Center, Denver, Colorado, USA, [22]Department of Environmental Chemistry, Institute of Environmental Assessment and Water Research, Spanish Council for Scientific Research, Barcelona, Spain, [23]Department of Earth Sciences, University of Cambridge, Cambridge, United Kingdom, [24]School of Earth, Atmospheric and Life Sciences, University of Wollongong, Wollongong, Australia, [25]US Geological Survey, St. Petersburg, Florida, USA, [26]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Oregon, USA, [27]School of Earth Sciences, University College Dublin, Dublin, Ireland, [28]US Forest Service, Rocky Mountain Research Station, Jemez Pueblo, New Mexico, USA, [29]Department of Geosciences, University of Arizona, Tucson, Arizona, USA, [30]University of Melbourne, Richmond, Victoria, Australia, [31]EDYTEM, Université Grenoble Alpes, University Savoie Mt Blanc, CNRS, Chambery, France, [32]Department of Earth and Planetary Sciences, Northwestern University, Evanston, Illinois, USA, [33]Department of Geography, University of Georgia, Athens, GA, USA, [34]Department of Computer Science, University of Colorado, Boulder, Colorado, USA, [35]Department of Geography, Indiana University – Bloomington, Bloomington, Indiana, USA, [36]Physics of Ice, Climate and Earth, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark, [37]Institute of Geological Sciences, University of Bern, Bern, Switzerland, [38]Department of Geosciences, National Taiwan University, Taipei City, Taiwan, [39]School of Earth and Atmospheric Sciences, Georgia Tech, Atlanta, Georgia, USA, [40]Université de Rouen Normandie, France, [41]Institute of Geosciences, University of Potsdam, Potsdam, Germany, [42]Earth and Environmental Sciences, University of Michigan, Ann Arbor, Michigan, USA, [43]Instituto de Geociências, Laboratório de Sistemas Cársticos, Universidade de São Paulo, São Paulo, Brazil, [44]The University of Adelaide, Adelaide, Australia, [45]ISEM, CNRS, University Montpellier, Montpellier, France, [46]Département de géographie, Université de Montréal, Montréal, Québec, Canada,

[47]Shcool of Human Evolution and Social Change, Arizona State University, Tempe, Arizona, USA, [48]National Center for Atmospheric Science (NCAS), Department of Meteorology, University of Reading, Reading, United Kingdom, [49]Department of Earth Sciences, University of Minnesota, Minneapolis, Minnesota, USA, [50]Department of Ocean, Earth, and Atmospheric Sciences, Old Dominion University, Norfolk, Virginia, USA, [51]Research School of Earth Sciences, The Australian National University, Canberra, Australia, [52]School of Geosciences, University of Louisiana at Lafayette, Lafayette, Louisiana, USA, [53]Antarctic Climate and Ecosystems Cooperative Research Center, University of Tasmania, Hobart, Australia, [54]Earth, Atmospheric, and Planetary Sciences Department, Purdue University, West Lafayette, Indiana, USA, [55]Department of Geography, School of Environment, Education, and Development, University of Manchester, Manchester, United Kingdom, [56]Bjerknes Center for Climate Research, University of Bergen, Bergen, Germany, [57]School of Earth Sciences and Engineering, Sun Yat-sen University, Guangzhou, China, [58]Department of Oceanography, Dalhousie University, Halifax, Canada, [59]Earth and Planetary Sciences, Washington University in St. Louis, St. Louis, Missouri, USA, [60]EcoLab UMR5245 CNRS-Université de Toulouse, France, [61]University of Maryland Center for Environmental Science, Cambridge, Maryland, USA, [62]Geology Department, University of Nairobi, Nairobi, Kenya, [63]Institute for Geophysics, the University of Texas at Austin, Austin, Texas, USA, [64]Department of Geoscience, Aarhus University, Aarhus, Denmark, [65]Institue for Marine and Antarctic Studies, University of Tasmania, Hobart, Australia, [66]Institute of Physics-CSE, Silesian University of Technology, Gliwice, Poland, [67]Department of Geography and Geology, Eastern Michigan University, Ypsilanti, Michigan, USA, [68]Institut für Geologie und Mineralogie, University of Cologne, Cologne, Germany, [69]Centro Polar e Climatico, UFRGS, Rio Grande do Sul, Brazil, [70]Alfred Wegener Institute – Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany, [71]School of Geography, University of Nottingham, Nottingham, United Kingdom, [72]Department of Biological Sciences, Bergen University, Bergen, Germany, [73]British Antarctic Survey, Cambridge, United Kingdom, [74]School of Biological, Earth, and Environmental Science, UNSW, Sydney, Australia, [75]Arizona State Parks and Trails, Benson, Arizona, USA, [76]PAGES International Project Office, Bern, Switzerland, [77]ANGARI Foundation, West Palm Beach, Florida, USA, [78]Institute of Environmental Physics, Heidelberg University, Heidelberg, Germany, [79]Université Grenoble Alpes, CNRS, IRD, Grenoble, INP, IGE, Grenoble, France, [80]Department of Geography, University of Wisconsin Madison, Madison, Wisconsin, USA, [81]Department of Social Sciences, Oxford Brookes University, Oxford, United Kingdom, [82]University of Copenhagen, Copenhagen, Denmark, [83]Max Planck Institute for Chemistry, Mainz, Germany, [84]Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, USA.

*Corresponding author: Deborah Khider (khider@usc.edu)

Key Points:

- First version of a crowdsourced reporting standard for paleoclimate data
- The standards arose through collective discussions, both in-person and online, and via an innovative social platform
- The standard helps meet the interoperability and reuse criteria of FAIR (Findable, Accessible, Interoperable, and Reusable).

Abstract

The progress of science is tied to the standardization of measurements, instruments, and data. This is especially true in the Big Data age, where analyzing large data volumes critically hinges on the data being standardized. Accordingly, the lack of community-sanctioned data standards in paleoclimatology has largely precluded the benefits of Big Data advances in the field. Building upon recent efforts to standardize the format and terminology of paleoclimate data, this article describes the Paleoclimate Community reporTing Standard (PaCTS), a crowdsourced reporting standard for such data. PaCTS captures which information should be included when reporting paleoclimate data, with the goal of maximizing the reuse value of paleoclimate datasets, particularly for synthesis work and comparison to climate model simulations. Initiated by the LinkedEarth project, the process to elicit a reporting standard involved an international workshop in 2016, various forms of digital community engagement over the next few years, and grassroots working groups. Participants in this process identified important properties across paleoclimate archives, in addition to the reporting of uncertainties and chronologies; they also identified archive-specific properties and distinguished reporting standards for new vs. legacy datasets. This work shows that at least 135 respondents overwhelmingly support a drastic increase in the amount of metadata accompanying paleoclimate datasets. Since such goals are at odds with present practices, we discuss a transparent path towards implementing or revising these recommendations in the near future, using both bottom-up and top-down approaches.

1. Introduction

Paleoclimatology is a highly integrative discipline, often requiring the comparison of multiple datasets and model simulations to reach fundamental insights about the climate system. Currently, such syntheses are hampered by the time and effort required to transform the data into a usable format for each application. This task, called "*data wrangling*", is estimated to consume up to 80% of researcher time in some scientific fields (Dasu and Johnson, 2003), an estimate commensurate with the experience of many paleoclimatologists, particularly at the early-career stage. Wrangling involves not only identifying missing values or outliers in the time series, but also searching multiple databases for the scattered records, contacting the original investigators for the missing data and metadata, and organizing the data into a machine-readable format. Further, this wrangling requires an understanding of each dataset's originating field and its unspoken practices, and so cannot be easily automated or outsourced to unskilled labor or software. There is therefore an acute need for standardizing paleoclimate datasets.

Indeed, standardization accelerates scientific progress, particularly in the era of Big Data, where data should be Findable, Accessible, Interoperable, and Reusable (FAIR, Wilkinson et al., 2016). Standardization is critical to efficiently query databases and analyze and plot results of analyses, to remove participation barriers for new scientists or people outside the specific field by explicitly describing the data rather than relying on unspoken conventions, to reduce unintended errors in data management, and to ensure appropriate and complete citations for the work of the original authors. While the paleoclimate community has made great strides in this direction (e.g., Williams et al., 2018), much work remains. The recent adoption of the FAIR data principles (Wilkinson et al., 2016) by the American Geophysical Union (Stall et al., 2017) elevates the urgency of defining what data and metadata should be archived, and how. This article proposes a community-recommended set of preliminary reporting standards and an open platform to determine which metadata are important for public archival, with an eye towards maximizing the long-term value of hard-earned paleoclimate observations and ensuring optimal reuse.

The need for standardization in paleoclimate research is beyond vocabulary agreement. Consider the editorial of Wolff (2007), which tackled the ambiguous definition of time in the paleoclimate community. The notation "before present (BP)" has become a *de fact*o "standard" in the community, although "present" means different things to different people. It is often taken as Common Era (CE) 1950 (especially within the radiocarbon community), undefined, or defined as some other date (e.g. CE 2000), or the year the study was performed/published. For studies spanning several million years with age uncertainties in excess of 1000 years, a 50-year difference is immaterial. However, for studies working at higher resolution (e.g, decadal to subannual), concentrating on recent millennia, this difference is consequential. Thus, an agreement over the precise meaning of the term "present" turns out to be critical to many uses of these datasets. The same can be said of many other metadata properties, underscoring the need for common practices in paleoclimate data reporting.

Given this acute need for standardization, the National Science Foundation (NSF) EarthCube-funded LinkedEarth project nucleated a discussion on data reporting practices. EarthCube (2015) defines a standard as "a public specification documenting some practice or technology that is adopted and used by a community." The emphasis on community and practice underlines the cooperative nature of standard development. If only one person uses a technical specification, it is not a standard. If it is voted on but not applied in practice, it is of little practical use.

Standardization requires three distinct elements: (1) a standard format for the data, (2) a standard terminology for metadata, and (3) standard guidelines for reporting paleoclimate data (i.e., reporting standards). We note that some prior knowledge of standardization practices (e.g., which data to include) can be useful in the planning stages of data collection. As an analogy, consider the organization of library cards into an old-fashioned file cabinet. For this system to function, one needs (1) a set of compartments and drawers to house the information; (2) labels to identify and classify the contents of the drawers; and (3) a disciplined adherence to the classification system. This entails including essential information required for application and re-use of the cards and the information they contain. In other words, every user follows similar guidelines to generate, use, and file the cards, otherwise the classification falls apart and the cards may as well be stored in a random pile.

This article focuses on the last requirement, namely the creation of standards for reporting paleo data and metadata. It builds upon recent efforts to address the first two points. On the first point, the Linked PaleoData format (LiPD, McKay and Emile-Geay, 2016) and derived vocabulary agreements to describe paleoclimate data (the LinkedEarth Ontology, Emile-Geay et al., 2019) provide a data container for paleoclimate data (Section 2), which is currently used in a range of data analysis software (Bradley et al., 2018, Khider et al., 2018a, McKay et al, 2018). On the second point, the National Oceanic and Atmospheric Administration (NOAA) World Data Service for Paleoclimatology (WDS-Paleo) has created a set of standard names to document paleoclimate variables, the Paleoenvironmental Standard Terms (PaST) Thesaurus (National Oceanographic and Atmospheric Administration, 2018).

This article's aim is twofold: firstly, to provide a snapshot of the first version of the Paleoclimate Community reporTing Standard (PaCTS), as of 2019, with the understanding that this standard will eventually evolve; secondly, to document the process of community elicitation of such guidelines, so as to provide maximum transparency on why and how these decisions were made. We start from the premise that sampling decisions predate these reporting decisions, so the standard aims to guide an investigator's decisions as to how they should report existing measurements, e.g., at the time of publication.

The remaining sections are organized as follows: Section 2 summarizes the relevant prior standardization efforts, which serve as the foundation for PaCTS v1.0. Section 3 describes the standardization process, including eliciting community feedback. Section 4 presents recommendation from a group of 135 international researchers actively engaged in paleoclimate research. Section 5 illustrates the application of PaCTS v1.0 to an existing paleoclimate record. Finally, Section 6 concludes with a plan to disseminate the first version of PaCTS within the paleoclimate community and provides a roadmap for further standards development and their future applications.

2. Background

2.1. The LinkedEarth Framework: an online approach to standard development

The LinkedEarth project established an online platform (Gil et al., 2017) that enables the curation of metadata for publicly accessible datasets by experts and fosters the development of terminology agreements and standards for paleoclimate metadata. Our approach builds on two synergistic elements: (1) the LinkedEarth Ontology (Emile-Geay et al., 2019), which provides an unambiguous structure and terminology to describe the metadata of a paleoclimate dataset; and 2) the LinkedEarth Platform (Gil et al, 2017), which enables the collaborative authoring of highly-structured metadata about paleoclimate datasets using the terms in the LinkedEarth Ontology.

The LinkedEarth Ontology represents vocabulary agreements to describe paleoclimate metadata. In a domain like paleoclimatology, we usually can distinguish the different kinds of objects that we want to describe (i.e., a sample, a measurement, a dataset, etc…) and the relationships used to describe those objects (e.g., a measurement is taken from a sample and therefore they are related, the measurement in is a dataset and therefore they are related, etc…). An ontology is a formal way to represent objects and their properties, and they represent consensual knowledge that helps a community describe major concepts in the domain using common terms. Specifically, an ontology formalism allows the representation of objects types as "classes", and relationships as "properties" of those classes. Classes can have subclasses, and a given class can be a subclass of several classes. For example, the class "proxy archive" can have "coral" as a subclass, and the class "repository item" can have "sample" as a subclass". A feature of ontologies is that they allow the creation of machine-readable metadata, i.e., data descriptions that can be queried programmatically by machines to retrieve datasets of interest. Thanks to the ontology, machines can navigate through metadata and discover data that otherwise would be hidden to them. LinkedEarth relies on semantic web technologies to represent ontologies, specifically the Web Ontology Language (OWL) standard of the World Wide Web Consortium (W3C) (W3C OWL Working Group, 2012). More details are provided in Emile-Geay et al. (2019).

The LinkedEarth Platform allows users to 1) describe paleoclimate datasets using the terms available in the LinkedEarth Ontology, and 2) propose new terms if they cannot find an appropriate one in the ontology. The LinkedEarth Platform is a *sociotechnical system*, and as such it provides technology infrastructure coupled with social processes that support terminology and standards convergence. When users describe a paleoclimate dataset, the terms in the existing LinkedEarth Ontology are offered to them as editable forms and completion commands, which promotes adoption. If a user does not find a term that is appropriate for their dataset, they can create a new term on the fly. Such new terms can then be discussed on the platform, building community consensus on their definitions and the essential status of their inclusion to a dataset. The social extensions of the LinkedEarth Platform allow working groups to organize activities by users with similar expertise to build a common vocabulary. Each working group was assigned a special page on the LinkedEarth Platform to nucleate their activities, including discussions and polls for rapid community feedback. The terms discussed within these working groups form the crowdsourced part of the LinkedEarth Ontology. The social editorial processes eventually will lead to a new version of the LinkedEarth Ontology. The LinkedEarth Platform and its associated social processes are described in detail in Gil et al. (2017).

The LinkedEarth Platform is implemented as an extension of the Semantic MediaWiki framework (Krötzsch and Vrandečić, 2011), Semantic wikis augment traditional wikis with the ability to structure information through: 1) semantic annotations, which enable the assignment of a class (or category) to an object in a wiki page, and properties (or qualifiers) that are useful to describe that object; and 2) automated reasoning capabilities that exploit those annotations to organize the wiki's knowledge (Gil, 2013). For example, if the page for “Los Angeles” is annotated as being in the class “city” and having a property “location=California”, and the page for “California” has a property that “location=US” then the semantic wiki can infer that Los

Angeles is in the US even though that was not explicitly stated. Semantic wiki pages can also include queries that are executed when the page is visited, so dynamic content is created in a way that is up to date with the latest additions. Semantic wikis also have facilities to track edits together with the data and contributor, so that the provenance of edits can be examined and undesirable ones can be easily undone. The content of semantic wikis becomes part of the open Semantic Web, as it can be published as a set of linked Web objects in the Web of Data, following Linked Data Principles (Heath and Bizer 2011). With this approach, the metadata for all paleoclimate datasets defined in the wiki becomes openly available on the Web, machine readable, and can be queried programmatically by any application. More details are provided in Gil et al. (2017).

2.2. Previous and concurrent efforts towards a data standard

The discussion below is non-exhaustive and only focuses on the relevant efforts that have sparked the discussion about PaCTS.

2.1.1. Origins of a standard format for paleoclimate data

Climate modeling has greatly benefitted from the netCDF data format (Unidata, 2019), designed to support the creation, access, and sharing of array-oriented data, including climate model output. Despite the importance of paleoclimate data availability for model evaluation (Masson-Delmotte et al, 2013), until recently there was no universal container to describe, store, and share these datasets. Emile-Geay & Eshleman (2013) first introduced the idea of a flexible container, where metadata would be stored semantically with the numeric data in tabular form. This concept was the basis for the Linked Paleo Data (LiPD) format (McKay and Emile-Geay, 2016).

LiPD is a universally readable data container that organizes paleoclimate data and metadata in a uniform way. It is based on JSON-LD (JavaScript Object Notation for Linked Data), a JSON-based format compliant with the Linked Data paradigm. JSON is a lightweight data interchange format that is easy for humans and machines alike to read and write. LiPD has six distinct components: root metadata (e.g., dataset name, investigator, version); geographic metadata (e.g., coordinates, descriptive location such as a country or city); publication metadata (e.g., authors, title, journal, DOI); funding metadata (e.g., funding agency and grant number); PaleoData, which includes all the measured (e.g., Mg/Ca) and inferred (e.g., sea surface temperature) paleoenvironmental data; and ChronData, which mirrors PaleoData for information pertaining to age. These components provide the rigidity necessary to write robust codes around the format, while remaining extensible enough to capture (meta)data as rich as the users want to provide for them. Utilities in Matlab, Python, and R (Heiser et al., 2018) allow users to interact with the files (specifically, to read, write, query, or filter datasets matching specified conditions).

In many ways, LiPD is intended to be the netCDF of paleoclimate observational data. However, although both LiPD and the LinkedEarth Ontology provide a standard way to describe a paleoclimate dataset, they say little about what information should be stored to ensure re-use. The endorsement of netCDF by a broad community further benefited from the adoption of the Climate and Forecast (CF) conventions (Gregory, 2003). The CF conventions define metadata describing what the data in each variable represents, and the spatial and temporal properties of the data. In other words, it defines both a set of common terms (a standard vocabulary) and a reporting standard. Efforts toward standardization of common terms have been undertaken by WDS-Paleo in the form of the PaST thesaurus (National Oceanographic and Atmospheric Administration, 2018), which provides the preferred option for a standardized name and definition. PaCTS details a crowdsourced approach for deciding what information should be included when reporting paleoclimate data, a “CF convention” for paleoclimate datasets.

2.1.2. Archive-focused initiatives

Attempts at paleoclimate data standardization have a long history. For datasets derived from wood archives, LinkedEarth relied on the tree-ring data standard, TRiDaS (Jansma et al., 2010), which complies with established data standards such as Dublin Core (DCMI Usage Board, 2008). The TRiDaS project aimed at defining the properties that are used in the dendro community and give them a consistent name (i.e., a controlled vocabulary) and identifying whether the quantity should be mandatory and repeatable (i.e., best practices). These efforts help inform the PaCTS one for wood archives, though it should be noted that tree-ring science is far broader than dendroclimatology, involving applications to paleofire, landscape evolution, paleoecology, art history, and archeology. Because PaCTS is focused on paleoclimate, we re-used the relevant subset of the TRiDaS standard.

A discussion regarding paleoceanographic data standards was started during the Paleoclimate Model Intercomparison Project (PMIP) Ocean Workshop 2013 - Understanding Changes Since the Last Glacial Maximum (hereafter, PMIP LGM) in Corvallis, Oregon in December 2013. Given the expertise of the working group members, the discussion focused on marine sedimentary archives and was summarized into a document, which is available on the LinkedEarth Platform (Kucera et al., 2013). Their recommendations served as the foundation for a preliminary reporting standard for records based on marine sedimentary archives. Although the group identified recommended properties to be included with marine datasets, they did not propose a complete vocabulary nor a subset of required properties for acceptance in a database.

The Marine Annually Resolved Proxy Archives (MARPA) working group, nucleated under the EarthCube umbrella, is one of the first grassroots efforts within the paleoclimate community to enhance and facilitate the archiving and sharing of paleoclimate data as they pertain to annually resolved archives (e.g., corals, mollusks, coralline algae, and sclerosponges; Dassié et al., 2017). Their efforts included a registry of physical samples as well as their associated geochemical data and metadata, which are our primary focus here. The MARPA group summarized their recommendations in a document that was circulated among the community and constitutes the backbone of the recommendations presented here. Most of these recommendations were also applicable to other archives, rather than MARPA-specific, underscoring that despite their diversity, paleoclimate datasets retain common core properties that facilitate multi-proxy syntheses and comparisons.

The Speleothem Isotopes Synthesis and Analysis (SISAL) group was formed under the international Past Global Changes (PAGES) project and aimed at bringing together speleothem scientists, process modelers, statisticians, and climate modelers to develop a global synthesis of speleothem isotopes that can be used to further our understanding of past climate variability and in model evaluation. As part of this initiative, a template was created, outlining the necessary metadata for speleothem-based records (Atsawawaranunt et al., 2018). This template (Comas-Bru & Harrison, 2019) forms the backbone of properties applicable to speleothems-based records presented here.

2.3. Workshop on paleoclimate data standards

The workshop on paleoclimate data standards held in Boulder, USA in June 2016 (Emile-Geay

& McKay, 2016, Figure 1) served as a focal point to initiate a broader process of community engagement and feedback solicitation, with the goal of generating a community-vetted standard for reporting paleoclimate data. Workshop participants identified the necessity to distinguish a set of essential, recommended, and desired properties for each dataset. By default, any and all information was considered *desired*, though we shall see exceptions to this principle. A subset of the archived information should be *recommended* to ensure optimal reuse of the dataset. Yet a smaller subset of this information is defined as *essential*, meaning that the dataset cannot be reused reliably or at all without these critical pieces of information.

A consensus emerged that these distinctions are archive-specific; for instance, what is needed to meaningfully reuse MARPA records could be quite different from what is needed to meaningfully reuse an ice core dataset. It was therefore decided that experts on particular paleoclimate archives organized into working groups (WGs) would be best positioned to elaborate and discuss the components of a data standard for their specific sub-field of paleoclimatology. Consequently, seven WGs were created on the LinkedEarth Platform centered around the main archives used in paleoclimate studies: historical documents, ice cores, lake sediments, marine sediments, MARPA, speleothems, and tree rings. A call for additional WGs was made in the fall of 2016. Observations common to two or more archives (e.g., alkenones) were discussed in one WG with a link to the discussion in other WGs. It is also critical to ensure interoperability among standards to enable investigations using multiple observations on the same archive as well as across archives; to that end, three longitudinal WGs were created to deal with information common to all archives (such as publication, geographical coordinates, funding information), to report uncertainties in the record, and to report how chronologies were established.

The workshop participants also identified the need to have a separate set of requirements for newly generated datasets and legacy datasets, for which less metadata would likely be available. In PaCTS v1.0, a legacy dataset is defined as a dataset that is not being archived by the author(s) of the original study.

3. Towards PaCTS

3.1. Working groups

Rules of engagement on the LinkedEarth Platform were published in the fall of 2016 along with the establishment of seven WGs (ice cores, lake sediments, marine sediments, MARPA, speleothems, trees, and uncertainties, Figure 1). Three WGs (chronologies, cross-archive, and historical documents) followed in the spring of 2017 as additional archives and common information to all archives were identified. Each WG leader was tasked to organize their subcommunity either directly on the platform, through videoconferences, meetings at conferences, and/or other working groups (e.g., MARPA group and the PAGES SISAL group). The WG leaders were tasked to regularly update the discussion directly on the LinkedEarth platform or provide a document for integration on the platform. One difficulty in defining desired, essential, and recommended properties was related to the expected use of the data: depending on what one wants to do with the data, one needs different metadata. By far, the most important and metadata-hungry task is to perform queries to find datasets pertinent to a scientific question.

As an example of finding datasets pertinent to a scientific question, consider a study conducted by a paleoceanographer who wants to characterize millennial-scale sea surface temperature (SST) variability during the Holocene epoch (Khider et al, 2016). In the current research ecosystem, a typical workflow would consist of querying several databases to find suitable records, extract the data, consult the original publication(s) for additional metadata (e.g. author's definition of 'present'), reformat the data into a coherent format for analysis, apply spectral analysis to examine the frequency content of the records, perform some statistical analysis of the results, and visualize them. In an ideal world, the query, preferably from a single database, should (1) find records that span the Holocene, (2) find the subset of those that primarily reflect SST, and (3) find the subset of that subset with a specified resolution (e.g., finer than 200 years) to have at least five data points per 1,000-year cycle (a permissive assumption for this sort of work). Simple though it may seem, this query requires the following (meta)data: (1) a measure of age (time) and minimum and maximum values of the time series; (2) an estimate of SST, as an inferred variable, and/or Mg/Ca, $U^{k'}_{37}$, TEX_{86}, or microfossil assemblages as measured variables from which SST can be inferred; and (3) temporal resolution, calculated from the data.

Other types of basic queries include: searching for a particular publication, using either the digital object identifier (DOI), title, journal, or authors; and searching by the type of archives. Defining the search parameters for these complex queries on the LinkedEarth platform (Khider & Garijo, 2018) sparked the discussion for the needed properties.

A standard helps not only with the menial task of searching for records in a database. Such a standard can also assist with doing the science *per se*, by ensuring that the required information is present in the dataset. For instance, making a simple map of all the records in a database by archive types (Figure 1a of PAGES 2k Consortium, 2017) requires each dataset to report latitude, longitude, and the archive type. More complex data analysis requires more information: to investigate the effect of age uncertainties (e.g. with the Bchron (Haslett and Parnell, 2008) or BACON (Blaauw and Christen, 2011) packages), or to establish new depth-age models (Blois et al., 2011; Giesecke et al., 2014), one needs the raw radiocarbon measurements, their measurement uncertainties, and associated depth in the archive.

3.2. Community surveys

To decide which of the properties identified within the various WGs should be considered essential, recommended, or desired, we first gathered input via the LinkedEarth platform (Figure 2a). As of August 1st 2018, it was home to 207 polls, with 796 votes given by 32 different users. On average, each question received 3 votes, with some questions receiving no votes and others as many as 27. Note that some questions were duplicated across different WGs and the final count presented here takes into account all votes received on the platform. The low number of votes can be partially attributed to the fact that voting was only possible after authentication onto the platform, creating a barrier to widespread participation. To broaden community involvement, the polls were then threaded on Twitter from the LinkedEarth account with voting allowed over a seven-day period (Figure 2b). The Twitter polls increased engagement (by a factor of 3 on average), and also led to discussions that were then moved to the LinkedEarth platform for traceability of decisions.

Finally, by request from the community, the questions were summarized in a survey distributed to the paleoclimate community through the ISOGEOCHEM, CLIMLIST, paleoclimate and cryolist list-servs as well as the PAGES e-news, website, and social media. The survey contained

603 questions across all working groups for which respondents were asked to determine whether each property is deemed essential, recommended, or desired for new and legacy datasets, in addition to open-ended questions and prompts for community feedback. The survey was more comprehensive than the polls on the LinkedEarth platform or Twitter since all questions were framed to allow for a response for legacy and new datasets. On the other hand, the LinkedEarth platform also contains duplicate questions across various WGs (e.g., "should depth be reported as essential, recommended, desired), polls aiming to define the scope of the datasets housed on LinkedEarth (e.g., "should the LinkedEarth platform only contain datasets that appear in peer-reviewed publications?"), and the operating definition of legacy versus new datasets that was then used in the survey. Ninety-five scientists participated in the survey. Each question on the survey received on average 54 answers.

Paleoclimatology is a multi-disciplinary effort where researchers typically have expertise in one or more proxy systems (e.g., different observations on the same archive, similar observations on different archives, or a mix of different sensors, observations and archives). Scientists are often led to compare their own datasets to others obtained from proxy systems with which they are less familiar. Consequently, the metadata they need tend to differ based on their level of expertise (it is easier to "fill in the blanks" in one's own area of expertise). For instance, an ice core expert interested in comparing their deuterium record with a nearby record of SST would most likely only require the age at each horizon and associated SST. On the other hand, an expert on foraminiferal Mg/Ca-based SST reconstruction may also need information about the cleaning methodology or the number of individual foraminifera in the sample. To ensure that both needs were represented, respondents were encouraged to complete the entire survey, rather than focus exclusively on their own areas of expertise.

3.3. Survey responses

The 95 survey responses were then combined with the Twitter and LinkedEarth platform poll answers (Figures 3, 4 and Supplementary Information). In total, 135 participants from North America (52%), Europe (36%), Australia (5%), Asia (4%), South America (2%) and Africa (1%) were identified across the survey and LinkedEarth platform. Since voting on Twitter is anonymous, it is impossible to identify these voters or establish whether they voted on other platforms. We are aware that some researchers may have answered the same question several times on the various platforms. Since the number of survey answers dwarfs the number of votes on Twitter and the LinkedEarth platform (Supplementary Information) and Twitter does not track the user names associated with the votes, we did not attempt to correct for multiple responses. Therefore, 135 contributors represent our best estimate for the number of total participants.

Most of the polls on Twitter and the LinkedEarth platform referenced legacy versus new datasets. However, in the cases where the dataset status was not specified, we assumed that the question referred to a new dataset only. Furthermore, if a question was repeated on various WGs (e.g., latitude, longitude), the number of votes were tallied and included in the total count for the cross-archive metadata reporting (see Section 4.1). Responses on the survey, Twitter, and the LinkedEarth platform were given equal weight.

For each of the properties, we identified respondents' recommendation for both new and legacy datasets as the majority vote. We used mind maps to visually organize the hierarchical information, keeping the relationship intact (Figures 5) and mosaic plots to display the frequencies of the essential, recommended, and desired categories for each working group (Figure 6). Overall, the community identified 208 properties (69% of polled properties) as essential, 82 (27%) properties as recommended, and 12 (4%) as desired for new datasets. For legacy datasets, fewer properties were deemed essential: 131 (44%) of polled properties versus 136 properties (45%) were considered recommended and 34 properties (11%) were identified as desired. This difference is not unexpected and highlights the fact that legacy datasets, although not as metadata-rich as new datasets, are still valuable to the community (Figure 6).

4. PaCTS v1.0: Paleoclimate Community reporTing Standard

This section is based on the recommendations made in the various WGs, which were then subject to polling through the LinkedEarth platform, Twitter, and the survey. We are aware that these recommendations may be incomplete for some archives, a point discussed in Section 6. A list of these properties, definitions, and associated recommendations are available on the LinkedEarth platform.

4.1. Cross-Archive Metadata

Despite their diversity, paleoclimate records (and compilations thereof) share common metadata properties such as contributors, geographical information (e.g., coordinates, site name), publication information (e.g, authors, title, journal, DOI), funding information, and general information about the paleoenvironmental and chronology data (e.g., “should the raw data be included?”). In total, the community identified 54 properties applicable to all archives (Figures 5 and 7).

For new datasets, 36 of these properties were identified as essential, 9 as recommended and 9 as desired. It is not surprising that 67% of the properties were voted as essential since these properties are critical for the data reuse with no expert knowledge about the proxy systems or paleoclimate. Likewise, 24 of these properties (44%) were identified as essential for legacy datasets. For a dataset to be reused, information regarding the location, publication, and interpreted chronology and paleoenvironmental variables is critical. Hence, several researchers commented that new datasets should contain both the raw and interpreted data. The bar for legacy datasets should be lower, recognizing that much of the desired data may no longer be available, and that interpreted data are still useful for many applications.

In addition to the properties identified, a dataset DOI and a dataset license would also promote data reuse. LinkedEarth is not setup to mint DOIs directly but they can be obtained through other platforms such as PANGAEA, Dryad, or FigShare. The registry of research data repositories, re3data, gives information on whether a repository provides persistent identifiers. The Creative Commons (CC-BY) license is recommended for paleoclimate data since under this license, other researchers are free to share and adapt materials while giving appropriate credit to the original contributor of the resource.

4.2. Archive-specific metadata

4.2.1. Ice cores

The ice core WG identified 16 properties specific to glacier ice, including information pertaining to the archive, such as melt in transport, storage conditions, the observations available for the archive, and the chronology. For new datasets, eight properties were deemed essential and eight recommended. The number of essential properties dropped to four for legacy datasets with three properties deemed recommended (Figures 5, 6 and 8).

As with historical documents, most survey respondents were not experts on records generated on ice cores and therefore only responded for properties they were likely to use.

4.2.2. Lake Sediments

The lake sediments WG reported 54 properties specific to this archive, which were grouped by proxy sensor/observation types: particle size, mineralogy, imagery data, accumulation rate, and compound specific isotopes. Whereas some properties were common across the various types of observations (i.e., units, interpretation, pre-treatment methods), many were observation-specific (e.g., source of compound for compound-specific isotopes), highlighting the necessity of detailed sets of guidelines down to the proxy observation level to meet researchers' needs.

For new datasets, 39 properties were identified as essential and 15 as recommended. For legacy datasets, 25 were seen as essential, 28 as recommended, and 1 as desired (Figures 5, 6, and 9). In addition to these 54 properties, the WG started a discussion on how to best report the concept of depth in the archive. Although several WGs identified depth (i.e., position in the archive sample) as an essential property, especially for new datasets, none had defined how this depth should be reported. The majority of the respondents indicated a preference to report top and bottom depth for both new and legacy datasets although several respondents proposed to lower the bar for legacy datasets to whatever is available for these records.

Respondents also noted that pictures of the core after the sampling process would be useful. Whether these pictures should be available with the data or stored in the database of the physical sample repository is a decision best left to individual researchers, based on their constraints and mandates by funding entities.

4.2.3. Marine sediments

The marine sediments WG identified 48 properties specific to this type of archives. These properties were divided into 6 groups, according to the type of observation: general sampling, bulk sediment geochemistry, foraminifera geochemistry, alkenones, the glycerol dialkyl glycerol tetraether (GDGT) proxies, and micropaleontology. The foraminifera geochemistry category was further subdivided into stable isotopes, boron isotopes, and trace elements. Although some of the requirements were common to all observations, this WG included several observation-specific properties such as the cleaning methodology for foraminiferal trace elements or raw peak areas for GDGTs.

For new datasets, 36 properties were identified as essential and 12 as recommended. The number of essential properties drops to 24 for legacy datasets, with the remainder considered recommended (Figures 5, 6 and 10).

4.2.4. Coral, mollusks, and other annually resolved marine records

The properties for these archives were taken from the spreadsheet the MARPA group had circulated online for feedback. Most of these properties were applicable to all archives reporting geochemical properties and were therefore incorporated into the cross-archive WG and questions. Two archive-specific properties were also identified: interpolated chronologies (i.e., distance from core top translated to time usually a calendar day for each sample then interpolated to even monthly intervals) and X-ray pictures (and associated drilling path). For both new and legacy datasets, the raw (distance from core top), interpolated chronologies, and X-ray pictures were considered essential and recommended, respectively (Figure 5 and 6). The reporting of growth increments in mollusks and corals is still an ongoing discussion within MARPA.

4.2.5. Speleothems

When constructing their database (Atsawawaranunt et al., 2018), the SISAL WG identified 23 properties specific to speleothem records. The SISAL database only focuses on stable isotopes in speleothems and these properties only apply to this proxy system. These properties can be further subdivided into four categories describing the cave and modern cave conditions, the physical sample, and information about the sample data. For new datasets, 11 properties were considered essential and 12 recommended. For legacy datasets, only 2 properties were considered essential and 21 were marked as recommended (Figures 5, 6 and 11).

Although “evidence for equilibrium” (e.g., the Hendy test; Hendy, 1971, or monitoring data that supports equilibrium precipitation of calcite) was narrowly voted as essential for new datasets and recommended for legacy datasets, three respondents (two on Twitter and one on the survey) expressed concerns about the value of this property as it “rarely shows up in monitoring data” and the Hendy test has been “abused” by the paleoclimate community. This illustrates the need for an evolving standard, one that fits the needs of the community and changes as our scientific understanding about proxy systems increases.

4.2.6. Tree-based records

The tree ring community has a long history of developing and adopting data standards; however, the metadata capacity or requirements in earlier data formats (e.g., Tucson, Heidelberg, Sheffield, CATRAS and Belfast amongst many others) were limited by the technology of the decade in which they were created (Brewer et al. 2011). The 35 properties in the survey were taken from TRiDaS (Jansma et al., 2010) and from the proposed tree-ring isotope databank (Csank, 2009). TRiDaS was chosen as a starting point as it was designed as a standard to represent dendrochronological data across its many subdisciplines, including dendroclimatology. TRiDaS therefore includes many (optional) properties as essential or recommended that are not applicable to datasets collected for paleoclimate reconstructions.

For new datasets, 26 properties were considered essential, 7 recommended, and 2 desired. For legacy datasets, 19 properties were voted on as essential, 9 as recommended, and 7 as desired (Figures 5, 6, and 12). Several researchers were confused about the terms used in TRiDaS, suggesting that the standard may be too broad for most paleoclimate applications and should be further refined if it is to be widely adopted. The reason for this confusion may be because TRiDaS was initiated by the cultural dendrochronology community (e.g., dendroarcheology, art and building history) in a response to the more pressing need for standardized metadata in these disciplines. Despite attempts to engage all subdisciplines of dendrochronology in the development of TRiDaS, the cultural aspects of the standard were more fully implemented due to the greater participation of users from these areas of research.

Nevertheless, a subset of the fields defined in TRiDaS were used as a starting point for discussion for PaCTS v1.0. Many fields within TRiDaS are already addressed in the cross-archive metadata and were disregarded, leaving only dendro-specific fields. These were then supplemented by fields for tree-ring isotope data taken from the tree-ring isotope databank proposed by Csank (2009). Regretfully, discussion of the suitability of these fields among the dendroclimatology community has been limited and the list of initial fields was not subsequently refined. The public voting process has resulted in a number of fields being marked as 'essential' that are not routinely (if ever) collected for dendroclimatological research. Furthermore, some of the quantities that are being proposed are difficult to measure or know, raising the issue of whether these properties are even desired. Some of the properties are a characteristic of the data themselves ('ring count') and not metadata *per se*. These may be useful as convenience fields when querying large data collections (rather than having to extract and calculate).

The confusion in the voting process could reflect confusion over whether PaCTS v1.0 is to be a data standard applicable to all dendrochronological datasets or exclusively to those collected for use in climate reconstructions, for which a smaller number of 'essential' fields would be required. It could also reflect sampling bias in the voting process related to the composition of the WG.

While the work described here is clearly an important step towards incorporating dendroclimatological data into a universally applicable paleoclimate data standard, there remains a great deal of work to be done. This work needs to begin with discussions that engage a much broader cross-section of the dendroclimatological community and refined criteria in subsequent surveys.

4.2.7. Documentary archives

Historical documents differ quite significantly from the other archive types presented in PaCTS v1.0. Documentary data are extracted from written sources (books, chronicles, newspaper, etc) and each of these sources in the dataset needs a reference to the publication metadata (in addition to the scientific publication of the data in a journal). The raw data most comparable to measurements on other archives are quotes, i.e., text strings in any language cited from the source from which location, time, and event are extracted. Every single data point in the set can thereby have a different location and a variety of parameters describing the event (Glaser, 1996). The time step can be, but is not necessarily, periodic. The quote might contain information regarding the temperature in a city, precipitation conditions, and the resulting water level in a river, as well as statements concerning harvest amount and quality of a certain crop. The resulting data type can be boolean (for presence/absence), integer (for indices), real numbers with units for measurements, or enumerations (Riemann et al., 2016).

The documentary archives WG identified nine properties which concerned the source material, including original scans of the documents, quote ID, language, and reference to the source material (e.g., DOI, license, page). Among these nine archive-specific properties, four (the quote, reference to the quote, the quote ID and the quote's DOI) were voted as essential and five as recommended for new datasets. For legacy datasets, only two (the quote and its reference) were identified as essential (Figures 5, 6 and 13). Four survey respondents indicated that they were least familiar with this type of archive, which may help explain why fewer properties compared to other archives were considered essential for optimal reuse of the resource by researchers not familiar with the intrinsic details of the archive.

4.3. Uncertainties

The Uncertainties WG identified seven properties applicable to most records. These properties fell into two broad categories concerning the uncertainty in the measured variable (analytical uncertainty, number of repeat measurements, and reproducibility) and the uncertainty associated with models to infer variables, including chronologies (output statistics, output ensembles along with the parameters and the publication in which the model is described). For new datasets, four properties (analytical uncertainty, number of repeat measurements, the publication and parameters of the model) were deemed essential and the other three recommended. For legacy datasets, only one was deemed essential (number of repeat measurements) while the rest were recommended. This highlights the commitment of the community to better characterize uncertainties in paleoclimate records and the acknowledgement that uncertainty has often been ignored when reporting datasets in the past, making it difficult to include metadata for legacy datasets (Figures 5, 6, and 14).

Respondents voted on reporting the analytical uncertainty and reproducibility as “2-sigma” (estimated as the standard error of the mean), although a point was raised that the reporting should be community-specific, following their own accepted standards (e.g., radiocarbon, Stuiver et al., 1977, Millard et al., 2014), but clearly indicated in the metadata. A compromise is to keep community-specific standards while encouraging 2-sigma reporting if there is no preexisting standard.

For models, the method used should be documented both in the papers and with the data, with publication information about the software and parameters used being considered essential for new datasets. For legacy datasets, all information about the model is considered recommended.

The Uncertainties WG has barely scratched the surface of uncertainty reporting in paleoclimate studies. Although several other WGs have reported that uncertainty should be an essential parameter, there is not yet a clear path forward as to how this uncertainty should be unambiguously reported. However, there is some consensus that the method of reporting does not matter as long as the method is clearly described. To do so, the LinkedEarth Ontology (Emile-Geay et al., 2019) offers several paths forward. The class "Uncertainty" can refer to a single value for all the data values, to a list of values of equal length as the uncertain variable, and to models output stored in ensemble, summary, and distribution tables.

Consider the example of radiocarbon dating. Each radiocarbon value is associated with an uncertainty that is often reported in a separate column of the measurement table. This radiocarbon-age uncertainty is then translated (via a calibration curve) into a calendar age uncertainty that is also stored in a separate column. In both of these cases, the uncertainty is a variable that can be described with the same richness as other columns in the data table. Furthermore, probabilistic age modeling software such as Bchron (Haslett and Parnell, 2008) and BACON (Blaauw et al., 2011) for radiocarbon, HMM-Match (Lin et al., 2014) for stratigraphic alignments, and the Banded Age Model (Comboul et al., 2014) return possible age distributions around the calendar age value as well as age model ensembles for each depth in the paleorecord. In this particular example, each measured value has at least one associated uncertainty value, possibly an entire probability distribution.

On the other hand, uncertainty associated with measurements of trace elements and stable isotopes is often reported as the uncertainty of the standard or a handful of replicates that are taken to represent the uncertainty for all values. The LinkedEarth Ontology (Emile-Geay et al., 2019) allows for the specification of not only the values and units of the uncertainty, but also how this uncertainty is estimated and the level at which it is being reported (e.g., one standard error of the mean).

4.4. Chronologies

The Chronologies WG identified 54 properties, 43 of which were deemed essential for new datasets, 10 recommended and 1 desired. For legacy datasets, 30 were identified as essential, 22 as recommended, and 2 as desired (Figures 5, 6 and 15).

Chronologies are obtained using two methods: absolute and relative. Relative chronologies often involve the alignment of one paleoclimate time series with another of known age. For instance, benthic foraminifera stable oxygen isotope ($\delta^{18}O$) records have often been aligned to the dated LR04 benthic $\delta^{18}O$ stack (Lisiecki and Raymo, 2005). For this type of chronology, the original measurements (e.g., benthic foraminifera $\delta^{18}O$), the alignment target (e.g., LR04 benthic $\delta^{18}O$ stack), its associated reference chronology (e.g., LR04 age model) and alignment method (e.g., HMM-Match (Lin et al., 2014)) should be clearly identified (*essential*) for both new and legacy datasets. We acknowledge that there is potentially more work to be done to devise a standard for relative chronologies, which should include an integration framework for biostratigraphy, paleomagnetism, stable isotopes chronologies, and orbitally-tuned chronologies.

Absolute chronologies are based on radiometric measurements (commonly radiocarbon, lead, and uranium-decay series, or terrestrial cosmogenic nuclide), layer-counting, counting of annual cycles in geochemical/isotopic proxies, dendro- or tephrochronological crossdating, or luminescence. In addition, some records are characterized by floating chronologies that are absolutely dated (within the uncertainty of the radiometrically derived age), but which have a precise internal chronology due to clear annual banding/cycles (e.g., U-series dated fossil corals, radiocarbon-dated tree chronologies).

The radiocarbon community has a long history of standardizing the reporting of their measurements. In 1977, Stuiver and Polach highlighted recommendations that have remained mostly unchanged (Stuiver and Polach, 1977). For chronological studies using the Libby half-life (Libby et al., 1949), Stuiver and Polach recommend reporting the $\delta^{13}C$ ratio, the conventional radiocarbon age (relative to CE 1950), associated error (expressed as ± one standard deviation), the estimated reservoir correction, and (optionally) the per mil depletion or enrichment with respect to 0.95 NBS Oxalic acid standard (Olson, 1970). For geochemical samples, dendrochronological samples, reservoir equilibria, and diffusion models, they recommend reporting the $\delta^{13}C$ ratio, percent modern, and $\delta^{14}C$ and $\Delta^{14}C$ based on the Cambridge half-life of 5730 years (Godwin, 1962). These guidelines were further extended to include post-bomb ^{14}C data (Reimer et al., 2004) and the reporting of calibrated dates (Millard, 2014) and formed the basis of the properties that were put to a vote. Given the long history of standardization, it is not surprising that legacy radiocarbon datasets are also held at a stringent reporting level.

For U-Th dating, the WG recommended the use of the standard proposed by Dutton et al. (2017), with most properties recognized as essential when reporting U-series dates.

Survey respondents also defined what information should be included when reporting the use of age modeling software. The method's name is deemed *essential* for both legacy and new datasets with most of the other properties identified as recommended. In addition, there is interest in storing ensembles of posterior draws from Bayesian approaches to ensure that the study is fully reproducible. The LiPD structure is already setup to handle multiple model output instances, allowing updates of chronologies for legacy datasets when raw data are available. They thus provide a natural container to store this information.

Finally, respondents were asked to define some nomenclature, including the use of "present" in paleoclimate studies. Over 80% of respondents voted on keeping the concepts of age and year separated. Age is represented on a time axis starting from the "present" and counting positively back in time. On the other hand, "year" follows the Gregorian calendar and is particularly useful for studies concentrating on the past 2,000 years. Over 60% of respondents also voted on reporting years relative to CE (Common Era) rather than AD.

Asking for a definition of "present" yielded diverse results. Sixty-eight percent of respondents voted in favor of using 1950 as the present, following the radiocarbon convention, 7% voted in favor as defining present as the last year in a record (with no mention of uncertainty), 12% voted in favor of using 2000 as the present, while the last 13% answered "other " This last category includes the use of 1950 for radiocarbon and either something else for the other chronologies or readjusting to 1950 to stay in tune with radiocarbon and the use of either 1950 or 2000 as long as it is clearly defined with the data. In summary, there is a consensus that "present" should be defined as an absolute date (and reported in the metadata), but it should be archive-dependent, with practitioners of U-series dating leaning towards CE 2000 and practitioners of radiocarbon dating leaning towards CE 1950.

One issue in reporting ages is, again, the lack of standards. The most common standard for time and date reporting (e.g., ISO 8601) does not accommodate for geologic time. The more recent OWL time ontology draws on the work of Cox and Richard (2015) and includes these concepts. However, these authors offer no finer division of geologic time than eras. This means that the vast majority of archived paleoclimate datasets (particularly, the totality of datasets archived on the LinkedEarth platform) would represent a single time point (the Quaternary era). To remedy this gap between ISO 8601 and the OWL time representation, we hereby propose a precise mechanism to report the time axis in paleoclimate datasets:

$$\text{Time (age)} = \textbf{significand}\ .\ 10^{\textbf{exponent}}\ \text{years}\ \textbf{direction datum}$$

Where “**significand**” and “**exponent**” are components of standard floating-point representation; “**direction”** indicates whether time flows forward (since a datum, as in the case of AD dates), or backwards (before a particular datum, as in the case of ages). “**Datum”** here refers to the origin point of the time (age) axis, which is arbitrary and (as recounted by Wolff, 2007) highly inconsistent among researchers.

Table 1 shows how this representation would work in practice. Note that variability in the datum for rows 1 (21 ky BP, a common date for the Last Glacial Maximum) and 4 (127 ky BP, a common date for Marine Isotope Stage 5e) could arise because of the date being reported from a radiocarbon vs. U-series chronology, and is usually impossible to infer without clarification from the original publication, or from its authors. The current proposal removes such ambiguities and can accommodate both observed and simulated datasets, potentially easing the task of model-data comparison if both communities start adopting it.

5. An example: MD98-2181

This section puts these recommendations into practice on a real-world dataset: the MD98-2181 marine sedimentary record from Khider et al. (2014). The purpose is twofold: (1) illustrate how to implement these recommendations in practice and (2) draw attention to practical difficulties that may impede large-scale adoption of PaCTS v1.0.

MD98-2181 is the most metadata-rich dataset currently available on the LinkedEarth platform since it was used as an example to further develop the LiPD framework and later the LinkedEarth Ontology. The dataset consists of measurements of Mg/Ca and $\delta^{18}O$ made on the planktic foraminifera *Globigerinoides ruber* (white, *sensu stricto* and *lato*) and $\delta^{18}O$ made on the benthic foraminifera *Cibicidoides mundulus* to infer surface and deep ocean variability in the western tropical Pacific over the Holocene. The age model is based on radiocarbon measurements for the Holocene and deglacial portion of the core.

Using the standards proposed for cross-archive metadata, Mg/Ca and $\delta^{18}O$ on foraminifera, radiocarbon-based chronology, and uncertainties, we calculated how many metadata properties in the essential and recommended categories were present in the MD98-2181 datasets (Figure

16). Since, by default, all metadata are desired, we ignored this category for the purpose of this example. In terms of its cross-archive metadata, the MD98-2181 record is nearly complete, with 95% of the essential metadata and 78% of the recommended metadata present in the record (Figure 16). The only missing component of essential metadata is the sample thickness. For the recommended category, the International Geo Sample Number (IGSN) for the sample and date at which the measurements were performed (i.e., analysis date) are missing. The core IGSN should be assigned by the core repository directly (e.g., Bremen Core Repository, Oregon State University core repository). Both analysis dates and sample thickness are metadata readily available at the time of collection. Although both were collected in either a physical notebook or by the instrument during analysis, they were not archived with the dataset on LinkedEarth since the information was not deemed by the metadata authors as essential for reproducibility.

The paleodata for the record consists of Mg/Ca and $\delta^{18}O$ measurements on foraminifera tests from sediment core subsamples. For the essential reporting of $\delta^{18}O$ on foraminifera, the MD98-2181 record lacks metadata regarding the taxonomy scheme being followed and equilibrium offsets. In the recommended category, only the volume of sediment analyzed is missing. For Mg/Ca reporting, the contamination indicator values (Mn/Ca and Fe/Ca; Khider et al., 2014) are missing from the archived record in addition to the taxonomy scheme being followed. Neither were deemed useful for reproducibility by the authors of the study at the time of reporting. In the recommended category, the volume of sediment analyzed and habitat depth have not been reported. In both cases, the values are unknown, either because they were not measured during sample preparation (sediment analyzed) or could not be accurately determined (habitat depth) from previous studies in the region.

The MD98-2181 chronology was based on radiocarbon measurements. Ninety percent of the raw radiocarbon dates used in Khider et al. (2014) were reported in Stott et al. (2004) and Stott et al. (2007). The raw data necessary for the repeatability and replicability of the age model in Khider et al. (2014) were re-reported in the later study. However, the archived record is missing information about the modern fraction (F14C), the sample ID, and the matrix, which are deemed essential. The archived record is also missing most of the recommended properties, only reporting the reservoir age correction (ΔR), the ensemble statistics, and the ensemble age models. The last two properties are essential in the context of the Khider et al. (2014) study to reproduce the age-uncertain spectral analysis. The Stott et al. (2004) and Stott et al. (2007) studies are also missing the essential and recommended properties with respect to reporting of raw measurements.

For uncertainty quantification, the record metadata lack the number of repeated measurements and the model parameters in the essential category, though it should be noted that the values of repeated measurements are reported in the measurement table itself. The record is complete in the recommended category.

This example highlights the difficulty of reporting all essential metadata, especially after the study has been completed. We therefore present version 1.0 of PaCTS as an aspirational standard, one that would theoretically ensure optimal reuse of paleoclimate datasets but is difficult to observe in practice. Clearly, being aware of these requirements at the start of a study would help scientists keep track of the necessary metadata and ensure that they are reported when the dataset is digitally published (e.g., on WDS-Paleo or PANGAEA). We therefore recommend that investigators plan ahead of time which properties they intend to report, and structure their lab notebooks so this information is easier to track at the time of publication.

6. Discussion

This paper describes the first effort by the global paleoclimate community to define standards for digitally archiving paleoclimate datasets. Such standards aim to make publicly archived paleoclimate data more re-usable by clearly describing them with comprehensive metadata. In combination with the LinkedEarth Ontology, these standards also help meet the interoperability principle by using a formal, accessible, shared, and broadly applicable language for knowledge representation. If the datasets are properly described using micro-data (e.g., Schema.org), they are also findable. Together, these standards bring such datasets closer to compliance with "FAIR" principles.

The standards arose through collective discussions, both in-person and online, and via an innovative social platform (Gil et al 2017). The results of this collective decision-making reveal an evident desire for archiving a rich set of metadata properties, with respondents identifying roughly two thirds of properties (208 out of 302) as *essential* for new datasets. Respondents also recognized that legacy datasets may not be as complete, so they identified less stringent requirements in order not to overlook valuable datasets. Nonetheless, respondents identified 131 properties as *essential* for legacy datasets, highlighting the fact that a dataset loses its usefulness if too many requirements are not met. Several respondents also indicated that, while some properties should theoretically be *essential* (or *recommended*), they may be hard to obtain in practice and/or variable in time. These include seasonality and habitat depth of foraminifera and many of the properties from TRiDaS. Furthermore, although rich metadata are always valuable, these requirements should be balanced with the researcher's time. Scans of historical documents or uploads of x-radiographs of archive samples would be highly valuable to the community, but these activities are time-consuming and this use of time is rarely, if ever, incentivized by funding agencies.

PaCTS v1.0 is also missing several proxy systems, including loess and continental records, faunal and floral counts in lake sediments and does not incorporate recent standards such as the one developed by Courtney Mustaphi et al. (2019) for ^{210}Pb dating. Finally, although cross-pollination was encouraged, common properties were not adequately identified across WGs, resulting in duplicates. This is especially apparent in the lake and marine sediment WGs.

Another salient outcome is that this first version of PaCTS can only be described as aspirational. Indeed, section 5 illustrates that even in the best of circumstances (the author describing their own dataset, generated less than a decade ago), the compliance rate was far from perfect. This points to the need for more realistic guidelines. It is indeed apparent that many participants misinterpreted what was meant by "essential." Further, the participation rate is still far below what is needed for this standard to be representative of the worldwide paleoclimate community, which would gain much from harmonization. How can this standard be collectively refined and more broadly adopted? How should the standard, and its future versions, be implemented in practice?

6.1 Broadening participation

The genesis of PacTS v1.0 serves as a useful template for future efforts. As detailed in section 2, the spark for the discussion came from the 2016 workshop on Paleo Data Standards. Nothing replaces the immediacy of in-person communication for this sort of work. However, it would be costly, carbon-intensive and unrealistic to expect large segments of the paleoclimate community to travel for such an event, should it happen again. We therefore advocate that further discussion take place within, or around, existing meetings. Examples include the annual meetings of the American Geophysical Union and the European Geosciences Union, the Goldschmidt conference, Ocean Sciences meeting, the PAGES Open Science Meeting, the International Conference on Paleoceanography, meetings of the International Union for Quaternary Research, as well as more focused meetings like WorldDendro, Karst Record, or the ASLO Aquatic Sciences Meeting. We have also found PAGES-sponsored workshops to be excellent opportunities to discuss data stewardship considerations, of which reporting standards are an important aspect. At the very least, an annual session at an international meeting would be useful for the community to touch base and take stock of progress and challenges, but more frequent interactions will be desirable until adoption reaches a critical threshold (e.g., 80% of submissions to public repositories like WDS-Paleo or PANGAEA).

Assuming such meetings will take place over the next few years in many corners of the community, there is still a need for more sustained forms of communication. The virtual working groups on the LinkedEarth platform is where many of our discussions took place, and they remain available to complement to in-person discussions. Membership is open, and we encourage interested readers to join LinkedEarth so they can participate in these forums or create their own forums on a platform of their choice (traceability and transparency being of paramount importance).

6.2 Roadmap to standardization

In practical terms, we recommend that the next iteration of PaCTS use the following steps:

(1) The procedure for ratification is developed in tandem with major stakeholders (scientific societies, data repositories, chief editors).

(2) The proposed procedure is widely distributed to the community (e.g., through the PAGES magazine, AGU and EGU communication channels, social media).

(3) The timeline for discussion and voting is clearly indicated, and voting occurs on the LinkedEarth platform.

(4) The vote outcome is presented at a major international meeting and any additional discussion is considered before the vote is certified at the meeting.

(5) The standard is widely disseminated and encouraged by appropriate incentives (see below).

6.3 Implementing Emerging Standards

We envision two main ways to encourage the adoption of the standard. The first is to use technical innovation to lower the barrier to metadata archiving; the second is to change the incentive structure to make it worthwhile for researchers to adopt the standard, despite the inevitable opportunity cost that comes with providing more complete data records.

On the first point, the LinkedEarth project has recently implemented a web interface to convert paleoclimate datasets into the LiPD format: the lipd.net "playground" (http://lipd.net/playground). To promote standardization, the reporting recommendations described herein will be flagged as users create LiPD files interactively on the lipd.net website, pulling data and metadata from native archival formats (e.g., Excel spreadsheets). Ideally, all records, especially those accepted on the LinkedEarth platform, will show their compliance rate with PaCTS. This rate can be computed during creation of the LiPD file, allowing “unavailable” as an answer for the essential fields. At present, the lipd.net playground displays the rate of required fields that have been entered, but is not set up to track archive or proxy-specific completeness, although this is possible with further development. The “unavailable” category serves two purposes: (1) to encourage researchers to gather these metadata during their next study and (2) to investigate how many of these essential properties are reported in practice. Alternatively, LinkedEarth could appoint a Board of Data Editors to approve the datasets for upload onto the platform. The Board presents several advantages over an automatic process: (1) to answer specific questions, therefore taking into consideration the intricacies of a dataset; (2) to identify needed changes to the reporting standards faster; and (3) to assist the community with the online web service when needed. The major drawback is the volunteer time of the Board of Data Editors. In our experience, the time of researchers is already stretched thin, and they have little incentive to commit more of it to the relatively thankless task of standardization.

How might the reward structure be changed? There are essentially two levers to activate. The first is funding agencies. In the United States, for instance, the National Science Foundation funds the vast majority of paleoclimate research. While the agency now requires a data management plan to be submitted for each proposal, its reporting guidelines are very broad. They could be made more specific, and point paleoclimate researchers to the latest version of PaCTS.

The European Research Council similarly supports Open Science, but with far less specific guidelines than PaCTS v1.0. To the best of our knowledge, the situation is similar for other countries (e.g., Canada, Australia). We therefore call on funding agencies to either endorse this standard or propose a meaningful alternative.

The second lever is publishers and editors: while each publishing house encourages digital data archiving to varying degrees, the decision of what (meta)data to include is ultimately up to the author, and often fails to consider the long-term value proposition of the dataset. Publishers could help ensure that the present standard is, at the very least, encouraged, if not mandatory. In particular, the American Geophysical Union and Copernicus publishers recently endorsed requirements to make data FAIR. Affiliated journals could use their leverage to promote more stringent reporting standards. As an example, the recent PAGES 2k special issue of the journal *Climate of the Past* piloted the implementation of open-data practices, which included some reporting standards, and reported the challenges faced when requiring such practices (Kaufman et al., 2018). Another avenue for promoting best practices, including adoption of reporting standards, is through professional paleoscience organizations such as PAGES and INQUA.

We expect the present reporting standard to evolve to meet the needs of the paleoclimate community. It is our hope that this publication will stimulate volunteers to join the effort and organize discussions at all community levels; there can be no community standard without community involvement. We are confident that improving paleoclimate data standards will promote collaboration on international data syntheses and encourage the development of software based on the new standards. In turn, such software will reduce the time to science, by compressing the time researchers spend on the menial task of data wrangling.

Acknowledgments, Samples, and Data

Code and data to reproduce the figures of this article are available on GitHub and released on Zenodo (doi:10.5281/zenodo.3165019). Definition of properties and recommendations are summarized here: http://wiki.linked.earth/PaCTS_v1.0. This work was supported by the National Science Foundation through the EarthCube Program with grant ICER-1541029. Feedback solicitation on the standard was facilitated by the Past Global Changes (PAGES) organization. The 2016 workshop on Paleoclimate Data Standards was hosted by the World Data Service for Paleoclimatology (WDS/NOAA-Paleo), and the participation of international attendees was made possible by a PAGES travel grant. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Atsawawaranunt, K., et al. (2018), The SISAL database: a global resource to document oxygen and carbon isotope records from speleothems, Earth System Science Data, 10(3), 1687-1713, doi: 10.5194/essd-10-1687-2018

Blaauw, M., and J. A. Christen (2011), Flexible Paleoclimate Age-Depth Models using an Autoregressive Gamma Process, Bayesian Analysis, 6(3), 457-474, doi: doi:10.1214/11-BA618

Blois, J. L., Williams, J. W. (Jack), Grimm, E. C., Jackson, S. T., & Graham, R. W. (2011). A methodological framework for assessing and reducing temporal uncertainty in paleovegetation mapping from late-Quaternary pollen records. *Quaternary Science Reviews*, *30*(15), 1926–1939. doi:10.1016/j.quascirev.2011.04.017

Bradley, E., K. Anderson, L. de Vesine, T. Nelson, S. Soti, I. Weiss, and R. Yadav (2018), CSciBox – building age models of paleorecords, Zenodo, doi:10.5281/zenodo.1245175

Brewer, P.W., Murphy, D. and Jansma, E., 2011. TRiCYCLE: a universal conversion tool for digital tree-ring data. Tree-Ring Research, 67(2), pp.135-145. DOI: 10.3959/2010-12.1

Comas-Bru, L. and Harrison S.P. (2019), SISAL: Bringing added value to speleothem research, Quaternary, 2(1), 7; doi: 10.3390/quat2010007

Courtney Mustaphi, C. J., Brahney, J., Aquino-López, M. A., Goring, S., Orton, K., Noronha, A., et al. (2019). Guidelines for reporting and archiving 210Pb sediment chronologies to improve fidelity and extend data lifecycle. Quaternary Geochronology, 52, 77-87, doi:10.1016/j.quageo.2019.04.003

Cox, S. J. D., and S. M. Richards (2015), A geologic timescale ontology and service, Earth Science Informatics, 8(1), 5-19, doi: 10.1007/s12145-014-0170-6

Csank, A.Z. (2009), An International Tree-Ring Isotope Data bank—A proposed repository for tree-ring isotopic data, Tree-Ring Research 65(2),163- 164, doi:10.3959/1536-1098-65.2.163

Dassié, E.P. , et al. (2017), Saving our marine archives, EOS, 98, doi: 10.1029/2017EO068159

Dasu, T., and T. Johnson (2003), Exploratory Data Mining and Data Cleaning, 203 pp., Wiley

DCMI Usage Board, 2008. Dublin Core Metadata Initiative (DCMI)metadata terms. Retrieved on Ausgust 6th 2019 at http://dublincore.org/documents/dcmi-terms/

Dutton, A., Rubin, K., McLean, N., Bowring, J., Bard, E., Edwards, R.L., Henderson, G.M., Reid, M.R., Richards, D.A., Sims, K.W.W., Walker, J.D., Yokoyama, Y. (2017) Data reporting standards for publication of U-series data for geochronology and timescale assessment in the earth sciences. Quat. Geochron., 39:142-149, doi:10.106/j.quageo.2017.03.001

EarthCube Technology and Architecture Committee Standards Working Group: Report of the EarthCube Standards Working Group, finalized 10/05/2015. Accessed online on 08/13/2018 at https://www.earthcube.org/document/2015/ecstandardsrecs

Emile-Geay, J., and J. A. Eshleman (2013), Toward a semantic web of paleoclimatology, Geochemistry, Geophysics, Geosystems, 14(2), 457-469, doi: 10.1002/ggge.20067

Emile-Geay, J., and N. P. McKay (2016), Paleoclimate data standards, Past Global Change Magazine, 24(1), doi: 10.22498/pages.24.1.47

Emile-Geay, J., D. Khider, D. Garijo, N. P. McKay, Y. Gil, V. Ratnakar, and E. Bradley (2019), The Linked Earth Ontology: A Modular, Extensible Representation of Open Paleoclimate Data, Zenodo. http://doi.org/10.5281/zenodo.2577604

Giesecke, T., Davis, B., Brewer, S., Finsinger, W., Wolters, S., Blaauw, M., De Beaulieu, J.-L., Binney, H., Fyfe, R.M., Gaillard, M.-J., Gil-Romera, G., Knaap, W.O., Kuneš, P., Kühl, N., Leeuwen, J.F.N., Leydet, M., Lotter, A.F., Ortu, E., Semmler, M., Bradshaw, R.H.W. (2014), Towards mapping the late Quaternary vegetation change of Europe. Vegetation History and Archaeobotany 23, 75–86. doi:10.1007/s00334-012-0390-y

Gil, Y. (2013). Social Knowledge Collection. In P. Michelucci (Ed.), *Handbook of Human Computation* (pp. 285-296): Springer.

Gil, Y., D. Garijo, V. Ratnakar, D. Khider, J. Emile-Geay, and N. P. McKay (2017), A Controlled Crowdsourcing Approach for Practical Ontology Extensions and Metadata Annotations, in The Semantic Web - ISWC 2017. ISWC 2107. Lecture Notes in Computer Science, edited by C. e. a. d'Amato, pp. 231-246, Springer, Cham.

Glaser, R. (1996). Data and Methods of Climatological Evaluation in Historical Climatology HSR Historical Social Research, 21 (4) : 56-88.

Godwin, H. (1962), Half-life of radiocarbon, Nature, 195, 984, doi: 10.1038/195984a0

Gregory, J. (2003), The CF metadata standard, Retrieved from http://cfconventions.org/Data/cf-documents/overview/article.pdf on May 28th 2019.

Haslett, J., and A. Parnell (2008), A simple monotone process with application to radiocarbon-dated depth chronologies, Journal of the Royal Statistical Society C, 57, 399-418, doi: 10.1111/j.1467-9876.2008.00623.x

Heath, T., Bizer, C., (2011), Linked Data: Evolving the Web into a Global Data Space (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.

Heiser, C, McKay, N., Emile-Geay, J., Khider, D. (2018). LiPD-utilities (Version 1.0.0). Zenodo. doi:10.5281/zenodo.60813.

Hendy, C. H. (1971), The isotopic geochemistry of speleothems-I: The calculation of the effects of different modes of formation on the isotopic composition of speleothems and their applicability as paleoclimate indicators, Geochimica and Cosmochimica Acta, 35, 801-824

Jansma, E., P. W. Brewer, and I. Zandhuis (2010), TRiDaS 1.1: The tree-ring data standard, Dendrochronologia, 28(2), 99-130, doi: 10.1016/j.dendro.2009.06.009

Kaufman, D.S., PAGES 2k Special Issue Editorial Team, 2018. Technical Note: Open-paleo-data implementation pilot – The PAGES 2k special issue. Climate of the Past 14, 593-600. doi: 10.5194/cp-14-593-2018

Khider, D., C. S. Jackson, and L. D. Stott (2014), Assessing millennial-scale variability during the Holocene: A perspective from the western tropical Pacific, Paleoceanography, 29(3), 143-159, doi: 10.1002/2013pa002534

Khider, D., Emile-Geay, J., McKay, N. P., Jackson, C., & Rouston, C. (2016). Testing the millennial-scale Holocene solar-climate connection in the Indo-Pacific Warm Pool. Paper presented at the American Geophysical Union Fall Meeting, San Francisco, CA.

Khider, D., F. Zhu, J. Hu, and J. Emile-Geay (2018a), LinkedEarth/Pyleoclim util: Pyleoclim release v0.4.0, Zenodo, doi:10.5281/zenodo.1205662

Khider, D., and D. Garijo (2018b), LinkedEarth Queries, edited, Zenodo, doi:10.5281/zenodo.1160672

Krötzsch, M., and D. Vrandečić (2011), Semantic MediaWiki. Foundations for the Web of Information and Services – A Review of 20 Years of Semantic Web Research, pp. 311–326. Springer

Kucera, M., D. Khider, and L. Lisiecki (2013), Reporting standards for Paleoceanographic/Paleoclimate data. Retrieved online from http://wiki.linked.earth/wiki/images/d/d4/Reporting_Standards_for_Paleoceanographic_PMIP3_Dec2013.docx on May 28th 2019.

Libby, W. F., E. C. Anderson, and J. R. Arnold (1949), Age determination by radiocarbon content: world-wide assay of natural radiocarbon, Science, 109(2827), 227-228, doi: 10.1126/science.109.2827.227

Lin, L., D. Khider, L. E. Lisiecki, and C. E. Lawrence (2014), Probabilistic sequence alignment of stratigraphic records, Paleoceanography, 29(976-989), 976-989, doi: 10.1002/2014PA002713

Lisiecki, L. E., and M. E. Raymo (2005), A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}O$ records, Paleoceanography, 20(PA1003), doi: 1010.1029/2004PA001071

McKay, N. P., and J. Emile-Geay (2016), Technical Note: The Linked Paleo Data framework – a common tongue for paleoclimatology, Climate of the Past, 12, 1093-1100, doi: 10.5194/cp-12-1093-2016

McKay, N., J. Emile-Geay, C. Heiser, and D. Khider (2018), GeoChronR, doi:10.5281/zenodo.60812

Masson-Delmotte, V., M. Schulz, A. Abe-Ouchi, J. Beer, A. Ganopolski, J. G. Rouco, E. Jansen, K. Lambeck, J. Luterbacher, T. Naish, T. Osborn, B. Otto-Bliesner, T. Quinn, R. Ramesh, M. Rojas, X. Shao, and A. Timmermann (2013), Information from paleoclimate archives, in Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley, chap. 5, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Millard, A. R. (2014). Conventions for reporting radiocarbon determinations. *Radiocarbon, 56*(2), 555-559. doi:10.2458/56.17455

National Oceanographic and Atmospheric Administration. (2018) PaST (Paleoenvironmental Standard Terms) Thesaurus. Retrieved from https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/past-thesaurus on May 28th 2019.

Olsson, I. U. (1970), The use of Oxalic acid as a standard, in Radiocarbon variations and absolute chronology, Nobel symposium, 12th Proc, edited by O. I.U., p. 17, John Wiley & Sons, New York.

PAGES2k Consortium (2017), A global multiproxy database for temperature reconstructions of the Common Era, Sci Data, 4, 170088, doi: 10.1038/sdata.2017.88

Reimer, P. J., T. A. Brown, and R. W. Reimer (2004), Discussion: Reporting and calibration of post-bomb 14C data, Radiocarbon, 46, 1299-1304. doi: 10.1017/S0033822200033154

Riemann, D., Glaser, R., Kahle, M., Vogt, S. (2016). The CRE tambora.org – new data and tools for collaborative research in climate and environmental history. Geoscience Data Journal 2(2):63-77. DOI:10.1002/gdj3.30

Stall, S., E. Robinson, L. Wyborn, L. R. Yarmey, M. A. Parsons, K. Lehnert, B. Cutcher-Gershenfeld, B. Nosek, and B. Hanson (2017), Enabling FAIR data across the Earth and space sciences, EOS, 98, doi: 10.1029/2017EO088425

Stott, L., Cannariato, K., Thunell, R., Haug, G. H., Koutavas, A., & Lund, S. (2004). Decline of surface temperature and salinity in the western tropical Pacific Ocean in the Holocene epoch. *Nature*, *431*, 56-59. doi:10.1038/nature02903

Stott, L., Timmermman, A., & Thunell, R. (2007). Southern Hemisphere and Deep-Sea Warming led to deglacial atmospheric CO_2 rise and tropical warming. *Science*, *318*, 435-438. doi:10.1126/science.1143791

Stuiver, M., and H. A. Polach (1977), Discussion: Reporting of 14C Data, Radiocarbon, 19(3), 355-363, doi: 10.1017/S0033822200003672

Unidata, (2019): Network Common Data Form version 4.7.0 [software]. Boulder, CO: UCAR/Unidata. doi:10.5065/D6H70CW6

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data, 3*, 160018. doi:10.1038/sdata.2016.18

Williams, J.W., Newton, A.J., Kaufman, D.S., von Gunten, L. (eds) (2018) Building and Harnessing open PaleoData, Past Global Changes Magazine, 26(2), 45-96, doi:10.22498/pages.26.2

Wolff, E. W. (2007), When is the "present"?, Quaternary Science Reviews, 26(25-28), 3023-3024, doi: 10.1016/j.quascirev.2007.10.008.

W3C OWL Working Group. (2012), OWL 2 Web Ontology Language Document Overview (Second Edition), Retrieved online on August 6th 2019 at https://www.w3.org/TR/owl2-overview/

Figure 1. Timeline of the community elicitation for best practices in paleoclimate data reporting. The Workshop on Paleoclimate Data Standard marks the official beginning of the endeavor.

PaCTS collects responses from the LinkedEarth platform, Twitter polls, and survey up to November 2017.

Figure 2. Example of polls on a. the LinkedEarth platform and b. Twitter (@Linked_Earth)

Figure 3. Example of a survey question for a new dataset. The histogram represents the number of votes on each platform (orange: LinkedEarth, purple: Twitter, and green: Google survey). The pie chart represents the fraction of the votes for essential (green), recommended (pink), and desired (blue).

Figure 4. Same as Figure 3 for a legacy dataset.

Figure 5. Mind map of the various properties identified by the WGs and associated vote. Colors represent the different WGs. Parentheses indicate a different reporting standard for legacy datasets when different from new datasets. Available online at: https://coggle.it/diagram/WqMd49MJtB8DbqfH/t/community-standards-for-paleoclimate-data-and-metadata.

Figure 6. Mosaic plots for a. new datasets and b. legacy datasets showing the number of essential, recommended, and desired metadata for the various WGs. The height of the bar represents the fraction of total occurrences for essential (e), recommended (r), and desired (d) votes, while the width of the bar represents the number of properties voted on in each WG.

Figure 7. Mind map of the various properties identified by the cross-archive WG and associated vote. Color is the same as in Figure 5. Parentheses indicate recommendations for legacy datasets when different from new datasets. Available online at: https://coggle.it/diagram/W4W9podcxp86PPvf/t/cross-archive-metadata

Figure 8. Mind map of the various properties identified by the ice core archives WG and associated vote. Color is the same as in Figure 5. Parentheses indicate recommendations for legacy datasets when different from new datasets. Available online at: https://coggle.it/diagram/W4XNNeGhIngfjHzB/t/historical-documents

Figure 9. Mind map of the various properties identified by the lake sediments archives WG and associated vote. Color is the same as in Figure 5. Parentheses indicate recommendations for legacy datasets when different from new datasets. Available online at: https://coggle.it/diagram/W4h9m-GhIjjbm3yX/t/lake-sediments

Figure 10. Mind map of the various properties identified by the marine sediments archives WG and associated vote. Color is the same as in Figure 5. Parentheses indicate recommendations for legacy datasets when different from new datasets. Available online at: https://coggle.it/diagram/W4iIkodcxlDKTK6v/t/marine-sediments

Figure 11. Mind map of the various properties identified by the speleothem archives WG and associated vote. Color is the same as in Figure 5. Parentheses indicate recommendations for legacy datasets when different from new datasets. Available online at: https://coggle.it/diagram/W4gwj-GhIl4VmfYP/t/speleothem

Figure 12. Mind map of the various properties identified by tree-based archives WG and associated vote. Color is the same as in Figure 5. Parentheses indicate recommendations for legacy datasets when different from new datasets. Available online at: https://coggle.it/diagram/W4huaYdcxhdzTB9z/t/trees

Figure 13. Mind map of the various properties identified by the documentary archives WG and associated vote. Color is the same as in Figure 5. Parentheses indicate recommendations for legacy datasets when different from new datasets. Available online at: https://coggle.it/diagram/W4XNNeGhIngfjHzB/t/historical-documents

Figure 14. Mind map of the various properties identified by the uncertainties WG and associated vote. Color is the same as in Figure 5. Parentheses indicate recommendations for legacy datasets when different from new datasets. Available online at: https://coggle.it/diagram/W4gttodcxjfvSst0/t/uncertainties

Figure 15. Mind map of the various properties identified by the chronologies WG and associated vote. Color is the same as in Figure 5. Parentheses indicate recommendations for legacy datasets when different from new datasets. Available online at: https://coggle.it/diagram/W4hzXeGhIi5Fm0q7/t/chronologies

Figure 16. Radar plot showing the completeness of the metadata reporting for core MD98-2181 (Khider et al., 2014) for properties considered a. essential and b. recommended in the current study. The axis refers to the working group standards recommendation applicable to the record.

Reported Age/year in manuscript	Significand	Exponent	Direction	Datum
21 ka BP	21	3	before	1950 CE
1816 AD	1816	0	since	0 CE[1]
2.7 Ma	2.7	6	before	1950 CE
127 ka BP	127	3	before	2000 CE

Table 1. Illustration of our proposed time representation with four time points. The first column gives examples of reported age/year in a paleoclimate paper while the last four columns show an implementation of the representation proposed here.

Figure 1.

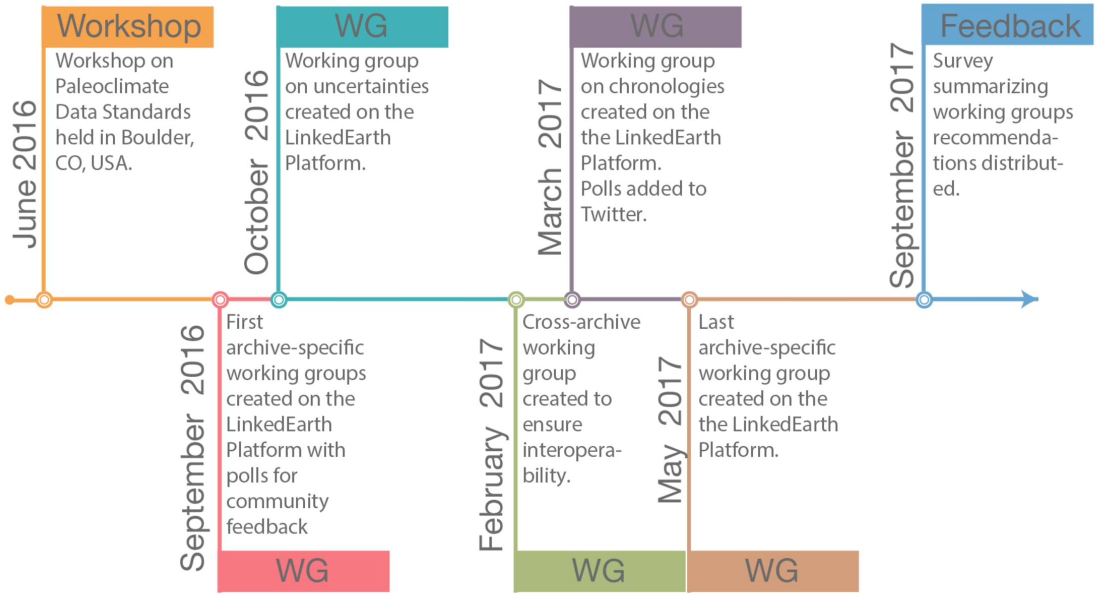

June 2016
Workshop
Workshop on Paleoclimate Data Standards held in Boulder, CO, USA.
September 2016
First archive-specific working groups created on the LinkedEarth Platform with polls for community feedback
WG
October 2016
WG
Working group on uncertainties created on the LinkedEarth Platform.
February 2017
Cross-archive working group created to ensure interopera-bility.
WG
March 2017
WG
Working group on chronologies created on the the LinkedEarth Platform.
Polls added to Twitter.
May 2017
Last archive-specific working group created on the the LinkedEarth Platform.
WG
September 2017
Feedback
Survey summarizing working groups recommenda-tions distribut-ed.

Figure 2.

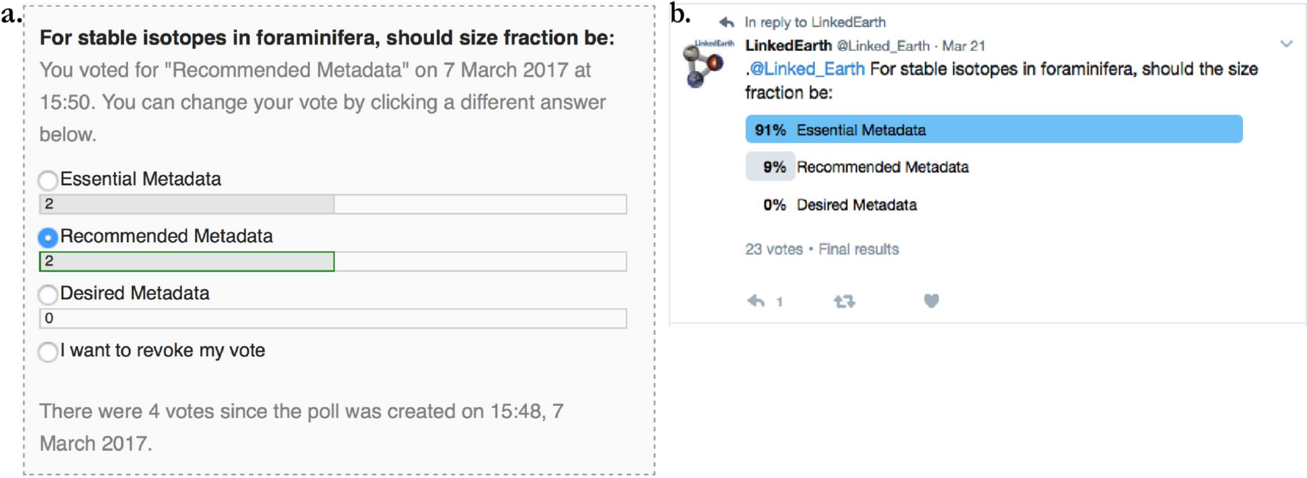
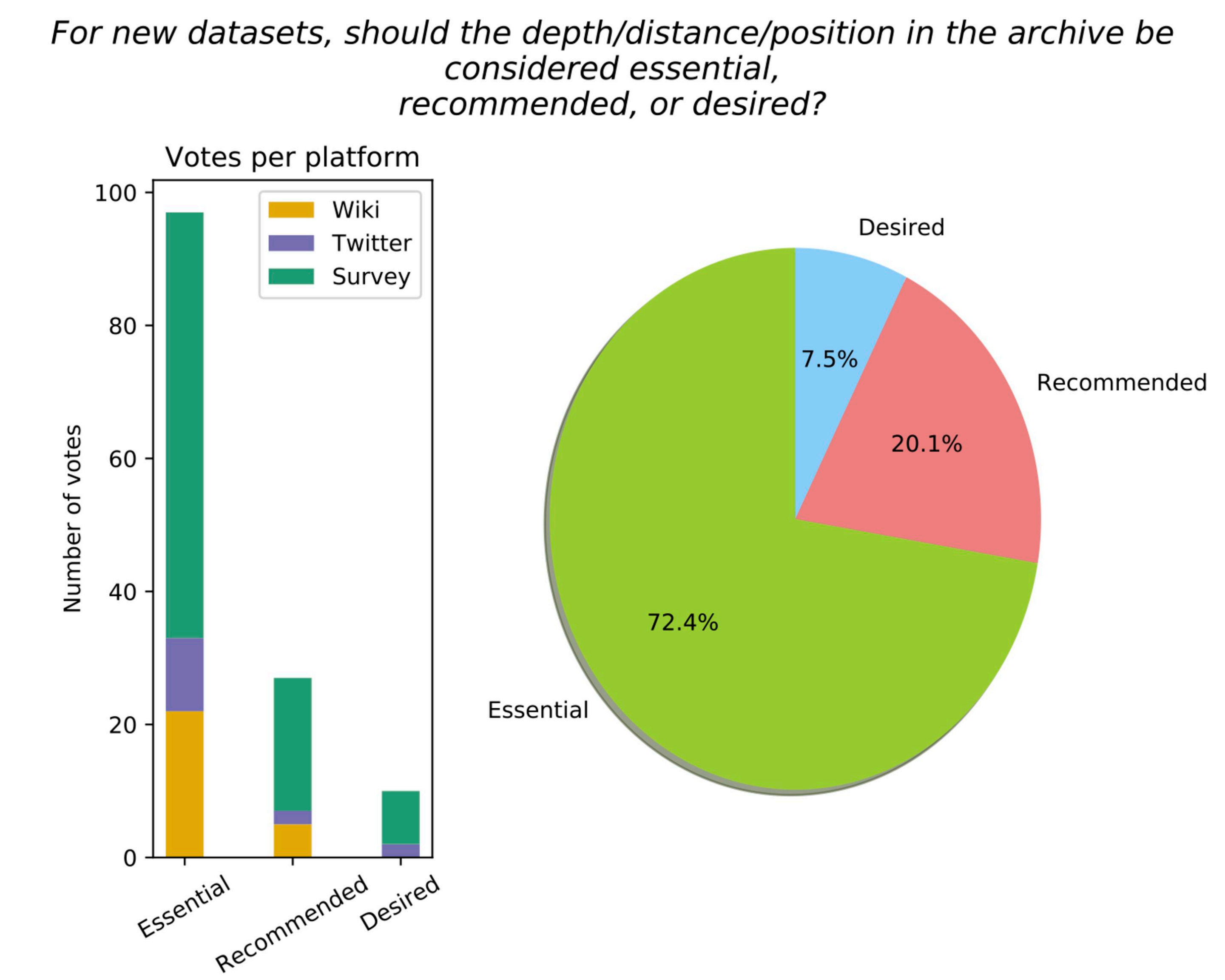

a.
For stable isotopes in foraminifera, should size fraction be:
You voted for "Recommended Metadata" on 7 March 2017 at 15:50. You can change your vote by clicking a different answer below.
Essential Metadata
2
Recommended Metadata
2
Desired Metadata
0
I want to revoke my vote
There were 4 votes since the poll was created on 15:48, 7 March 2017.
b.
In reply to LinkedEarth
LinkedEarth @Linked_Earth · Mar 21
.@Linked_Earth For stable isotopes in foraminifera, should the size fraction be:
91% Essential Metadata
9% Recommended Metadata
0% Desired Metadata
23 votes • Final results
1

Figure 3.



For new datasets, should the depth/distance/position in the archive be considered essential, recommended, or desired?
Votes per platform
Number of votes
100
80
60
40
20
0
Wiki
Twitter
Survey
Essential
Recommended
Desired
Desired
Recommended
Essential
7.5%
20.1%
72.4%

Figure 4.

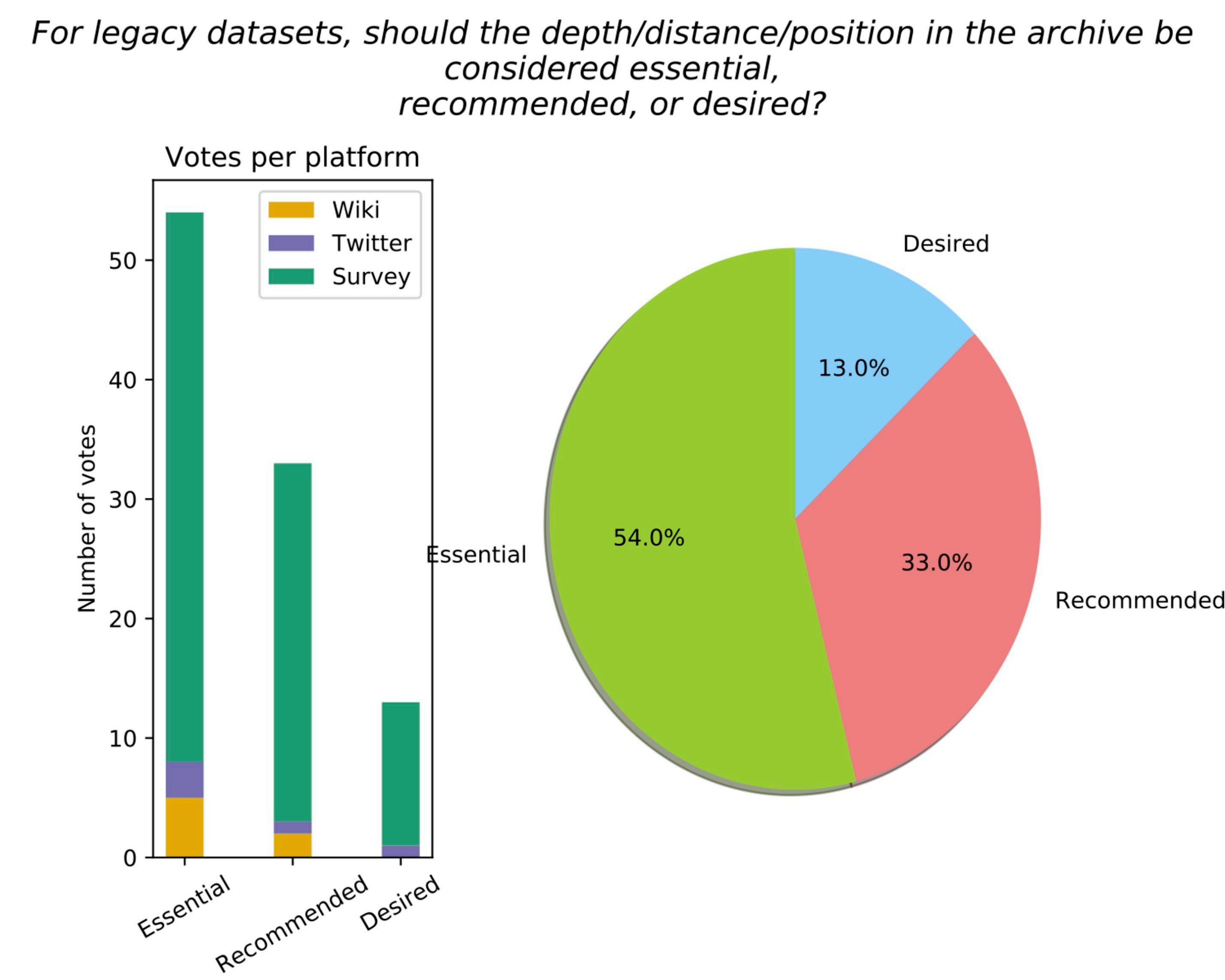
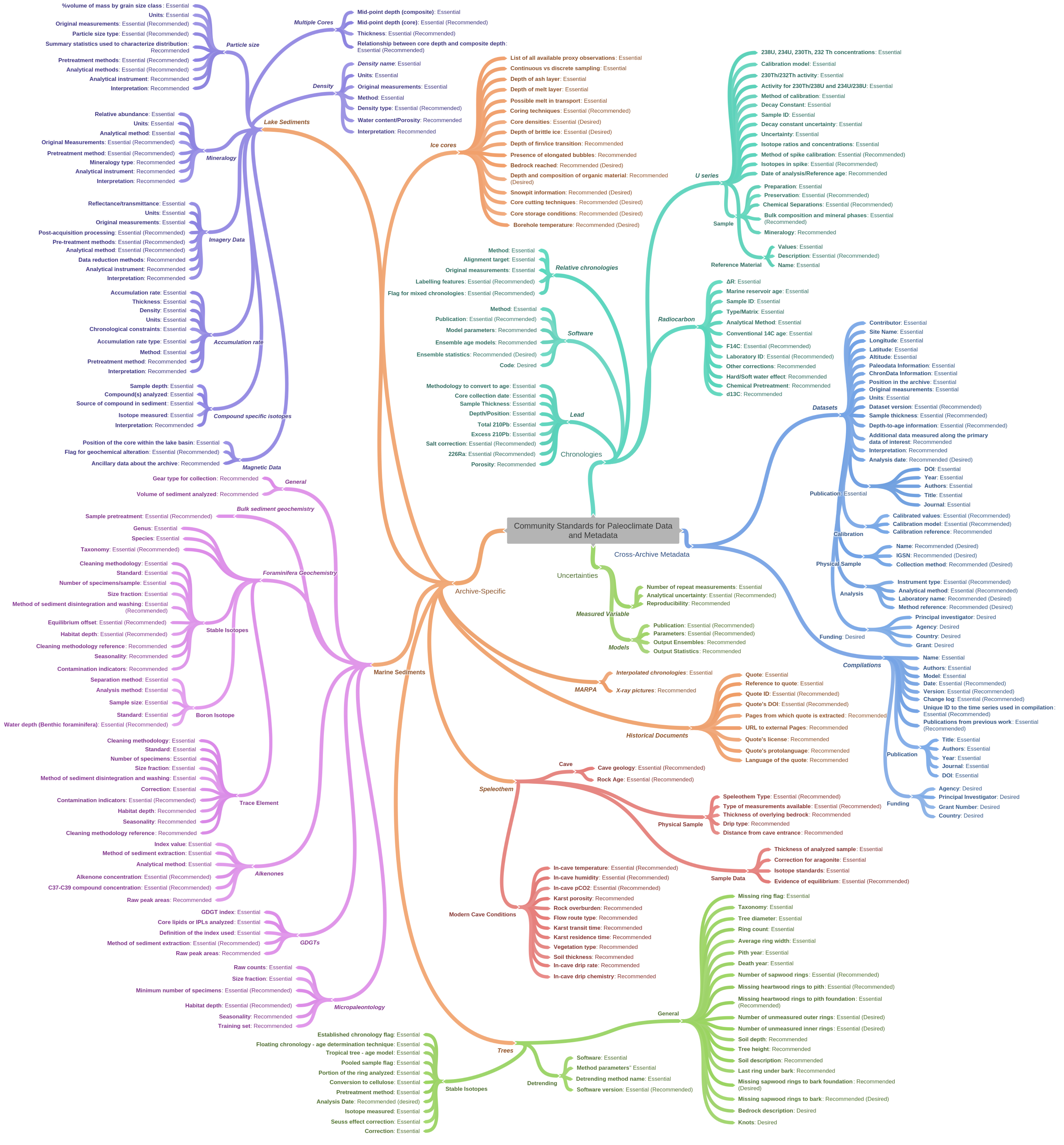

For legacy datasets, should the depth/distance/position in the archive be considered essential,
recommended, or desired?
Votes per platform
Wiki
Twitter
Survey
Number of votes
0
10
20
30
40
50
Essential
Recommended
Desired
Desired
13.0%
Essential
54.0%
33.0%
Recommended

Figure 5.



Community Standards for Paleoclimate Data and Metadata
Archive-Specific
Lake Sediments
Particle size
%volume of mass by grain size class: Essential
Units: Essential
Original measurements: Essential (Recommended)
Particle size type: Essential (Recommended)
Summary statistics used to characterize distribution: Recommended
Pretreatment methods: Essential (Recommended)
Analytical methods: Essential (Recommended)
Analytical instrument: Recommended
Interpretation: Recommended
Multiple Cores
Mid-point depth (composite): Essential
Mid-point depth (core): Essential (Recommended)
Thickness: Essential (Recommended)
Relationship between core depth and composite depth: Essential (Recommended)
Density
Density name: Essential
Units: Essential
Original measurements: Essential
Method: Essential
Density type: Essential (Recommended)
Water content/Porosity: Recommended
Interpretation: Recommended
Mineralogy
Relative abundance: Essential
Units: Essential
Analytical method: Essential
Original Measurements: Essential (Recommended)
Pretreatment method: Essential (Recommended)
Mineralogy type: Recommended
Analytical instrument: Recommended
Interpretation: Recommended
Imagery Data
Reflectance/transmittance: Essential
Units: Essential
Original measurements: Essential
Post-acquisition processing: Essential (Recommended)
Pre-treatment methods: Essential (Recommended)
Analytical method: Essential (Recommended)
Data reduction methods: Recommended
Analytical instrument: Recommended
Interpretation: Recommended
Accumulation rate
Accumulation rate: Essential
Thickness: Essential
Density: Essential
Units: Essential
Chronological constraints: Essential
Accumulation rate type: Essential
Method: Essential
Pretreatment method: Recommended
Interpretation: Recommended
Compound specific isotopes
Sample depth: Essential
Compound(s) analyzed: Essential
Source of compound in sediment: Essential
Isotope measured: Essential
Interpretation: Recommended
Magnetic Data
Position of the core within the lake basin: Essential
Flag for geochemical alteration: Essential (Recommended)
Ancillary data about the archive: Recommended
Ice cores
List of all available proxy observations: Essential
Continuous vs discrete sampling: Essential
Depth of ash layer: Essential
Depth of melt layer: Essential
Possible melt in transport: Essential
Coring techniques: Essential (Recommended)
Core densities: Essential (Desired)
Depth of brittle ice: Essential (Desired)
Depth of firn/ice transition: Recommended
Presence of elongated bubbles: Recommended
Bedrock reached: Recommended (Desired)
Depth and composition of organic material: Recommended (Desired)
Snowpit information: Recommended (Desired)
Core cutting techniques: Recommended (Desired)
Core storage conditions: Recommended (Desired)
Borehole temperature: Recommended (Desired)
Marine Sediments
General
Gear type for collection: Recommended
Volume of sediment analyzed: Recommended
Bulk sediment geochemistry
Sample pretreatment: Essential (Recommended)
Foraminifera Geochemistry
Genus: Essential
Species: Essential
Taxonomy: Essential (Recommended)
Stable Isotopes
Cleaning methodology: Essential
Standard: Essential
Number of specimens/sample: Essential
Size fraction: Essential
Method of sediment disintegration and washing: Essential (Recommended)
Equilibrium offset: Essential (Recommended)
Habitat depth: Essential (Recommended)
Cleaning methodology reference: Recommended
Seasonality: Recommended
Contamination indicators: Recommended
Boron Isotope
Separation method: Essential
Analysis method: Essential
Sample size: Essential
Standard: Essential
Water depth (Benthic foraminifera): Essential (Recommended)
Trace Element
Cleaning methodology: Essential
Standard: Essential
Number of specimens: Essential
Size fraction: Essential
Method of sediment disintegration and washing: Essential
Correction: Essential
Contamination indicators: Essential (Recommended)
Habitat depth: Recommended
Seasonality: Recommended
Cleaning methodology reference: Recommended
Alkenones
Index value: Essential
Method of sediment extraction: Essential
Analytical method: Essential
Alkenone concentration: Essential (Recommended)
C37-C39 compound concentration: Essential (Recommended)
Raw peak areas: Recommended
GDGTs
GDGT index: Essential
Core lipids or IPLs analyzed: Essential
Definition of the index used: Essential
Method of sediment extraction: Essential (Recommended)
Raw peak areas: Recommended
Micropaleontology
Raw counts: Essential
Size fraction: Essential
Minimum number of specimens: Essential (Recommended)
Habitat depth: Essential (Recommended)
Seasonality: Recommended
Training set: Recommended
MARPA
Interpolated chronologies: Essential
X-ray pictures: Recommended
Historical Documents
Quote: Essential
Reference to quote: Essential
Quote ID: Essential (Recommended)
Quote's DOI: Essential (Recommended)
Pages from which quote is extracted: Recommended
URL to external Pages: Recommended
Quote's license: Recommended
Quote's protolanguage: Recommended
Language of the quote: Recommended
Speleothem
Cave
Cave geology: Essential (Recommended)
Rock Age: Essential (Recommended)
Physical Sample
Speleothem Type: Essential (Recommended)
Type of measurements available: Essential (Recommended)
Thickness of overlying bedrock: Recommended
Drip type: Recommended
Distance from cave entrance: Recommended
Sample Data
Thickness of analyzed sample: Essential
Correction for aragonite: Essential
Isotope standards: Essential
Evidence of equilibrium: Essential (Recommended)
Modern Cave Conditions
In-cave temperature: Essential (Recommended)
In-cave humidity: Essential (Recommended)
In-cave pCO2: Essential (Recommended)
Karst porosity: Recommended
Rock overburden: Recommended
Flow route type: Recommended
Karst transit time: Recommended
Karst residence time: Recommended
Vegetation type: Recommended
Soil thickness: Recommended
In-cave drip rate: Recommended
In-cave drip chemistry: Recommended
Trees
General
Missing ring flag: Essential
Taxonomy: Essential
Tree diameter: Essential
Ring count: Essential
Average ring width: Essential
Pith year: Essential
Death year: Essential
Number of sapwood rings: Essential (Recommended)
Missing heartwood rings to pith: Essential (Recommended)
Missing heartwood rings to pith foundation: Essential (Recommended)
Number of unmeasured outer rings: Essential (Desired)
Number of unmeasured inner rings: Essential (Desired)
Soil depth: Recommended
Tree height: Recommended
Soil description: Recommended
Last ring under bark: Recommended
Missing sapwood rings to bark foundation: Recommended (Desired)
Missing sapwood rings to bark: Recommended (Desired)
Bedrock description: Desired
Knots: Desired
Detrending
Software: Essential
Method parameters" Essential
Detrending method name: Essential
Software version: Essential (Recommended)
Stable Isotopes
Established chronology flag: Essential
Floating chronology - age determination technique: Essential
Tropical tree - age model: Essential
Pooled sample flag: Essential
Portion of the ring analyzed: Essential
Conversion to cellulose: Essential
Pretreatment method: Essential
Analysis Date: Recommended (desired)
Isotope measured: Essential
Seuss effect correction: Essential
Correction: Essential
Chronologies
Relative chronologies
Method: Essential
Alignment target: Essential
Original measurements: Essential
Labelling features: Essential (Recommended)
Flag for mixed chronologies: Essential (Recommended)
Software
Method: Essential
Publication: Essential (Recommended)
Model parameters: Recommended
Ensemble age models: Recommended
Ensemble statistics: Recommended (Desired)
Code: Desired
Lead
Methodology to convert to age: Essential
Core collection date: Essential
Sample Thickness: Essential
Depth/Position: Essential
Total 210Pb: Essential
Excess 210Pb: Essential
Salt correction: Essential (Recommended)
226Ra: Essential (Recommended)
Porosity: Recommended
U series
238U, 234U, 230Th, 232 Th concentrations: Essential
Calibration model: Essential
230Th/232Th activity: Essential
Activity for 230Th/238U and 234U/238U: Essential
Method of calibration: Essential
Decay Constant: Essential
Sample ID: Essential
Decay constant uncertainty: Essential
Uncertainty: Essential
Isotope ratios and concentrations: Essential
Method of spike calibration: Essential (Recommended)
Isotopes in spike: Essential (Recommended)
Date of analysis/Reference age: Recommended
Sample
Preparation: Essential
Preservation: Essential (Recommended)
Chemical Separations: Essential (Recommended)
Bulk composition and mineral phases: Essential (Recommended)
Mineralogy: Recommended
Reference Material
Values: Essential
Description: Essential (Recommended)
Name: Essential
Radiocarbon
ΔR: Essential
Marine reservoir age: Essential
Sample ID: Essential
Type/Matrix: Essential
Analytical Method: Essential
Conventional 14C age: Essential
F14C: Essential (Recommended)
Laboratory ID: Essential (Recommended)
Other corrections: Recommended
Hard/Soft water effect: Recommended
Chemical Pretreatment: Recommended
d13C: Recommended
Uncertainties
Measured Variable
Number of repeat measurements: Essential
Analytical uncertainty: Essential (Recommended)
Reproducibility: Recommended
Models
Publication: Essential (Recommended)
Parameters: Essential (Recommended)
Output Ensembles: Recommended
Output Statistics: Recommended
Cross-Archive Metadata
Datasets
Contributor: Essential
Site Name: Essential
Longitude: Essential
Latitude: Essential
Altitude: Essential
Paleodata Information: Essential
ChronData Information: Essential
Position in the archive: Essential
Original measurements: Essential
Units: Essential
Dataset version: Essential (Recommended)
Sample thickness: Essential (Recommended)
Depth-to-age information: Essential (Recommended)
Additional data measured along the primary data of interest: Recommended
Interpretation: Recommended
Analysis date: Recommended (Desired)
Publication: Essential
DOI: Essential
Year: Essential
Authors: Essential
Title: Essential
Journal: Essential
Calibration
Calibrated values: Essential (Recommended)
Calibration model: Essential (Recommended)
Calibration reference: Recommended
Physical Sample
Name: Recommended (Desired)
IGSN: Recommended (Desired)
Collection method: Recommended (Desired)
Analysis
Instrument type: Essential (Recommended)
Analytical method: Essential (Recommended)
Laboratory name: Recommended (Desired)
Method reference: Recommended (Desired)
Funding: Desired
Principal investigator: Desired
Agency: Desired
Country: Desired
Grant: Desired
Compilations
Name: Essential
Authors: Essential
Model: Essential
Date: Essential (Recommended)
Version: Essential (Recommended)
Change log: Essential (Recommended)
Unique ID to the time series used in compilation: Essential (Recommended)
Publications from previous work: Essential (Recommended)
Publication
Title: Essential
Authors: Essential
Year: Essential
Journal: Essential
DOI: Essential
Funding
Agency: Desired
Principal Investigator: Desired
Grant Number: Desired
Country: Desired

Figure 6.

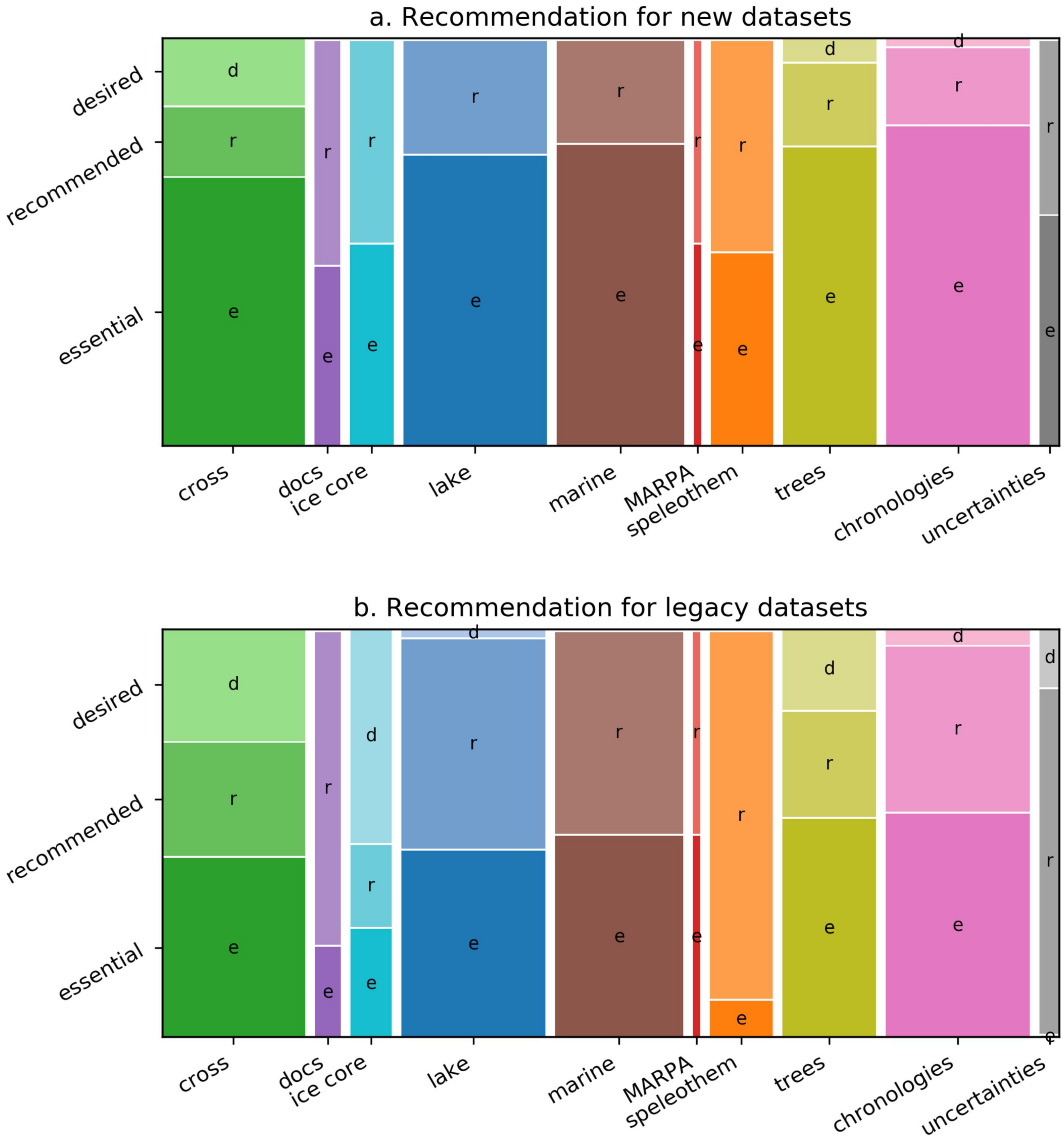
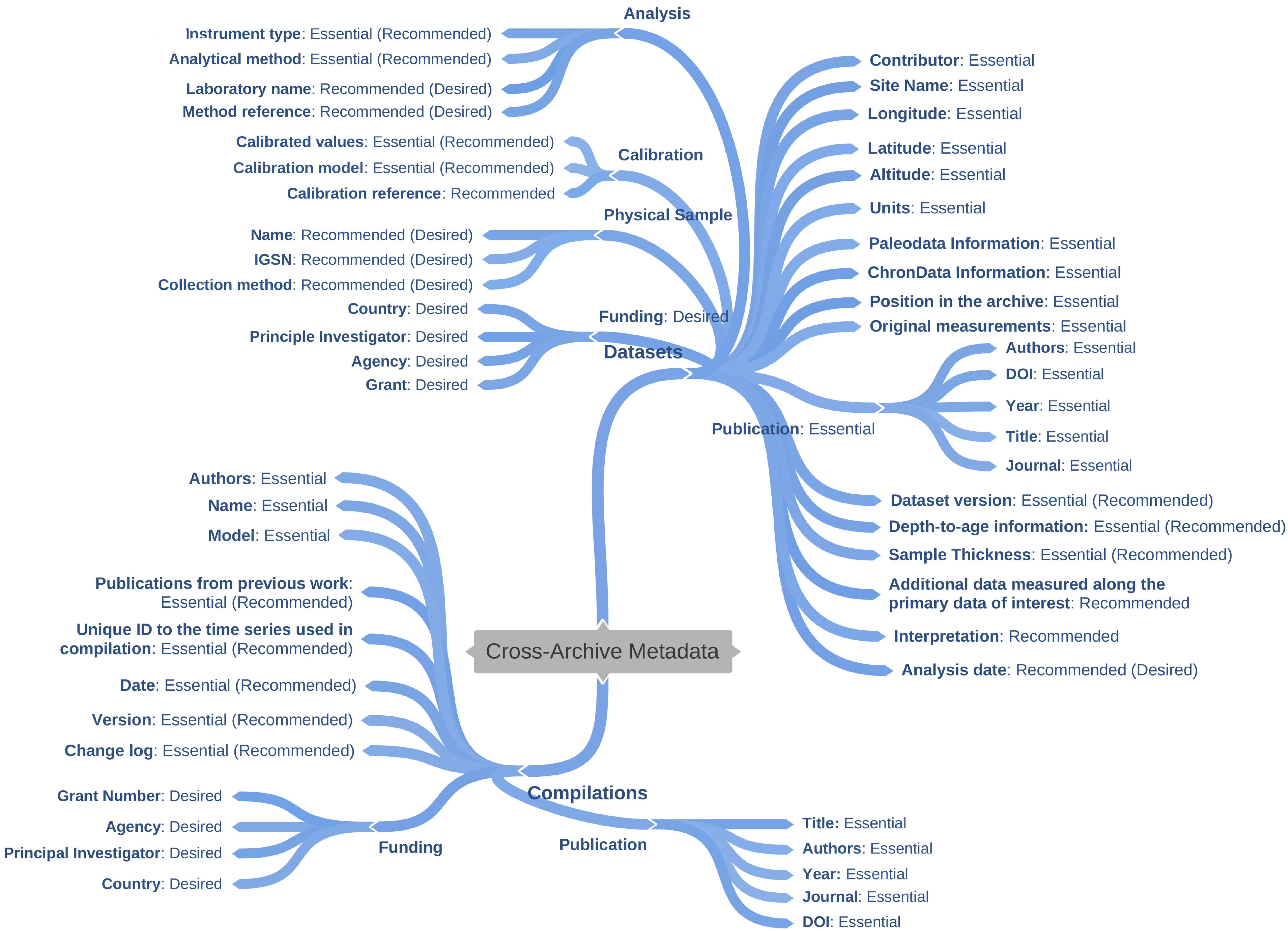

a. Recommendation for new datasets
desired
recommended
essential
d
r
e
cross
docs
ice core
lake
marine
MARPA
speleothem
trees
chronologies
uncertainties
b. Recommendation for legacy datasets

Figure 7.



Cross-Archive Metadata
Datasets
Analysis
Instrument type: Essential (Recommended)
Analytical method: Essential (Recommended)
Laboratory name: Recommended (Desired)
Method reference: Recommended (Desired)
Calibration
Calibrated values: Essential (Recommended)
Calibration model: Essential (Recommended)
Calibration reference: Recommended
Physical Sample
Name: Recommended (Desired)
IGSN: Recommended (Desired)
Collection method: Recommended (Desired)
Funding: Desired
Country: Desired
Principle Investigator: Desired
Agency: Desired
Grant: Desired
Contributor: Essential
Site Name: Essential
Longitude: Essential
Latitude: Essential
Altitude: Essential
Units: Essential
Paleodata Information: Essential
ChronData Information: Essential
Position in the archive: Essential
Original measurements: Essential
Publication: Essential
Authors: Essential
DOI: Essential
Year: Essential
Title: Essential
Journal: Essential
Dataset version: Essential (Recommended)
Depth-to-age information: Essential (Recommended)
Sample Thickness: Essential (Recommended)
Additional data measured along the primary data of interest: Recommended
Interpretation: Recommended
Analysis date: Recommended (Desired)
Compilations
Authors: Essential
Name: Essential
Model: Essential
Publications from previous work: Essential (Recommended)
Unique ID to the time series used in compilation: Essential (Recommended)
Date: Essential (Recommended)
Version: Essential (Recommended)
Change log: Essential (Recommended)
Funding
Grant Number: Desired
Agency: Desired
Principal Investigator: Desired
Country: Desired
Publication
Title: Essential
Authors: Essential
Year: Essential
Journal: Essential
DOI: Essential

Figure 8.

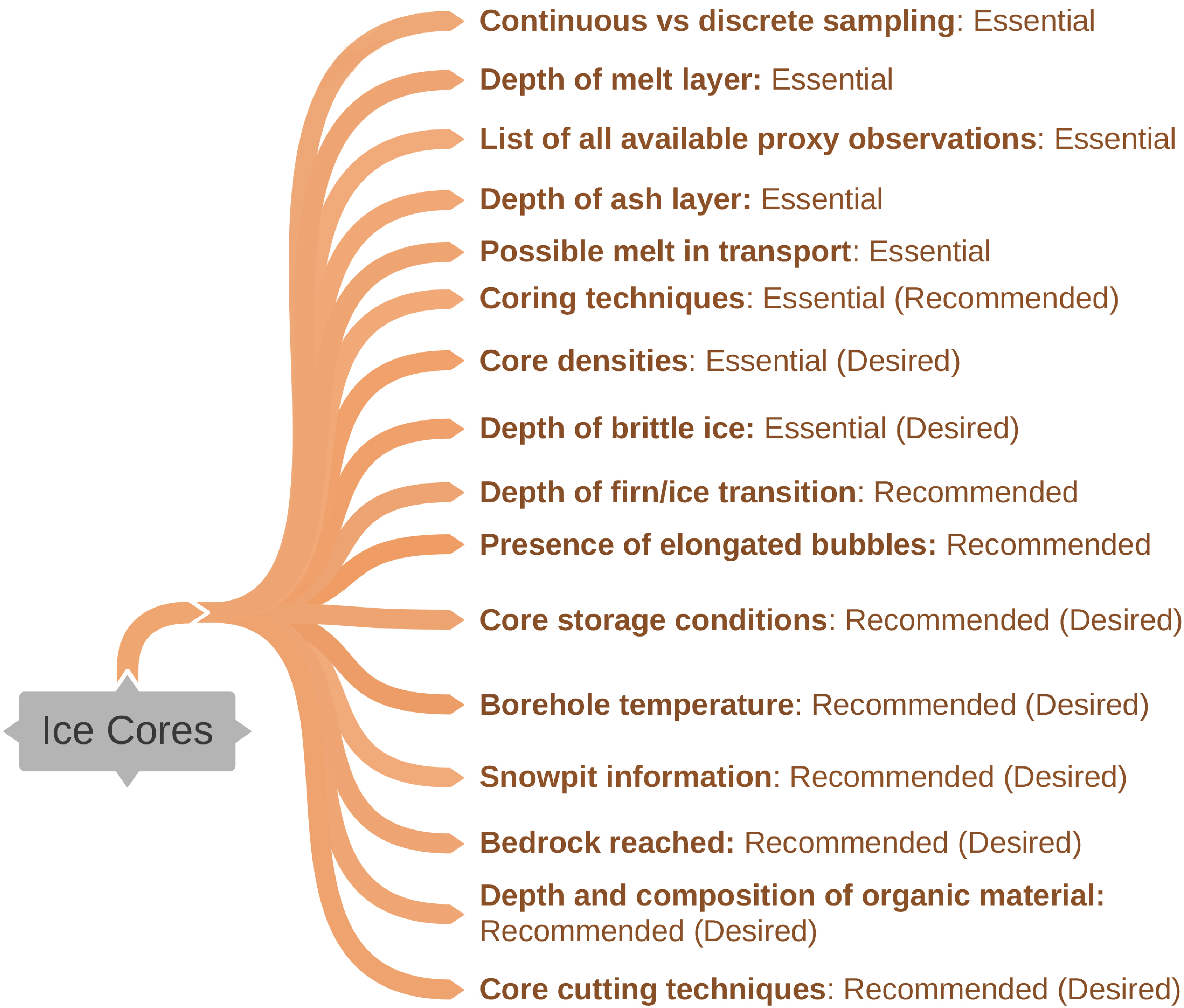

Ice Cores
Continuous vs discrete sampling: Essential
Depth of melt layer: Essential
List of all available proxy observations: Essential
Depth of ash layer: Essential
Possible melt in transport: Essential
Coring techniques: Essential (Recommended)
Core densities: Essential (Desired)
Depth of brittle ice: Essential (Desired)
Depth of firn/ice transition: Recommended
Presence of elongated bubbles: Recommended
Core storage conditions: Recommended (Desired)
Borehole temperature: Recommended (Desired)
Snowpit information: Recommended (Desired)
Bedrock reached: Recommended (Desired)
Depth and composition of organic material: Recommended (Desired)
Core cutting techniques: Recommended (Desired)

Figure 9.

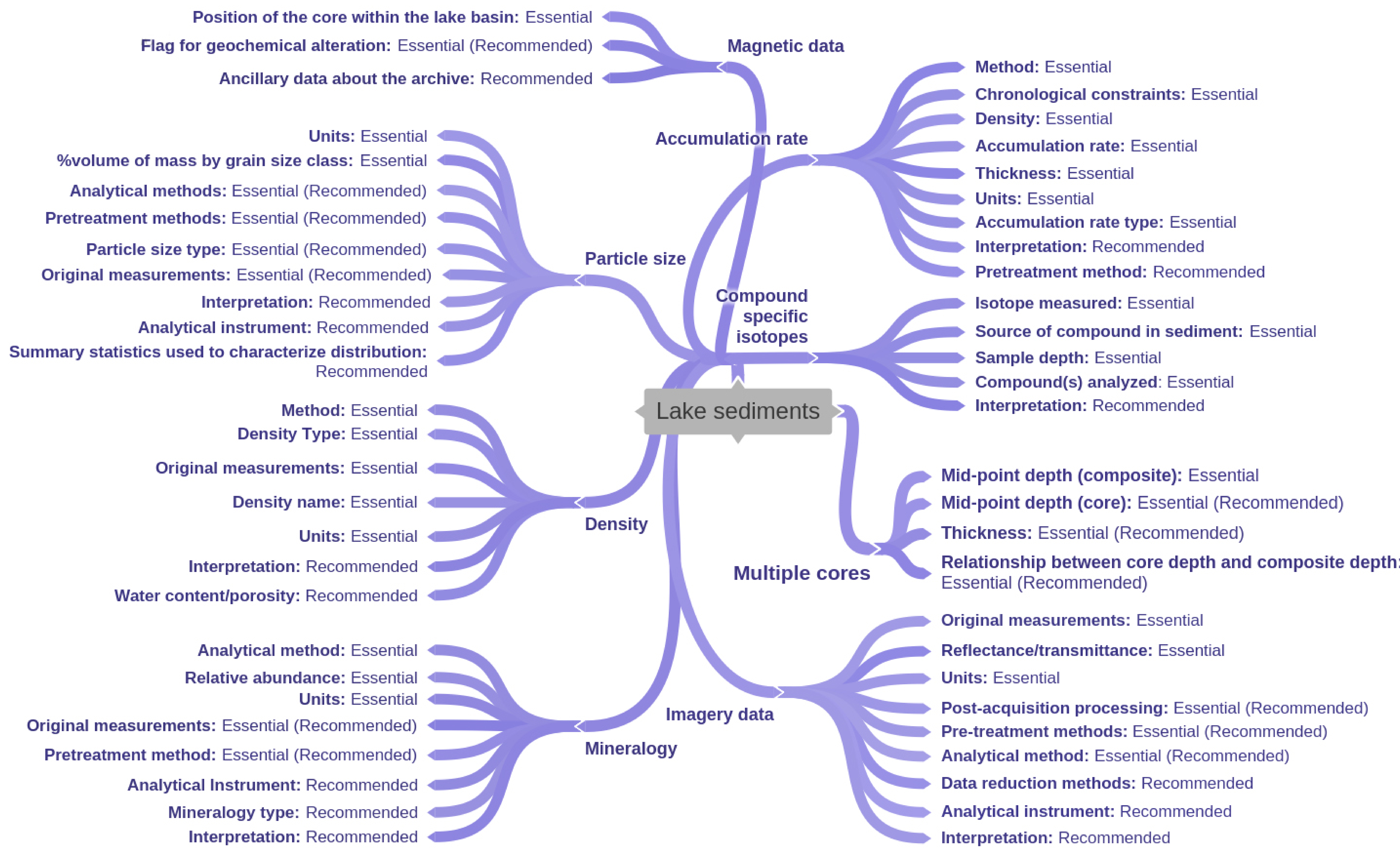

Lake sediments
Magnetic data
Position of the core within the lake basin: Essential
Flag for geochemical alteration: Essential (Recommended)
Ancillary data about the archive: Recommended
Accumulation rate
Method: Essential
Chronological constraints: Essential
Density: Essential
Accumulation rate: Essential
Thickness: Essential
Units: Essential
Accumulation rate type: Essential
Interpretation: Recommended
Pretreatment method: Recommended
Particle size
Units: Essential
%volume of mass by grain size class: Essential
Analytical methods: Essential (Recommended)
Pretreatment methods: Essential (Recommended)
Particle size type: Essential (Recommended)
Original measurements: Essential (Recommended)
Interpretation: Recommended
Analytical instrument: Recommended
Summary statistics used to characterize distribution: Recommended
Compound specific isotopes
Isotope measured: Essential
Source of compound in sediment: Essential
Sample depth: Essential
Compound(s) analyzed: Essential
Interpretation: Recommended
Density
Method: Essential
Density Type: Essential
Original measurements: Essential
Density name: Essential
Units: Essential
Interpretation: Recommended
Water content/porosity: Recommended
Multiple cores
Mid-point depth (composite): Essential
Mid-point depth (core): Essential (Recommended)
Thickness: Essential (Recommended)
Relationship between core depth and composite depth: Essential (Recommended)
Mineralogy
Analytical method: Essential
Relative abundance: Essential
Units: Essential
Original measurements: Essential (Recommended)
Pretreatment method: Essential (Recommended)
Analytical Instrument: Recommended
Mineralogy type: Recommended
Interpretation: Recommended
Imagery data
Original measurements: Essential
Reflectance/transmittance: Essential
Units: Essential
Post-acquisition processing: Essential (Recommended)
Pre-treatment methods: Essential (Recommended)
Analytical method: Essential (Recommended)
Data reduction methods: Recommended
Analytical instrument: Recommended
Interpretation: Recommended

Figure 10.

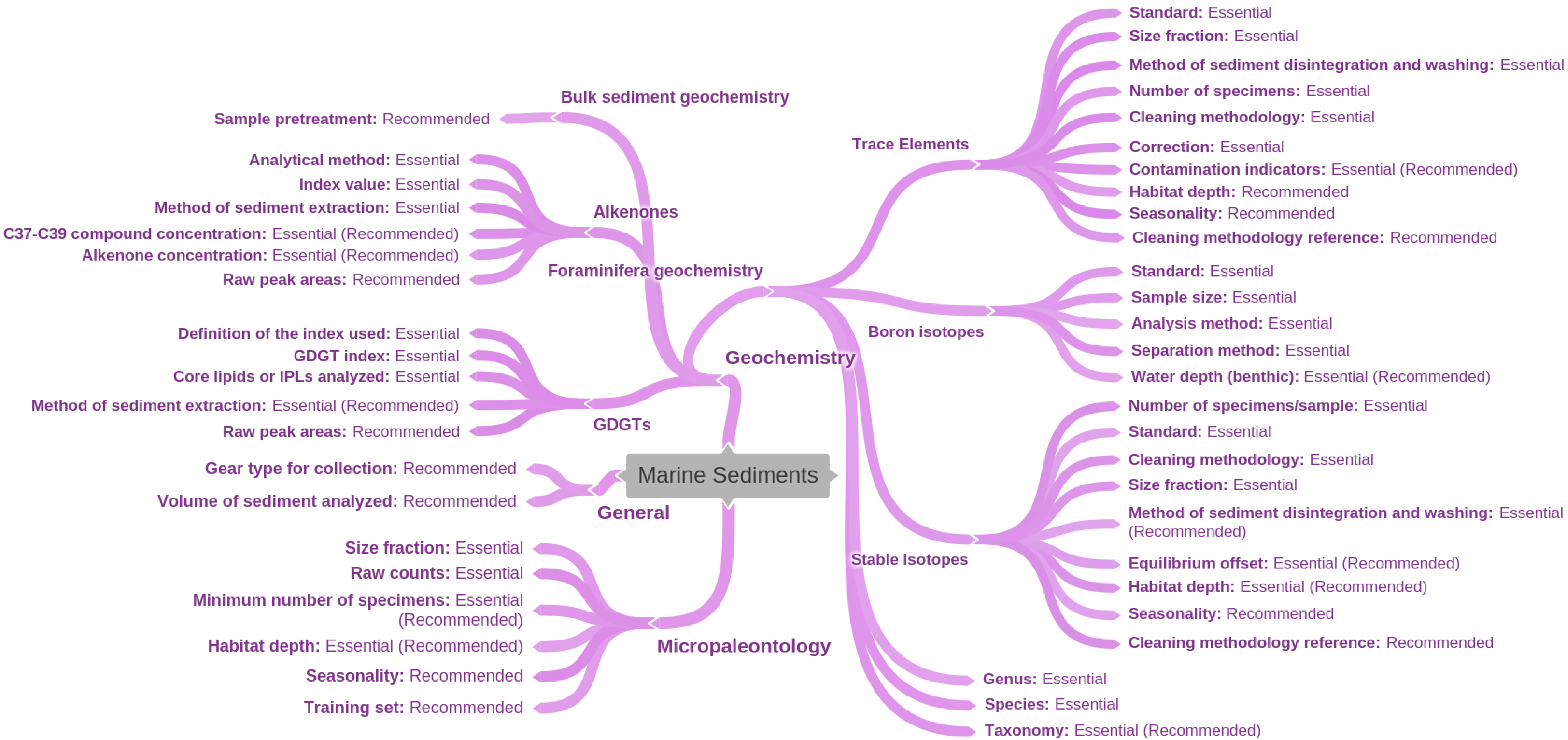

Marine Sediments
General
Gear type for collection: Recommended
Volume of sediment analyzed: Recommended
Micropaleontology
Size fraction: Essential
Raw counts: Essential
Minimum number of specimens: Essential (Recommended)
Habitat depth: Essential (Recommended)
Seasonality: Recommended
Training set: Recommended
Geochemistry
Bulk sediment geochemistry
Sample pretreatment: Recommended
Alkenones
Analytical method: Essential
Index value: Essential
Method of sediment extraction: Essential
C37-C39 compound concentration: Essential (Recommended)
Alkenone concentration: Essential (Recommended)
Raw peak areas: Recommended
GDGTs
Definition of the index used: Essential
GDGT index: Essential
Core lipids or IPLs analyzed: Essential
Method of sediment extraction: Essential (Recommended)
Raw peak areas: Recommended
Foraminifera geochemistry
Trace Elements
Standard: Essential
Size fraction: Essential
Method of sediment disintegration and washing: Essential
Number of specimens: Essential
Cleaning methodology: Essential
Correction: Essential
Contamination indicators: Essential (Recommended)
Habitat depth: Recommended
Seasonality: Recommended
Cleaning methodology reference: Recommended
Boron isotopes
Standard: Essential
Sample size: Essential
Analysis method: Essential
Separation method: Essential
Water depth (benthic): Essential (Recommended)
Stable Isotopes
Number of specimens/sample: Essential
Standard: Essential
Cleaning methodology: Essential
Size fraction: Essential
Method of sediment disintegration and washing: Essential (Recommended)
Equilibrium offset: Essential (Recommended)
Habitat depth: Essential (Recommended)
Seasonality: Recommended
Cleaning methodology reference: Recommended
Genus: Essential
Species: Essential
Taxonomy: Essential (Recommended)

Figure 11.

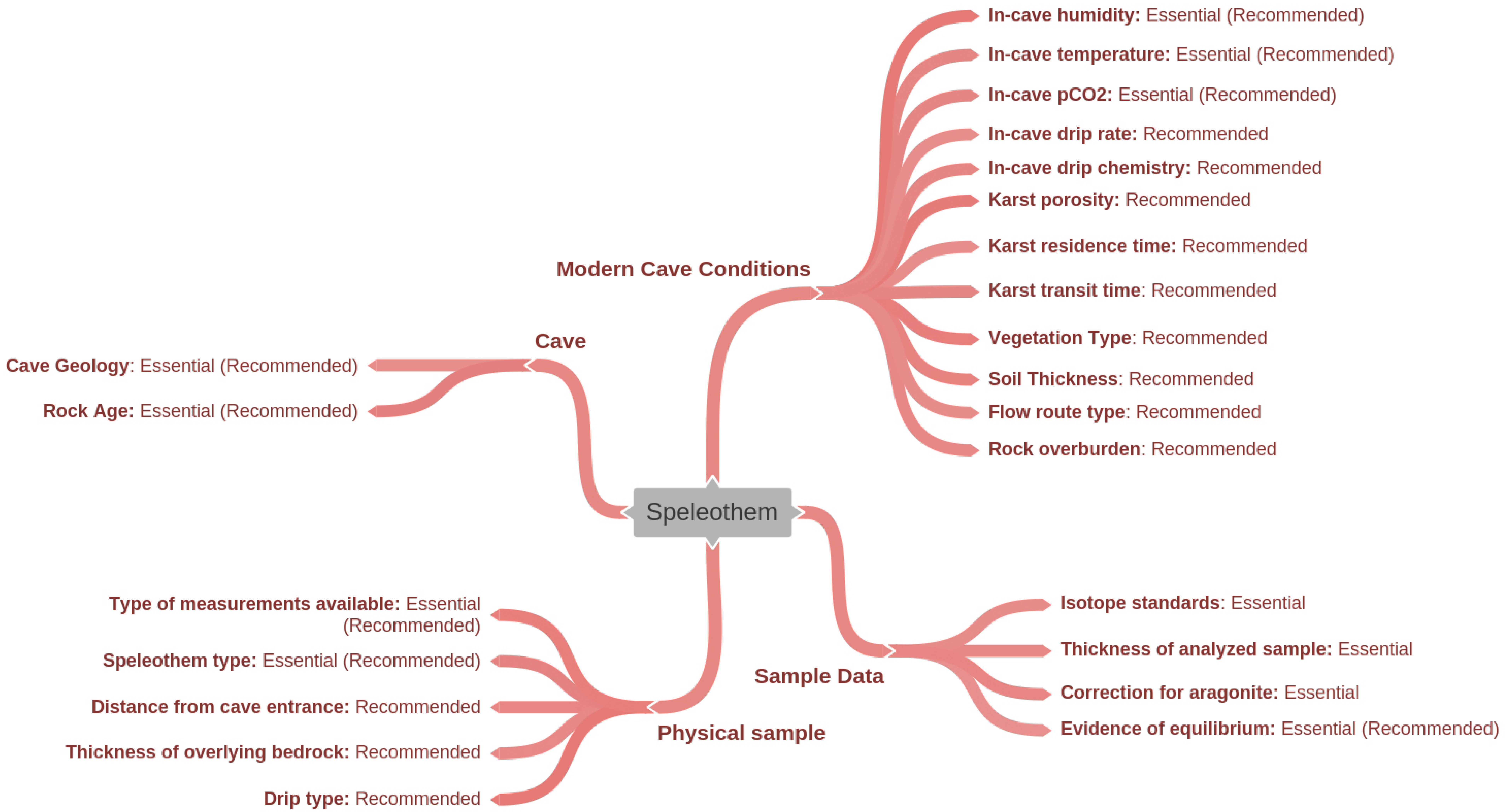

Speleothem
Modern Cave Conditions
In-cave humidity: Essential (Recommended)
In-cave temperature: Essential (Recommended)
In-cave pCO2: Essential (Recommended)
In-cave drip rate: Recommended
In-cave drip chemistry: Recommended
Karst porosity: Recommended
Karst residence time: Recommended
Karst transit time: Recommended
Vegetation Type: Recommended
Soil Thickness: Recommended
Flow route type: Recommended
Rock overburden: Recommended
Cave
Cave Geology: Essential (Recommended)
Rock Age: Essential (Recommended)
Sample Data
Isotope standards: Essential
Thickness of analyzed sample: Essential
Correction for aragonite: Essential
Evidence of equilibrium: Essential (Recommended)
Physical sample
Type of measurements available: Essential (Recommended)
Speleothem type: Essential (Recommended)
Distance from cave entrance: Recommended
Thickness of overlying bedrock: Recommended
Drip type: Recommended

Figure 12.

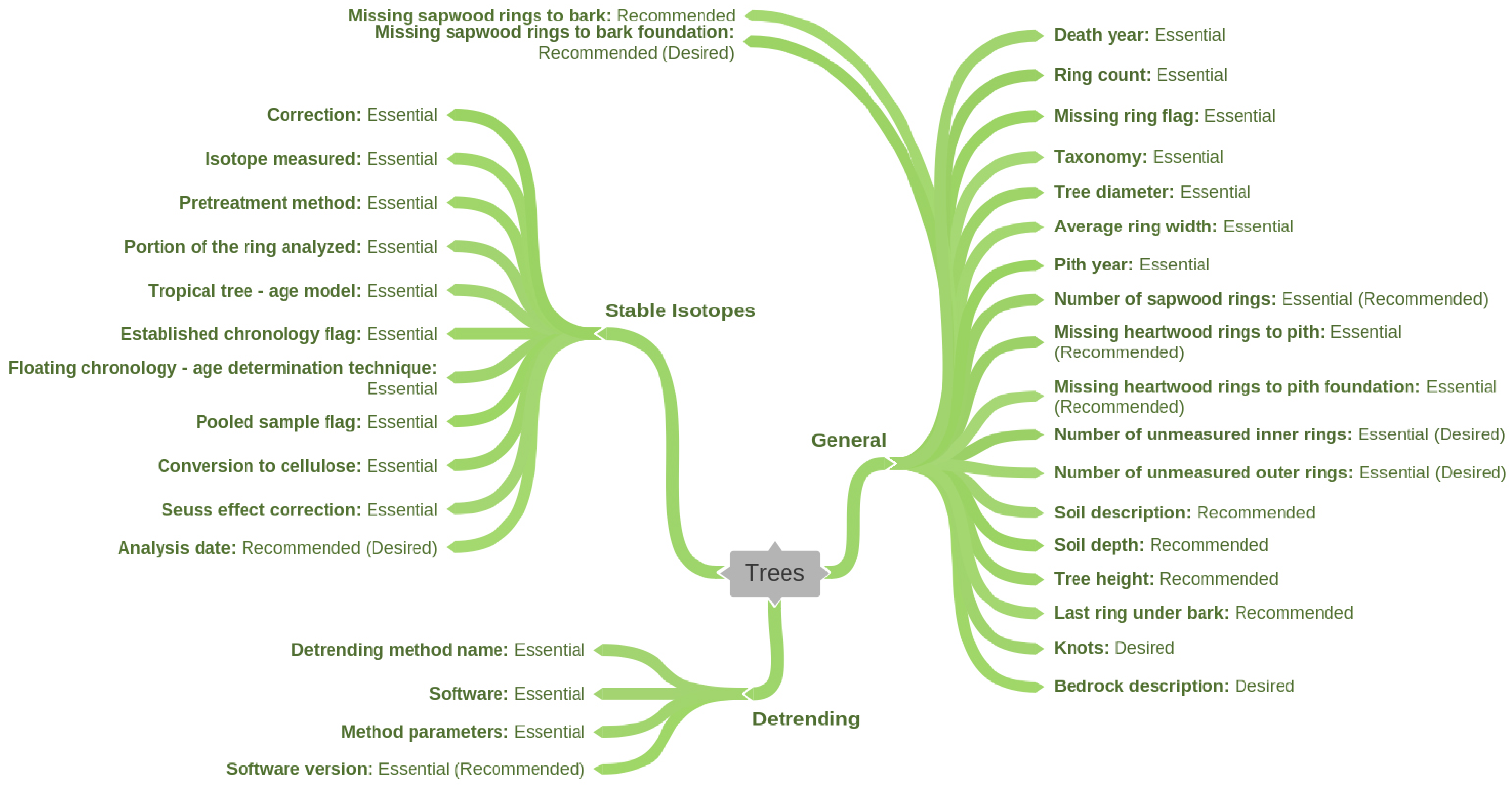

Trees
Stable Isotopes
Correction: Essential
Isotope measured: Essential
Pretreatment method: Essential
Portion of the ring analyzed: Essential
Tropical tree - age model: Essential
Established chronology flag: Essential
Floating chronology - age determination technique: Essential
Pooled sample flag: Essential
Conversion to cellulose: Essential
Seuss effect correction: Essential
Analysis date: Recommended (Desired)
General
Missing sapwood rings to bark: Recommended
Missing sapwood rings to bark foundation: Recommended (Desired)
Death year: Essential
Ring count: Essential
Missing ring flag: Essential
Taxonomy: Essential
Tree diameter: Essential
Average ring width: Essential
Pith year: Essential
Number of sapwood rings: Essential (Recommended)
Missing heartwood rings to pith: Essential (Recommended)
Missing heartwood rings to pith foundation: Essential (Recommended)
Number of unmeasured inner rings: Essential (Desired)
Number of unmeasured outer rings: Essential (Desired)
Soil description: Recommended
Soil depth: Recommended
Tree height: Recommended
Last ring under bark: Recommended
Knots: Desired
Bedrock description: Desired
Detrending
Detrending method name: Essential
Software: Essential
Method parameters: Essential
Software version: Essential (Recommended)

Figure 13.


Historical Documents
Reference to quote: Essential
Quote: Essential
Quote ID: Essential (Recommended)
Quote's DOI: Essential (Recommended)
URL to external pages: Recommended
Language of the quote: Recommended
Quote's license: Recommended
Quote's protolanguage: Recommended
Pages from which quote is extracted: Recommended

Figure 14.

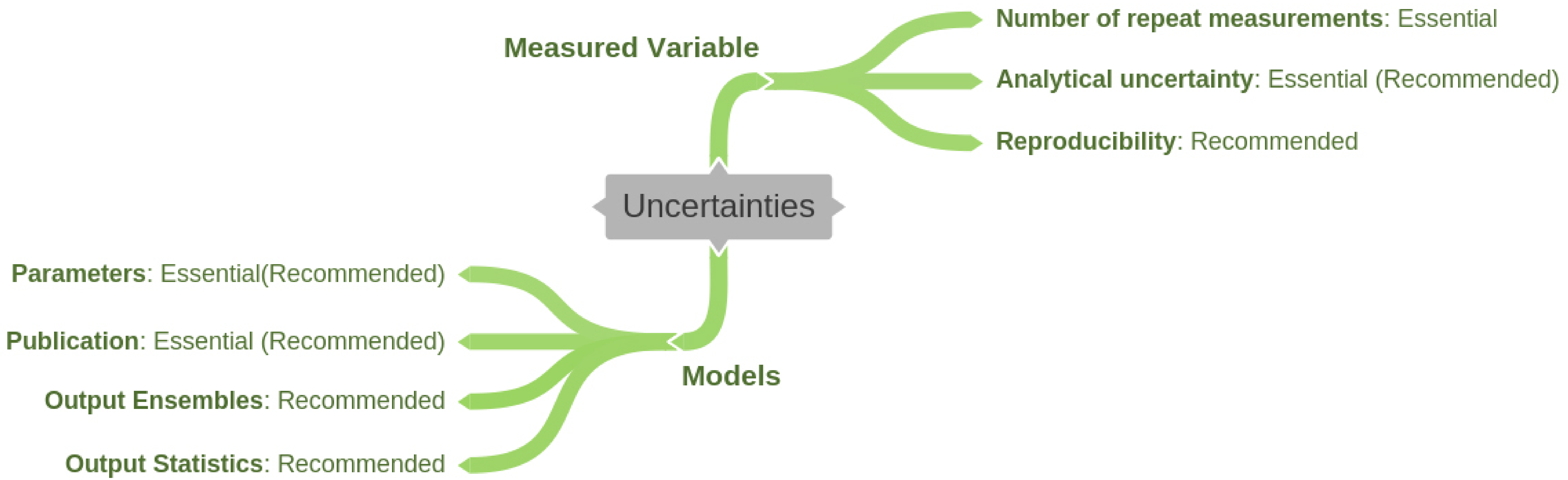

Measured Variable
Number of repeat measurements: Essential
Analytical uncertainty: Essential (Recommended)
Reproducibility: Recommended
Uncertainties
Parameters: Essential(Recommended)
Publication: Essential (Recommended)
Output Ensembles: Recommended
Output Statistics: Recommended
Models

Figure 15.

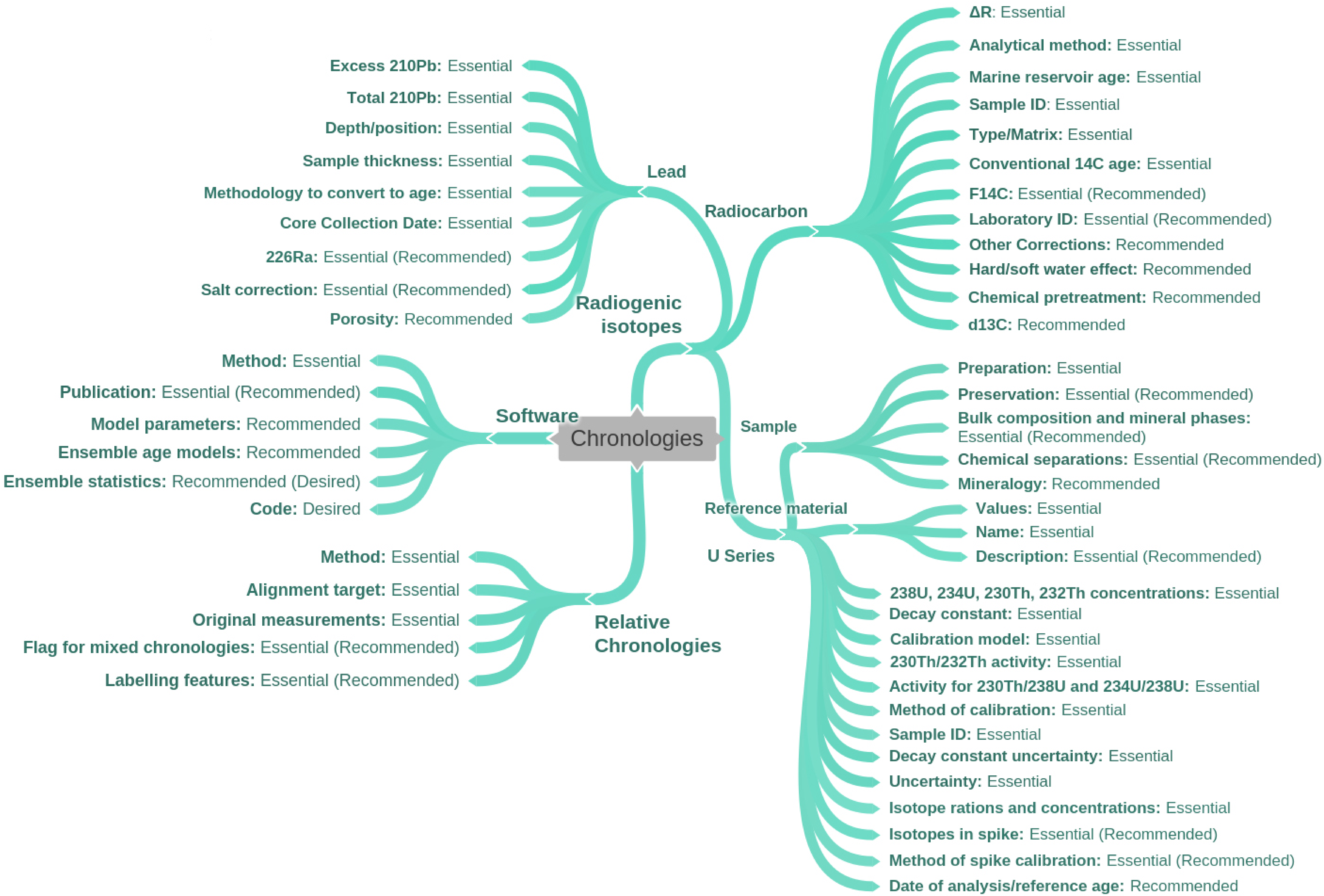
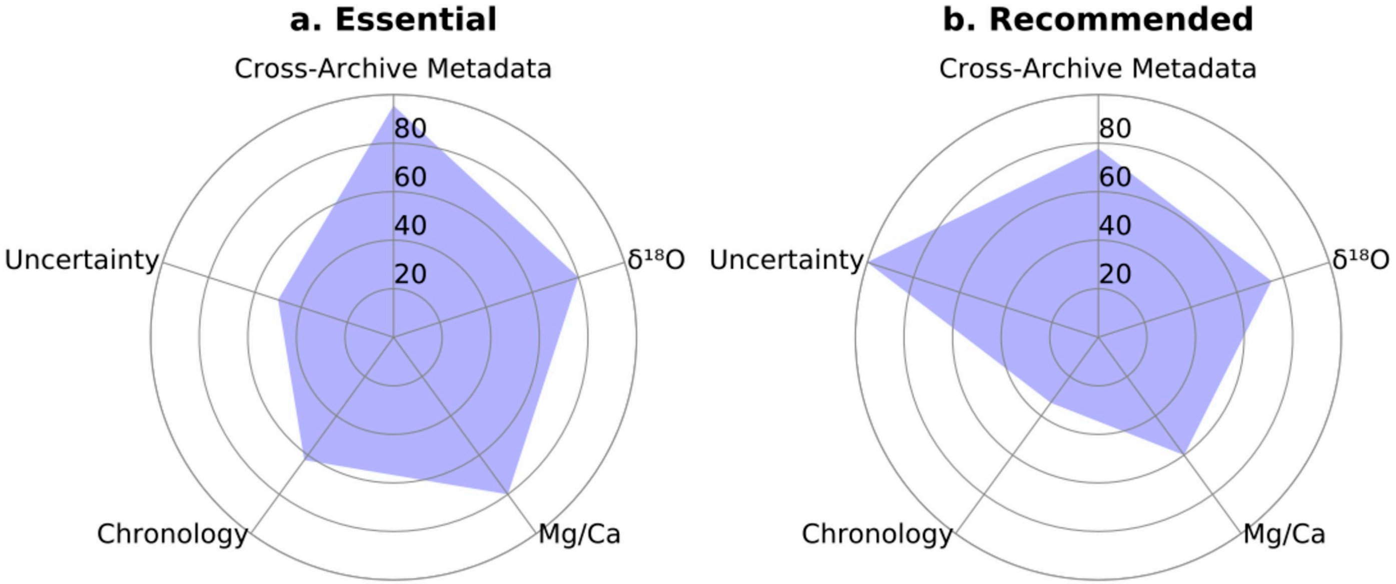

Chronologies
Radiogenic isotopes
Lead
Excess 210Pb: Essential
Total 210Pb: Essential
Depth/position: Essential
Sample thickness: Essential
Methodology to convert to age: Essential
Core Collection Date: Essential
226Ra: Essential (Recommended)
Salt correction: Essential (Recommended)
Porosity: Recommended
Radiocarbon
ΔR: Essential
Analytical method: Essential
Marine reservoir age: Essential
Sample ID: Essential
Type/Matrix: Essential
Conventional 14C age: Essential
F14C: Essential (Recommended)
Laboratory ID: Essential (Recommended)
Other Corrections: Recommended
Hard/soft water effect: Recommended
Chemical pretreatment: Recommended
d13C: Recommended
U Series
Sample
Preparation: Essential
Preservation: Essential (Recommended)
Bulk composition and mineral phases: Essential (Recommended)
Chemical separations: Essential (Recommended)
Mineralogy: Recommended
Reference material
Values: Essential
Name: Essential
Description: Essential (Recommended)
238U, 234U, 230Th, 232Th concentrations: Essential
Decay constant: Essential
Calibration model: Essential
230Th/232Th activity: Essential
Activity for 230Th/238U and 234U/238U: Essential
Method of calibration: Essential
Sample ID: Essential
Decay constant uncertainty: Essential
Uncertainty: Essential
Isotope rations and concentrations: Essential
Isotopes in spike: Essential (Recommended)
Method of spike calibration: Essential (Recommended)
Date of analysis/reference age: Recommended
Software
Method: Essential
Publication: Essential (Recommended)
Model parameters: Recommended
Ensemble age models: Recommended
Ensemble statistics: Recommended (Desired)
Code: Desired
Relative Chronologies
Method: Essential
Alignment target: Essential
Original measurements: Essential
Flag for mixed chronologies: Essential (Recommended)
Labelling features: Essential (Recommended)

Figure 16.


a. Essential
Cross-Archive Metadata
80
60
40
20
δ18O
Mg/Ca
Chronology
Uncertainty
b. Recommended
Cross-Archive Metadata
80
60
40
20
δ18O
Mg/Ca
Chronology
Uncertainty