

PaCTS 1.0: a crowdsourced reporting standard for paleoclimate data

Article

Accepted Version

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. ORCID: <https://orcid.org/0000-0002-7882-4996>, 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. ORCID: <https://orcid.org/0000-0002-0396-0299>, 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, A., Weitzel, N., Wilhelm, B., Williams, J., Williams, J. J., Winstrup, M., Zhao, N. and Zhou, Y. (2019) PaCTS 1.0: a crowdsourced reporting standard for paleoclimate data. *Paleoceanography*, 34 (10). pp. 1570-1596. ISSN 1944-9186 doi:

<https://doi.org/10.1029/2019PA003632> Available at
<https://centaur.reading.ac.uk/85563/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1029/2019PA003632>

Publisher: American Geophysical Union

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

PaCTS 1.0: A Crowdsourced Reporting Standard for Paleoclimate Data

D. Khider^{1,2,*}, J. Emile-Geay², N.P. McKay³, Y. Gil¹, D. Garijo¹, V. Ratnakar¹, M. Alonso-Garcia⁴, S. Bertrand⁵, O. Bothe⁶, P. Brewer⁷, A. Bunn⁸, M. Chevalier⁹, L. Comas-Bru^{10,11}, A. Csank¹², E. Dassié¹³, K. DeLong¹⁴, T. Felis¹⁵, P. Francus¹⁶, A. Frappier¹⁷, W. Gray¹⁸, S. Goring¹⁹, L. Jonkers¹⁵, M. Kahle²⁰, D. Kaufman³, N. M. Kehrwald²¹, B. Martrat^{22,23}, H. McGregor²⁴, J. Richey²⁵, A. Schmittner²⁶, N. Scroxton²⁷, E. Sutherland²⁸, K. Thirumalai²⁹, K. Allen³⁰, F. Arnaud³¹, Y. Axford³², Timothy T. Barrows²⁴, L. Bazin¹⁸, S.E. Pilaar Birch³³, E. Bradley³⁴, J. Bregy³⁵, E. Capron³⁶, O. Cartapanis³⁷, H.-W. Chiang³⁸, K. Cobb³⁹, M. Debret⁴⁰, R. Dommain⁴¹, J. Du²⁶, K. Dyez⁴², S. Emerick⁴³, M. P. Erb³, G. Falster⁴⁴, W. Finsinger⁴⁵, D. Fortier⁴⁶, Nicolas Gauthier⁴⁷, S. George⁴⁸, E. Grimm⁴⁹, J. Hertzberg⁵⁰, F. Hibbert⁵¹, A. Hillman⁵², W. Hobbs⁵³, M. Huber⁵⁴, A.L.C. Hughes^{55,56}, S. Jaccard³⁷, J. Ruan⁵⁷, M. Kienast⁵⁸, B. Konecky⁵⁹, G. Le Roux⁶⁰, V. Lyubchich⁶¹, V.F. Novello⁴³, L. Olaka⁶², J.W. Partin⁶³, C. Pearce⁶⁴, S.J. Phipps⁶⁵, C. Pignol³¹, N. Piotrowska⁶⁶, M.-S. Poli⁶⁷, A. Prokopenko⁶⁸, F. Schwanck⁶⁹, C. Stepanek⁷⁰, G. E. A. Swann⁷¹, R. Telford⁷², E. Thomas⁷³, Z. Thomas⁷⁴, S. Truebe⁷⁵, L. von Gunten⁷⁶, A. Waite⁷⁷, N. Weitzel⁷⁸, B. Wilhelm⁷⁹, J. Williams⁸⁰, J.J. Williams⁸¹, M. Winstrup⁸², N. Zhao⁸³, Y. Zhou⁸⁴.

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

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

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

81 **Key Points:**

- 82 • First version of a crowdsourced reporting standard for paleoclimate data
 - 83 • The standards arose through collective discussions, both in-person and online, and via an
84 innovative social platform
 - 85 • The standard helps meet the interoperability and reuse criteria of FAIR (Findable,
86 Accessible, Interoperable, and Reusable).
- 87

88 **Abstract**

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

107 **1. Introduction**

108 Paleoclimatology is a highly integrative discipline, often requiring the comparison of multiple
109 datasets and model simulations to reach fundamental insights about the climate system.
110 Currently, such syntheses are hampered by the time and effort required to transform the data into

111 a usable format for each application. This task, called “*data wrangling*”, is estimated to consume
112 up to 80% of researcher time in some scientific fields (Dasu and Johnson, 2003), an estimate
113 commensurate with the experience of many paleoclimatologists, particularly at the early-career
114 stage. Wrangling involves not only identifying missing values or outliers in the time series, but
115 also searching multiple databases for the scattered records, contacting the original investigators
116 for the missing data and metadata, and organizing the data into a machine-readable format.
117 Further, this wrangling requires an understanding of each dataset’s originating field and its
118 unspoken practices, and so cannot be easily automated or outsourced to unskilled labor or
119 software. There is therefore an acute need for standardizing paleoclimate datasets.

120

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

134

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

147

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

154

155 Standardization requires three distinct elements: (1) a standard format for the data, (2) a standard
156 terminology for metadata, and (3) standard guidelines for reporting paleoclimate data (i.e.,

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

166

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

177

178 This article's aim is twofold: firstly, to provide a snapshot of the first version of the Paleoclimate
179 Community reporTing Standard (PaCTS), as of 2019, with the understanding that this standard

180 will eventually evolve; secondly, to document the process of community elicitation of such
181 guidelines, so as to provide maximum transparency on why and how these decisions were made.
182 We start from the premise that sampling decisions predate these reporting decisions, so the
183 standard aims to guide an investigator's decisions as to how they should report existing
184 measurements, e.g., at the time of publication.

185
186 The remaining sections are organized as follows: Section 2 summarizes the relevant prior
187 standardization efforts, which serve as the foundation for PaCTS v1.0. Section 3 describes the
188 standardization process, including eliciting community feedback. Section 4 presents
189 recommendation from a group of 135 international researchers actively engaged in paleoclimate
190 research. Section 5 illustrates the application of PaCTS v1.0 to an existing paleoclimate record.
191 Finally, Section 6 concludes with a plan to disseminate the first version of PaCTS within the
192 paleoclimate community and provides a roadmap for further standards development and their
193 future applications.

194 **2. Background**

195 **2.1. The LinkedEarth Framework: an online approach to standard development**

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

202 highly-structured metadata about paleoclimate datasets using the terms in the LinkedEarth
203 Ontology.

204

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

222

223 The LinkedEarth Platform allows users to 1) describe paleoclimate datasets using the terms
224 available in the LinkedEarth Ontology, and 2) propose new terms if they cannot find an

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

239

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

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

258 **2.2. Previous and concurrent efforts towards a data standard**

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

261 **2.1.1. Origins of a standard format for paleoclimate data**

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

269

270 LiPD is a universally readable data container that organizes paleoclimate data and metadata in a

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

283

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

294 definition. PaCTS details a crowdsourced approach for deciding what information should be
295 included when reporting paleoclimate data, a “CF convention” for paleoclimate datasets.

296 **2.1.2. Archive-focused initiatives**

297 Attempts at paleoclimate data standardization have a long history. For datasets derived from
298 wood archives, LinkedEarth relied on the tree-ring data standard, TRiDaS (Jansma et al., 2010),
299 which complies with established data standards such as Dublin Core (DCMI Usage Board,
300 2008). The TRiDaS project aimed at defining the properties that are used in the dendro
301 community and give them a consistent name (i.e., a controlled vocabulary) and identifying
302 whether the quantity should be mandatory and repeatable (i.e., best practices). These efforts help
303 inform the PaCTS one for wood archives, though it should be noted that tree-ring science is far
304 broader than dendroclimatology, involving applications to paleofire, landscape evolution,
305 paleoecology, art history, and archeology. Because PaCTS is focused on paleoclimate, we re-
306 used the relevant subset of the TRiDaS standard.

307
308 A discussion regarding paleoceanographic data standards was started during the Paleoclimate
309 Model Intercomparison Project (PMIP) Ocean Workshop 2013 - Understanding Changes Since
310 the Last Glacial Maximum (hereafter, PMIP LGM) in Corvallis, Oregon in December 2013.
311 Given the expertise of the working group members, the discussion focused on marine
312 sedimentary archives and was summarized into a document, which is available on the
313 LinkedEarth Platform (Kucera et al., 2013). Their recommendations served as the foundation for
314 a preliminary reporting standard for records based on marine sedimentary archives. Although the
315 group identified recommended properties to be included with marine datasets, they did not
316 propose a complete vocabulary nor a subset of required properties for acceptance in a database.

317

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

329

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

338 **2.3. Workshop on paleoclimate data standards**

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

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

348

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

363 established.

364

365 The workshop participants also identified the need to have a separate set of requirements for
366 newly generated datasets and legacy datasets, for which less metadata would likely be available.

367 In PaCTS v1.0, a legacy dataset is defined as a dataset that is not being archived by the author(s)
368 of the original study.

369 **3. Towards PaCTS**

370 **3.1. Working groups**

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

384

385 As an example of finding datasets pertinent to a scientific question, consider a study conducted

386 by a paleoceanographer who wants to characterize millennial-scale sea surface temperature
387 (SST) variability during the Holocene epoch (Khider et al, 2016). In the current research
388 ecosystem, a typical workflow would consist of querying several databases to find suitable
389 records, extract the data, consult the original publication(s) for additional metadata (e.g. author's
390 definition of 'present'), reformat the data into a coherent format for analysis, apply spectral
391 analysis to examine the frequency content of the records, perform some statistical analysis of the
392 results, and visualize them. In an ideal world, the query, preferably from a single database,
393 should (1) find records that span the Holocene, (2) find the subset of those that primarily reflect
394 SST, and (3) find the subset of that subset with a specified resolution (e.g., finer than 200 years)
395 to have at least five data points per 1,000-year cycle (a permissive assumption for this sort of
396 work). Simple though it may seem, this query requires the following (meta)data: (1) a measure of
397 age (time) and minimum and maximum values of the time series; (2) an estimate of SST, as an
398 inferred variable, and/or Mg/Ca, $U^{k'}$ ₃₇, TEX₈₆, or microfossil assemblages as measured variables
399 from which SST can be inferred; and (3) temporal resolution, calculated from the data.

400

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

405

406 A standard helps not only with the menial task of searching for records in a database. Such a
407 standard can also assist with doing the science *per se*, by ensuring that the required information
408 is present in the dataset. For instance, making a simple map of all the records in a database by

409 archive types (Figure 1a of PAGES 2k Consortium, 2017) requires each dataset to report latitude,
410 longitude, and the archive type. More complex data analysis requires more information: to
411 investigate the effect of age uncertainties (e.g. with the Bchron (Haslett and Parnell, 2008) or
412 BACON (Blaauw and Christen, 2011) packages), or to establish new depth-age models (Blois et
413 al., 2011; Giesecke et al., 2014), one needs the raw radiocarbon measurements, their
414 measurement uncertainties, and associated depth in the archive.

415 **3.2. Community surveys**

416 To decide which of the properties identified within the various WGs should be considered
417 essential, recommended, or desired, we first gathered input via the LinkedEarth platform (Figure
418 2a). As of August 1st 2018, it was home to 207 polls, with 796 votes given by 32 different users.
419 On average, each question received 3 votes, with some questions receiving no votes and others
420 as many as 27. Note that some questions were duplicated across different WGs and the final
421 count presented here takes into account all votes received on the platform. The low number of
422 votes can be partially attributed to the fact that voting was only possible after authentication onto
423 the platform, creating a barrier to widespread participation. To broaden community involvement,
424 the polls were then threaded on Twitter from the LinkedEarth account with voting allowed over a
425 seven-day period (Figure 2b). The Twitter polls increased engagement (by a factor of 3 on
426 average), and also led to discussions that were then moved to the LinkedEarth platform for
427 traceability of decisions.

428

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

432 603 questions across all working groups for which respondents were asked to determine whether
433 each property is deemed essential, recommended, or desired for new and legacy datasets, in
434 addition to open-ended questions and prompts for community feedback. The survey was more
435 comprehensive than the polls on the LinkedEarth platform or Twitter since all questions were
436 framed to allow for a response for legacy and new datasets. On the other hand, the LinkedEarth
437 platform also contains duplicate questions across various WGs (e.g., “should depth be reported
438 as essential, recommended, desired), polls aiming to define the scope of the datasets housed on
439 LinkedEarth (e.g., “should the LinkedEarth platform only contain datasets that appear in peer-
440 reviewed publications?”), and the operating definition of legacy versus new datasets that was
441 then used in the survey. Ninety-five scientists participated in the survey. Each question on the
442 survey received on average 54 answers.

443
444 Paleoclimatology is a multi-disciplinary effort where researchers typically have expertise in one
445 or more proxy systems (e.g., different observations on the same archive, similar observations on
446 different archives, or a mix of different sensors, observations and archives). Scientists are often
447 led to compare their own datasets to others obtained from proxy systems with which they are less
448 familiar. Consequently, the metadata they need tend to differ based on their level of expertise (it
449 is easier to “fill in the blanks” in one’s own area of expertise). For instance, an ice core expert
450 interested in comparing their deuterium record with a nearby record of SST would most likely
451 only require the age at each horizon and associated SST. On the other hand, an expert on
452 foraminiferal Mg/Ca-based SST reconstruction may also need information about the cleaning
453 methodology or the number of individual foraminifera in the sample. To ensure that both needs
454 were represented, respondents were encouraged to complete the entire survey, rather than focus

455 exclusively on their own areas of expertise.

456 **3.3. Survey responses**

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

468

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

475

476 For each of the properties, we identified respondents' recommendation for both new and legacy
477 datasets as the majority vote. We used mind maps to visually organize the hierarchical

478 information, keeping the relationship intact (Figures 5) and mosaic plots to display the
479 frequencies of the essential, recommended, and desired categories for each working group
480 (Figure 6). Overall, the community identified 208 properties (69% of polled properties) as
481 essential, 82 (27%) properties as recommended, and 12 (4%) as desired for new datasets. For
482 legacy datasets, fewer properties were deemed essential: 131 (44%) of polled properties versus
483 136 properties (45%) were considered recommended and 34 properties (11%) were identified as
484 desired. This difference is not unexpected and highlights the fact that legacy datasets, although
485 not as metadata-rich as new datasets, are still valuable to the community (Figure 6).

486 **4. PaCTS v1.0: Paleoclimate Community reporting Standard**

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

492 **4.1. Cross-Archive Metadata**

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

499

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

509

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

517 **4.2. Archive-specific metadata**

518 **4.2.1. Ice cores**

519 The ice core WG identified 16 properties specific to glacier ice, including information pertaining
520 to the archive, such as melt in transport, storage conditions, the observations available for the
521 archive, and the chronology. For new datasets, eight properties were deemed essential and eight

522 recommended. The number of essential properties dropped to four for legacy datasets with three
523 properties deemed recommended (Figures 5, 6 and 8).

524

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

527 **4.2.2. Lake Sediments**

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

534

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

543

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

548 **4.2.3. Marine sediments**

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

557

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

561 **4.2.4. Coral, mollusks, and other annually resolved marine records**

562 The properties for these archives were taken from the spreadsheet the MARPA group had
563 circulated online for feedback. Most of these properties were applicable to all archives reporting
564 geochemical properties and were therefore incorporated into the cross-archive WG and
565 questions. Two archive-specific properties were also identified: interpolated chronologies (i.e.,
566 distance from core top translated to time usually a calendar day for each sample then interpolated

567 to even monthly intervals) and X-ray pictures (and associated drilling path). For both new and
568 legacy datasets, the raw (distance from core top), interpolated chronologies, and X-ray pictures
569 were considered essential and recommended, respectively (Figure 5 and 6). The reporting of
570 growth increments in mollusks and corals is still an ongoing discussion within MARPA.

571 **4.2.5. Speleothems**

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

579

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

587 **4.2.6. Tree-based records**

588 The tree ring community has a long history of developing and adopting data standards; however,
589 the metadata capacity or requirements in earlier data formats (e.g., Tucson, Heidelberg,

590 Sheffield, CATRAS and Belfast amongst many others) were limited by the technology of the
591 decade in which they were created (Brewer et al. 2011). The 35 properties in the survey were
592 taken from TRiDaS (Jansma et al., 2010) and from the proposed tree-ring isotope databank
593 (Csank, 2009). TRiDaS was chosen as a starting point as it was designed as a standard to
594 represent dendrochronological data across its many subdisciplines, including dendroclimatology.
595 TRiDaS therefore includes many (optional) properties as essential or recommended that are not
596 applicable to datasets collected for paleoclimate reconstructions.

597

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

608

609 Nevertheless, a subset of the fields defined in TRiDaS were used as a starting point for
610 discussion for PaCTS v1.0. Many fields within TRiDaS are already addressed in the cross-
611 archive metadata and were disregarded, leaving only dendro-specific fields. These were then
612 supplemented by fields for tree-ring isotope data taken from the tree-ring isotope databank

613 proposed by Csank (2009). Regretfully, discussion of the suitability of these fields among the
614 dendroclimatology community has been limited and the list of initial fields was not subsequently
615 refined. The public voting process has resulted in a number of fields being marked as ‘essential’
616 that are not routinely (if ever) collected for dendroclimatological research. Furthermore, some of
617 the quantities that are being proposed are difficult to measure or know, raising the issue of
618 whether these properties are even desired. Some of the properties are a characteristic of the data
619 themselves (‘ring count’) and not metadata *per se*. These may be useful as convenience fields
620 when querying large data collections (rather than having to extract and calculate).

621
622 The confusion in the voting process could reflect confusion over whether PaCTS v1.0 is to be a
623 data standard applicable to all dendrochronological datasets or exclusively to those collected for
624 use in climate reconstructions, for which a smaller number of ‘essential’ fields would be
625 required. It could also reflect sampling bias in the voting process related to the composition of
626 the WG.

627
628 While the work described here is clearly an important step towards incorporating
629 dendroclimatological data into a universally applicable paleoclimate data standard, there remains
630 a great deal of work to be done. This work needs to begin with discussions that engage a much
631 broader cross-section of the dendroclimatological community and refined criteria in subsequent
632 surveys.

633 **4.2.7. Documentary archives**

634 Historical documents differ quite significantly from the other archive types presented in PaCTS
635 v1.0. Documentary data are extracted from written sources (books, chronicles, newspaper, etc)

636 and each of these sources in the dataset needs a reference to the publication metadata (in addition
637 to the scientific publication of the data in a journal). The raw data most comparable to
638 measurements on other archives are quotes, i.e., text strings in any language cited from the
639 source from which location, time, and event are extracted. Every single data point in the set can
640 thereby have a different location and a variety of parameters describing the event (Glaser, 1996).
641 The time step can be, but is not necessarily, periodic. The quote might contain information
642 regarding the temperature in a city, precipitation conditions, and the resulting water level in a
643 river, as well as statements concerning harvest amount and quality of a certain crop. The
644 resulting data type can be boolean (for presence/absence), integer (for indices), real numbers
645 with units for measurements, or enumerations (Riemann et al., 2016).

646

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

656 **4.3. Uncertainties**

657 The Uncertainties WG identified seven properties applicable to most records. These properties
658 fell into two broad categories concerning the uncertainty in the measured variable (analytical

659 uncertainty, number of repeat measurements, and reproducibility) and the uncertainty associated
660 with models to infer variables, including chronologies (output statistics, output ensembles along
661 with the parameters and the publication in which the model is described). For new datasets, four
662 properties (analytical uncertainty, number of repeat measurements, the publication and
663 parameters of the model) were deemed essential and the other three recommended. For legacy
664 datasets, only one was deemed essential (number of repeat measurements) while the rest were
665 recommended. This highlights the commitment of the community to better characterize
666 uncertainties in paleoclimate records and the acknowledgement that uncertainty has often been
667 ignored when reporting datasets in the past, making it difficult to include metadata for legacy
668 datasets (Figures 5, 6, and 14).

669

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

676

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

680

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

689

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

701

702 On the other hand, uncertainty associated with measurements of trace elements and stable
703 isotopes is often reported as the uncertainty of the standard or a handful of replicates that are

704 taken to represent the uncertainty for all values. The LinkedEarth Ontology (Emile-Geay et al.,
705 2019) allows for the specification of not only the values and units of the uncertainty, but also
706 how this uncertainty is estimated and the level at which it is being reported (e.g., one standard
707 error of the mean).

708 **4.4. Chronologies**

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

712
713 Chronologies are obtained using two methods: absolute and relative. Relative chronologies often
714 involve the alignment of one paleoclimate time series with another of known age. For instance,
715 benthic foraminifera stable oxygen isotope ($\delta^{18}\text{O}$) records have often been aligned to the dated
716 LR04 benthic $\delta^{18}\text{O}$ stack (Lisiecki and Raymo, 2005). For this type of chronology, the original
717 measurements (e.g., benthic foraminifera $\delta^{18}\text{O}$), the alignment target (e.g., LR04 benthic $\delta^{18}\text{O}$
718 stack), its associated reference chronology (e.g., LR04 age model) and alignment method (e.g.,
719 HMM-Match (Lin et al., 2014)) should be clearly identified (*essential*) for both new and legacy
720 datasets. We acknowledge that there is potentially more work to be done to devise a standard for
721 relative chronologies, which should include an integration framework for biostratigraphy,
722 paleomagnetism, stable isotopes chronologies, and orbitally-tuned chronologies.

723
724 Absolute chronologies are based on radiometric measurements (commonly radiocarbon, lead,

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

731
732 The radiocarbon community has a long history of standardizing the reporting of their
733 measurements. In 1977, Stuiver and Polach highlighted recommendations that have remained
734 mostly unchanged (Stuiver and Polach, 1977). For chronological studies using the Libby half-life
735 (Libby et al., 1949), Stuiver and Polach recommend reporting the $\delta^{13}\text{C}$ ratio, the conventional
736 radiocarbon age (relative to CE 1950), associated error (expressed as \pm one standard deviation),
737 the estimated reservoir correction, and (optionally) the per mil depletion or enrichment with
738 respect to 0.95 NBS Oxalic acid standard (Olson, 1970). For geochemical samples,
739 dendrochronological samples, reservoir equilibria, and diffusion models, they recommend
740 reporting the $\delta^{13}\text{C}$ ratio, percent modern, and $\delta^{14}\text{C}$ and $\Delta^{14}\text{C}$ based on the Cambridge half-life of
741 5730 years (Godwin, 1962). These guidelines were further extended to include post-bomb ^{14}C
742 data (Reimer et al., 2004) and the reporting of calibrated dates (Millard, 2014) and formed the
743 basis of the properties that were put to a vote. Given the long history of standardization, it is not
744 surprising that legacy radiocarbon datasets are also held at a stringent reporting level.

745
746 For U-Th dating, the WG recommended the use of the standard proposed by Dutton et al. (2017),
747 with most properties recognized as essential when reporting U-series dates.

748

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

756

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

763

764 Asking for a definition of “present” yielded diverse results. Sixty-eight percent of respondents
765 voted in favor of using 1950 as the present, following the radiocarbon convention, 7% voted in
766 favor as defining present as the last year in a record (with no mention of uncertainty), 12% voted
767 in favor of using 2000 as the present, while the last 13% answered “other ” This last category
768 includes the use of 1950 for radiocarbon and either something else for the other chronologies or
769 readjusting to 1950 to stay in tune with radiocarbon and the use of either 1950 or 2000 as long as
770 it is clearly defined with the data. In summary, there is a consensus that “present” should be

771 defined as an absolute date (and reported in the metadata), but it should be archive-dependent,
 772 with practitioners of U-series dating leaning towards CE 2000 and practitioners of radiocarbon
 773 dating leaning towards CE 1950.

774

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

783

784
$$\text{Time (age)} = \text{significand} \cdot 10^{\text{exponent}} \text{ years direction datum}$$

785

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

791

792 Table 1 shows how this representation would work in practice. Note that variability in the datum
 793 for rows 1 (21 ky BP, a common date for the Last Glacial Maximum) and 4 (127 ky BP, a

794 common date for Marine Isotope Stage 5e) could arise because of the date being reported from a
795 radiocarbon vs. U-series chronology, and is usually impossible to infer without clarification from
796 the original publication, or from its authors. The current proposal removes such ambiguities and
797 can accommodate both observed and simulated datasets, potentially easing the task of model-
798 data comparison if both communities start adopting it.

799

800 **5. An example: MD98-2181**

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

805

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

813

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

817 16). Since, by default, all metadata are desired, we ignored this category for the purpose of this
818 example. In terms of its cross-archive metadata, the MD98-2181 record is nearly complete, with
819 95% of the essential metadata and 78% of the recommended metadata present in the record
820 (Figure 16). The only missing component of essential metadata is the sample thickness. For the
821 recommended category, the International Geo Sample Number (IGSN) for the sample and date at
822 which the measurements were performed (i.e., analysis date) are missing. The core IGSN should
823 be assigned by the core repository directly (e.g., Bremen Core Repository, Oregon State
824 University core repository). Both analysis dates and sample thickness are metadata readily
825 available at the time of collection. Although both were collected in either a physical notebook or
826 by the instrument during analysis, they were not archived with the dataset on LinkedEarth since
827 the information was not deemed by the metadata authors as essential for reproducibility.

828

829 The paleodata for the record consists of Mg/Ca and $\delta^{18}\text{O}$ measurements on foraminifera tests
830 from sediment core subsamples. For the essential reporting of $\delta^{18}\text{O}$ on foraminifera, the MD98-
831 2181 record lacks metadata regarding the taxonomy scheme being followed and equilibrium
832 offsets. In the recommended category, only the volume of sediment analyzed is missing. For
833 Mg/Ca reporting, the contamination indicator values (Mn/Ca and Fe/Ca; Khider et al., 2014) are
834 missing from the archived record in addition to the taxonomy scheme being followed. Neither
835 were deemed useful for reproducibility by the authors of the study at the time of reporting. In
836 the recommended category, the volume of sediment analyzed and habitat depth have not been
837 reported. In both cases, the values are unknown, either because they were not measured during
838 sample preparation (sediment analyzed) or could not be accurately determined (habitat depth)
839 from previous studies in the region.

840

841 The MD98-2181 chronology was based on radiocarbon measurements. Ninety percent of the raw
842 radiocarbon dates used in Khider et al. (2014) were reported in Stott et al. (2004) and Stott et al.
843 (2007). The raw data necessary for the repeatability and replicability of the age model in Khider
844 et al. (2014) were re-reported in the later study. However, the archived record is missing
845 information about the modern fraction (F14C), the sample ID, and the matrix, which are deemed
846 essential. The archived record is also missing most of the recommended properties, only
847 reporting the reservoir age correction (ΔR), the ensemble statistics, and the ensemble age
848 models. The last two properties are essential in the context of the Khider et al. (2014) study to
849 reproduce the age-uncertain spectral analysis. The Stott et al. (2004) and Stott et al. (2007)
850 studies are also missing the essential and recommended properties with respect to reporting of
851 raw measurements.

852

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

857

858 This example highlights the difficulty of reporting all essential metadata, especially after the
859 study has been completed. We therefore present version 1.0 of PaCTS as an aspirational
860 standard, one that would theoretically ensure optimal reuse of paleoclimate datasets but is
861 difficult to observe in practice. Clearly, being aware of these requirements at the start of a study
862 would help scientists keep track of the necessary metadata and ensure that they are reported

863 when the dataset is digitally published (e.g., on WDS-Paleo or PANGAEA). We therefore
864 recommend that investigators plan ahead of time which properties they intend to report, and
865 structure their lab notebooks so this information is easier to track at the time of publication.

866 **6. Discussion**

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

875

876 The standards arose through collective discussions, both in-person and online, and via an
877 innovative social platform (Gil et al 2017). The results of this collective decision-making reveal
878 an evident desire for archiving a rich set of metadata properties, with respondents identifying
879 roughly two thirds of properties (208 out of 302) as *essential* for new datasets. Respondents also
880 recognized that legacy datasets may not be as complete, so they identified less stringent
881 requirements in order not to overlook valuable datasets. Nonetheless, respondents identified 131
882 properties as *essential* for legacy datasets, highlighting the fact that a dataset loses its usefulness
883 if too many requirements are not met. Several respondents also indicated that, while some
884 properties should theoretically be *essential* (or *recommended*), they may be hard to obtain in
885 practice and/or variable in time. These include seasonality and habitat depth of foraminifera and

886 many of the properties from TRiDaS. Furthermore, although rich metadata are always valuable,
887 these requirements should be balanced with the researcher's time. Scans of historical documents
888 or uploads of x-radiographs of archive samples would be highly valuable to the community, but
889 these activities are time-consuming and this use of time is rarely, if ever, incentivized by funding
890 agencies.

891
892 PaCTS v1.0 is also missing several proxy systems, including loess and continental records,
893 faunal and floral counts in lake sediments and does not incorporate recent standards such as the
894 one developed by Courtney Mustaphi et al. (2019) for ^{210}Pb dating. Finally, although cross-
895 pollination was encouraged, common properties were not adequately identified across WGs,
896 resulting in duplicates. This is especially apparent in the lake and marine sediment WGs.

897
898 Another salient outcome is that this first version of PaCTS can only be described as aspirational.
899 Indeed, section 5 illustrates that even in the best of circumstances (the author describing their
900 own dataset, generated less than a decade ago), the compliance rate was far from perfect. This
901 points to the need for more realistic guidelines. It is indeed apparent that many participants
902 misinterpreted what was meant by "essential." Further, the participation rate is still far below
903 what is needed for this standard to be representative of the worldwide paleoclimate community,
904 which would gain much from harmonization. How can this standard be collectively refined and
905 more broadly adopted? How should the standard, and its future versions, be implemented in
906 practice?

907 **6.1 Broadening participation**

908 The genesis of PaCTS v1.0 serves as a useful template for future efforts. As detailed in section 2,

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

924

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

932 **6.2 Roadmap to standardization**

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

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

936

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

939

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

942

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

945

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

947 **6.3 Implementing Emerging Standards**

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

952

953 On the first point, the LinkedEarth project has recently implemented a web interface to convert
954 paleoclimate datasets into the LiPD format: the lipd.net “playground”

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

972

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

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

982

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

994

995 We expect the present reporting standard to evolve to meet the needs of the paleoclimate
996 community. It is our hope that this publication will stimulate volunteers to join the effort and
997 organize discussions at all community levels; there can be no community standard without
998 community involvement. We are confident that improving paleoclimate data standards will
999 promote collaboration on international data syntheses and encourage the development of
1000 software based on the new standards. In turn, such software will reduce the time to science, by

1001 compressing the time researchers spend on the menial task of data wrangling.

1002 **Acknowledgments, Samples, and Data**

1003 Code and data to reproduce the figures of this article are available on GitHub and released on
1004 Zenodo (doi:10.5281/zenodo.3165019). Definition of properties and recommendations are
1005 summarized here: http://wiki.linked.earth/PaCTS_v1.0. This work was supported by the National
1006 Science Foundation through the EarthCube Program with grant ICER-1541029. Feedback
1007 solicitation on the standard was facilitated by the Past Global Changes (PAGES) organization.
1008 The 2016 workshop on Paleoclimate Data Standards was hosted by the World Data Service for
1009 Paleoclimatology (WDS/NOAA-Paleo), and the participation of international attendees was
1010 made possible by a PAGES travel grant. Any use of trade, firm, or product names is for
1011 descriptive purposes only and does not imply endorsement by the U.S. Government.

1012 **References**

- 1013 Atsawawaranunt, K., et al. (2018), The SISAL database: a global resource to document oxygen
1014 and carbon isotope records from speleothems, *Earth System Science Data*, 10(3), 1687-1713,
1015 doi: 10.5194/essd-10-1687-2018
- 1016 Blaauw, M., and J. A. Christen (2011), Flexible Paleoclimate Age-Depth Models using an
1017 Autoregressive Gamma Process, *Bayesian Analysis*, 6(3), 457-474, doi: doi:10.1214/11-
1018 BA618
- 1019 Blois, J. L., Williams, J. W. (Jack), Grimm, E. C., Jackson, S. T., & Graham, R. W. (2011). A
1020 methodological framework for assessing and reducing temporal uncertainty in
1021 paleovegetation mapping from late-Quaternary pollen records. *Quaternary Science Reviews*,
1022 30(15), 1926–1939. doi:10.1016/j.quascirev.2011.04.017
- 1023 Bradley, E., K. Anderson, L. de Vesine, T. Nelson, S. Soti, I. Weiss, and R. Yadav (2018),
1024 CSciBox – building age models of paleorecords, Zenodo, doi:10.5281/zenodo.1245175
- 1025 Brewer, P.W., Murphy, D. and Jansma, E., 2011. TRiCYCLE: a universal conversion tool for
1026 digital tree-ring data. *Tree-Ring Research*, 67(2), pp.135-145. DOI: 10.3959/2010-12.1
- 1027 Comas-Bru, L. and Harrison S.P. (2019), SISAL: Bringing added value to speleothem research,
1028 *Quaternary*, 2(1), 7; doi: 10.3390/quat2010007
- 1029 Courtney Mustaphi, C. J., Brahney, J., Aquino-López, M. A., Goring, S., Orton, K., Noronha, A.,
1030 et al. (2019). Guidelines for reporting and archiving 210Pb sediment chronologies to improve
1031 fidelity and extend data lifecycle. *Quaternary Geochronology*, 52, 77-87,
1032 doi:10.1016/j.quageo.2019.04.003

- 1033 Cox, S. J. D., and S. M. Richards (2015), A geologic timescale ontology and service, Earth
1034 Science Informatics, 8(1), 5-19, doi: 10.1007/s12145-014-0170-6
- 1035 Csank, A.Z. (2009), An International Tree-Ring Isotope Data bank—A proposed repository for
1036 tree-ring isotopic data, *Tree-Ring Research* 65(2), 163- 164, doi:10.3959/1536-1098-65.2.163
- 1037 Dassié, E.P. , et al. (2017), Saving our marine archives, *EOS*, 98, doi: 10.1029/2017EO068159
- 1038 Dasu, T., and T. Johnson (2003), *Exploratory Data Mining and Data Cleaning*, 203 pp., Wiley
- 1039 DCMI Usage Board, 2008. Dublin Core Metadata Initiative (DCMI) metadata terms. Retrieved
1040 on August 6th 2019 at <http://dublincore.org/documents/dcmi-terms/>
- 1041 Dutton, A., Rubin, K., McLean, N., Bowring, J., Bard, E., Edwards, R.L., Henderson, G.M.,
1042 Reid, M.R., Richards, D.A., Sims, K.W.W., Walker, J.D., Yokoyama, Y. (2017) Data
1043 reporting standards for publication of U-series data for geochronology and timescale
1044 assessment in the earth sciences. *Quat. Geochron.*, 39:142-149,
1045 doi:10.106/j.quageo.2017.03.001
- 1046 EarthCube Technology and Architecture Committee Standards Working Group: Report of the
1047 EarthCube Standards Working Group, finalized 10/05/2015. Accessed online on 08/13/2018
1048 at <https://www.earthcube.org/document/2015/ecstandardsrecs>
- 1049 Emile-Geay, J., and J. A. Eshleman (2013), Toward a semantic web of paleoclimatology,
1050 *Geochemistry, Geophysics, Geosystems*, 14(2), 457-469, doi: 10.1002/ggge.20067
- 1051 Emile-Geay, J., and N. P. McKay (2016), Paleoclimate data standards, *Past Global Change*
1052 *Magazine*, 24(1), doi: 10.22498/pages.24.1.47
- 1053 Emile-Geay, J., D. Khider, D. Garijo, N. P. McKay, Y. Gil, V. Ratnakar, and E. Bradley (2019),
1054 The Linked Earth Ontology: A Modular, Extensible Representation of Open Paleoclimate
1055 Data, Zenodo. <http://doi.org/10.5281/zenodo.2577604>
- 1056 Giesecke, T., Davis, B., Brewer, S., Finsinger, W., Wolters, S., Blaauw, M., De Beaulieu, J.-L.,
1057 Binney, H., Fyfe, R.M., Gaillard, M.-J., Gil-Romera, G., Knaap, W.O., Kuneš, P., Kühl, N.,
1058 Leeuwen, J.F.N., Leydet, M., Lotter, A.F., Ortu, E., Semmler, M., Bradshaw, R.H.W.
1059 (2014), Towards mapping the late Quaternary vegetation change of Europe. *Vegetation*
1060 *History and Archaeobotany* 23, 75–86. doi:10.1007/s00334-012-0390-y
- 1061 Gil, Y. (2013). Social Knowledge Collection. In P. Michelucci (Ed.), *Handbook of Human*
1062 *Computation* (pp. 285-296): Springer.
- 1063 Gil, Y., D. Garijo, V. Ratnakar, D. Khider, J. Emile-Geay, and N. P. McKay (2017), A
1064 Controlled Crowdsourcing Approach for Practical Ontology Extensions and Metadata
1065 Annotations, in *The Semantic Web - ISWC 2017*. ISWC 2107. Lecture Notes in Computer
1066 Science, edited by C. e. a. d'Amato, pp. 231-246, Springer, Cham.
- 1067 Glaser, R. (1996). Data and Methods of Climatological Evaluation in Historical Climatology
1068 *HSR Historical Social Research*, 21 (4) : 56-88.
- 1069 Godwin, H. (1962), Half-life of radiocarbon, *Nature*, 195, 984, doi: 10.1038/195984a0
- 1070 Gregory, J. (2003), The CF metadata standard, Retrieved from <http://cfconventions.org/Data/cf->
1071 [documents/overview/article.pdf](http://cfconventions.org/Data/cf-documents/overview/article.pdf) on May 28th 2019.
- 1072 Haslett, J., and A. Parnell (2008), A simple monotone process with application to radiocarbon-
1073 dated depth chronologies, *Journal of the Royal Statistical Society C*, 57, 399-418, doi:
1074 10.1111/j.1467-9876.2008.00623.x
- 1075 Heath, T., Bizer, C., (2011), *Linked Data: Evolving the Web into a Global Data Space* (1st
1076 edition). *Synthesis Lectures on the Semantic Web: Theory and Technology*, 1:1, 1-136.
1077 Morgan & Claypool.

- 1078 Heiser, C, McKay, N., Emile-Geay, J., Khider, D. (2018). LiPD-utilities (Version 1.0.0).
1079 Zenodo. doi:10.5281/zenodo.60813.
- 1080 Hendy, C. H. (1971), The isotopic geochemistry of speleothems-I: The calculation of the effects
1081 of different modes of formation on the isotopic composition of speleothems and their
1082 applicability as paleoclimate indicators, *Geochimica and Cosmochimica Acta*, 35, 801-824
- 1083 Jansma, E., P. W. Brewer, and I. Zandhuis (2010), TRiDaS 1.1: The tree-ring data standard,
1084 *Dendrochronologia*, 28(2), 99-130, doi: 10.1016/j.dendro.2009.06.009
- 1085 Kaufman, D.S., PAGES 2k Special Issue Editorial Team, 2018. Technical Note: Open-paleo-
1086 data implementation pilot – The PAGES 2k special issue. *Climate of the Past* 14, 593-600.
1087 doi: 10.5194/cp-14-593-2018
- 1088 Khider, D., C. S. Jackson, and L. D. Stott (2014), Assessing millennial-scale variability during
1089 the Holocene: A perspective from the western tropical Pacific, *Paleoceanography*, 29(3),
1090 143-159, doi: 10.1002/2013pa002534
- 1091 Khider, D., Emile-Geay, J., McKay, N. P., Jackson, C., & Rouston, C. (2016). Testing the
1092 millennial-scale Holocene solar-climate connection in the Indo-Pacific Warm Pool. Paper
1093 presented at the American Geophysical Union Fall Meeting, San Francisco, CA.
- 1094 Khider, D., F. Zhu, J. Hu, and J. Emile-Geay (2018a), LinkedEarth/Pyleoclim util: Pyleoclim
1095 release v0.4.0, Zenodo, doi:10.5281/zenodo.1205662
- 1096 Khider, D., and D. Garijo (2018b), LinkedEarth Queries, edited, Zenodo,
1097 doi:10.5281/zenodo.1160672
- 1098 Krötzsch, M., and D. Vrandečić (2011), Semantic MediaWiki. *Foundations for the Web of*
1099 *Information and Services – A Review of 20 Years of Semantic Web Research*, pp. 311–326.
1100 Springer
- 1101 Kucera, M., D. Khider, and L. Lisiecki (2013), Reporting standards for
1102 Paleoceanographic/Paleoclimate data. Retrieved online from
1103 [http://wiki.linked.earth/wiki/images/d/d4/Reporting_Standards_for_Paleoceanographic_PMI](http://wiki.linked.earth/wiki/images/d/d4/Reporting_Standards_for_Paleoceanographic_PMI_P3_Dec2013.docx)
1104 [P3_Dec2013.docx](http://wiki.linked.earth/wiki/images/d/d4/Reporting_Standards_for_Paleoceanographic_PMI_P3_Dec2013.docx) on May 28th 2019.
- 1105 Libby, W. F., E. C. Anderson, and J. R. Arnold (1949), Age determination by radiocarbon
1106 content: world-wide assay of natural radiocarbon, *Science*, 109(2827), 227-228, doi:
1107 10.1126/science.109.2827.227
- 1108 Lin, L., D. Khider, L. E. Lisiecki, and C. E. Lawrence (2014), Probabilistic sequence alignment
1109 of stratigraphic records, *Paleoceanography*, 29(976-989), 976-989, doi:
1110 10.1002/2014PA002713
- 1111 Lisiecki, L. E., and M. E. Raymo (2005), A Pliocene-Pleistocene stack of 57 globally distributed
1112 benthic $\delta^{18}\text{O}$ records, *Paleoceanography*, 20(PA1003), doi: 10.1029/2004PA001071
- 1113 McKay, N. P., and J. Emile-Geay (2016), Technical Note: The Linked Paleo Data framework – a
1114 common tongue for paleoclimatology, *Climate of the Past*, 12, 1093-1100, doi: 10.5194/cp-
1115 12-1093-2016
- 1116 McKay, N., J. Emile-Geay, C. Heiser, and D. Khider (2018), GeoChronR,
1117 doi:10.5281/zenodo.60812
- 1118 Masson-Delmotte, V., M. Schulz, A. Abe-Ouchi, J. Beer, A. Ganopolski, J. G. Rouco, E. Jansen,
1119 K. Lambeck, J. Luterbacher, T. Naish, T. Osborn, B. Otto-Bliesner, T. Quinn, R. Ramesh, M.
1120 Rojas, X. Shao, and A. Timmermann (2013), Information from paleoclimate archives, in
1121 *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the*
1122 *Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T.
1123 Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex,

- 1124 and P. Midgley, chap. 5, Cambridge University Press, Cambridge, United Kingdom and New
 1125 York, NY, USA.
- 1126 Millard, A. R. (2014). Conventions for reporting radiocarbon determinations. *Radiocarbon*,
 1127 56(2), 555-559. doi:10.2458/56.17455
- 1128 National Oceanographic and Atmospheric Administration. (2018) PaST (Paleoenvironmental
 1129 Standard Terms) Thesaurus. Retrieved from [https://www.ncdc.noaa.gov/data-](https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/past-thesaurus)
 1130 [access/paleoclimatology-data/past-thesaurus](https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/past-thesaurus) on May 28th 2019.
- 1131 Olsson, I. U. (1970), The use of Oxalic acid as a standard, in Radiocarbon variations and
 1132 absolute chronology, Nobel symposium, 12th Proc, edited by O. I.U., p. 17, John Wiley &
 1133 Sons, New York.
- 1134 PAGES2k Consortium (2017), A global multiproxy database for temperature reconstructions of
 1135 the Common Era, *Sci Data*, 4, 170088, doi: 10.1038/sdata.2017.88
- 1136 Reimer, P. J., T. A. Brown, and R. W. Reimer (2004), Discussion: Reporting and calibration of
 1137 post-bomb 14C data, *Radiocarbon*, 46, 1299-1304. doi: 10.1017/S0033822200033154
- 1138 Riemann, D., Glaser, R., Kahle, M., Vogt, S. (2016). The CRE tambora.org – new data and tools
 1139 for collaborative research in climate and environmental history. *Geoscience Data Journal*
 1140 2(2):63-77. DOI:10.1002/gdj3.30
- 1141 Stall, S., E. Robinson, L. Wyborn, L. R. Yarmey, M. A. Parsons, K. Lehnert, B. Cutcher-
 1142 Gershenfeld, B. Nosek, and B. Hanson (2017), Enabling FAIR data across the Earth and
 1143 space sciences, *EOS*, 98, doi: 10.1029/2017EO088425
- 1144 Stott, L., Cannariato, K., Thunell, R., Haug, G. H., Koutavas, A., & Lund, S. (2004). Decline of
 1145 surface temperature and salinity in the western tropical Pacific Ocean in the Holocene epoch.
 1146 *Nature*, 431, 56-59. doi:10.1038/nature02903
- 1147 Stott, L., Timmerman, A., & Thunell, R. (2007). Southern Hemisphere and Deep-Sea Warming
 1148 led to deglacial atmospheric CO₂ rise and tropical warming. *Science*, 318, 435-438.
 1149 doi:10.1126/science.1143791
- 1150 Stuiver, M., and H. A. Polach (1977), Discussion: Reporting of 14C Data, *Radiocarbon*, 19(3),
 1151 355-363, doi: 10.1017/S0033822200003672
- 1152 Unidata, (2019): Network Common Data Form version 4.7.0 [software]. Boulder, CO:
 1153 UCAR/Unidata. doi:10.5065/D6H70CW6
- 1154 Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al.
 1155 (2016). The FAIR Guiding Principles for scientific data management and stewardship.
 1156 *Sci Data*, 3, 160018. doi:10.1038/sdata.2016.18
- 1157 Williams, J.W., Newton, A.J., Kaufman, D.S., von Gunten, L. (eds) (2018) Building and
 1158 Harnessing open PaleoData, *Past Global Changes Magazine*, 26(2), 45-96,
 1159 doi:10.22498/pages.26.2
- 1160 Wolff, E. W. (2007), When is the “present”?, *Quaternary Science Reviews*, 26(25-28), 3023-
 1161 3024, doi: 10.1016/j.quascirev.2007.10.008.
- 1162 W3C OWL Working Group. (2012), OWL 2 Web Ontology Language Document Overview
 1163 (Second Edition), Retrieved online on August 6th 2019 at [https://www.w3.org/TR/owl2-](https://www.w3.org/TR/owl2-overview/)
 1164 [overview/](https://www.w3.org/TR/owl2-overview/)

1165 **Figure 1.** Timeline of the community elicitation for best practices in paleoclimate data reporting.
 1166 The Workshop on Paleoclimate Data Standard marks the official beginning of the endeavor.

1167 PaCTS collects responses from the LinkedEarth platform, Twitter polls, and survey up to
1168 November 2017.

1169 **Figure 2.** Example of polls on a. the LinkedEarth platform and b. Twitter (@Linked_Earth)

1170 **Figure 3.** Example of a survey question for a new dataset. The histogram represents the number
1171 of votes on each platform (orange: LinkedEarth, purple: Twitter, and green: Google survey). The
1172 pie chart represents the fraction of the votes for essential (green), recommended (pink), and
1173 desired (blue).

1174 **Figure 4.** Same as Figure 3 for a legacy dataset.

1175 **Figure 5.** Mind map of the various properties identified by the WGs and associated vote. Colors
1176 represent the different WGs. Parentheses indicate a different reporting standard for legacy
1177 datasets when different from new datasets. Available online at:
1178 <https://coggle.it/diagram/WqMd49MJtB8DbqfH/t/community-standards-for-paleoclimate-data-and-metadata>
1179 and-metadata.

1180 **Figure 6.** Mosaic plots for a. new datasets and b. legacy datasets showing the number of
1181 essential, recommended, and desired metadata for the various WGs. The height of the bar
1182 represents the fraction of total occurrences for essential (e), recommended (r), and desired (d)
1183 votes, while the width of the bar represents the number of properties voted on in each WG.

1184 **Figure 7.** Mind map of the various properties identified by the cross-archive WG and associated
1185 vote. Color is the same as in Figure 5. Parentheses indicate recommendations for legacy datasets
1186 when different from new datasets. Available online at:
1187 <https://coggle.it/diagram/W4W9podcxp86PPvf/t/cross-archive-metadata>

1188 **Figure 8.** Mind map of the various properties identified by the ice core archives WG and
1189 associated vote. Color is the same as in Figure 5. Parentheses indicate recommendations for
1190 legacy datasets when different from new datasets. Available online at:
1191 <https://coggle.it/diagram/W4XNNeGhIngfjHzB/t/historical-documents>

1192 **Figure 9.** Mind map of the various properties identified by the lake sediments archives WG and
1193 associated vote. Color is the same as in Figure 5. Parentheses indicate recommendations for
1194 legacy datasets when different from new datasets. Available online at:
1195 <https://coggle.it/diagram/W4h9m-GhIjbm3yX/t/lake-sediments>

1196 **Figure 10.** Mind map of the various properties identified by the marine sediments archives WG
1197 and associated vote. Color is the same as in Figure 5. Parentheses indicate recommendations for
1198 legacy datasets when different from new datasets. Available online at:
1199 <https://coggle.it/diagram/W4iIkodcxlDKTK6v/t/marine-sediments>

1200 **Figure 11.** Mind map of the various properties identified by the speleothem archives WG and
1201 associated vote. Color is the same as in Figure 5. Parentheses indicate recommendations for
1202 legacy datasets when different from new datasets. Available online at:
1203 <https://coggle.it/diagram/W4gwj-GhI14VmfYP/t/speleothem>

1204 **Figure 12.** Mind map of the various properties identified by tree-based archives WG and
 1205 associated vote. Color is the same as in Figure 5. Parentheses indicate recommendations for
 1206 legacy datasets when different from new datasets. Available online at:
 1207 <https://coggle.it/diagram/W4huaYdcxhdzTB9z/t/trees>

1208 **Figure 13.** Mind map of the various properties identified by the documentary archives WG and
 1209 associated vote. Color is the same as in Figure 5. Parentheses indicate recommendations for
 1210 legacy datasets when different from new datasets. Available online at:
 1211 <https://coggle.it/diagram/W4XNNeGhIngfjHzB/t/historical-documents>

1212 **Figure 14.** Mind map of the various properties identified by the uncertainties WG and associated
 1213 vote. Color is the same as in Figure 5. Parentheses indicate recommendations for legacy datasets
 1214 when different from new datasets. Available online at:
 1215 <https://coggle.it/diagram/W4gttodcxjfvSst0/t/uncertainties>

1216 **Figure 15.** Mind map of the various properties identified by the chronologies WG and associated
 1217 vote. Color is the same as in Figure 5. Parentheses indicate recommendations for legacy datasets
 1218 when different from new datasets. Available online at:
 1219 <https://coggle.it/diagram/W4hzXeGhIi5Fm0q7/t/chronologies>

1220 **Figure 16.** Radar plot showing the completeness of the metadata reporting for core MD98-2181
 1221 (Khider et al., 2014) for properties considered a. essential and b. recommended in the current
 1222 study. The axis refers to the working group standards recommendation applicable to the record.

1223

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

1224 **Table 1.** Illustration of our proposed time representation with four time points. The first column
 1225 gives examples of reported age/year in a paleoclimate paper while the last four columns show an
 1226 implementation of the representation proposed here.

Figure 1.

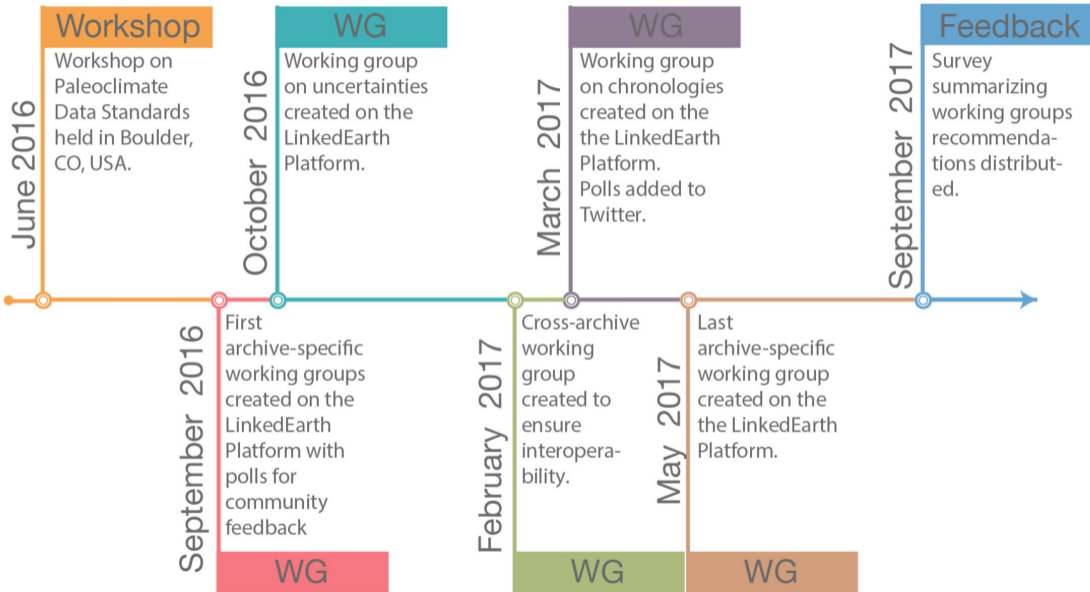


Figure 2.

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?

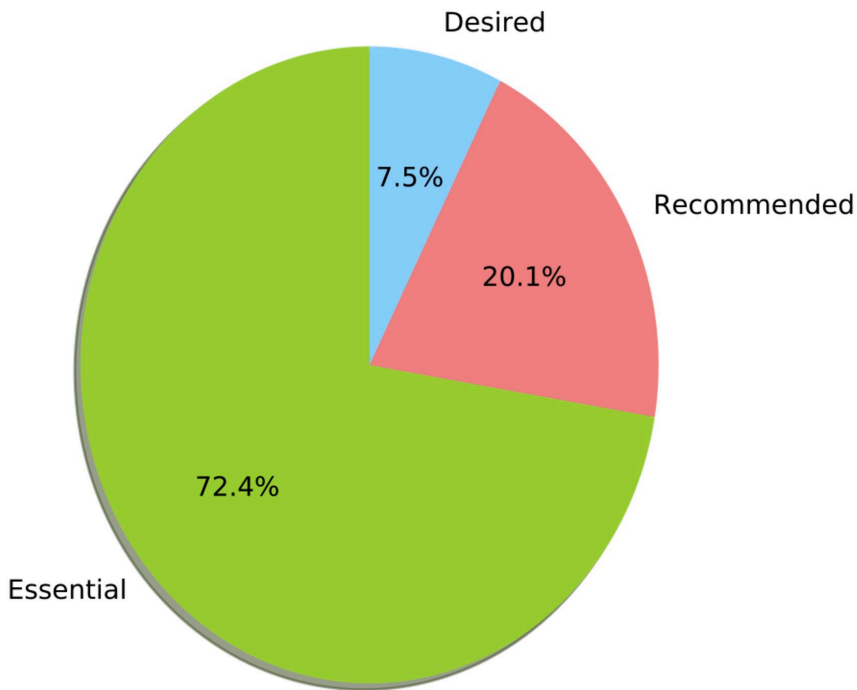
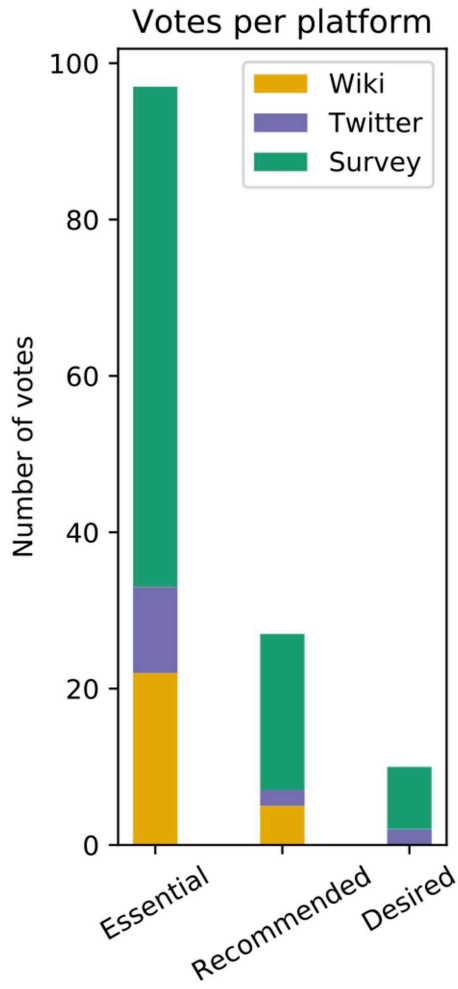


Figure 4.

For legacy datasets, should the depth/distance/position in the archive be considered essential, recommended, or desired?

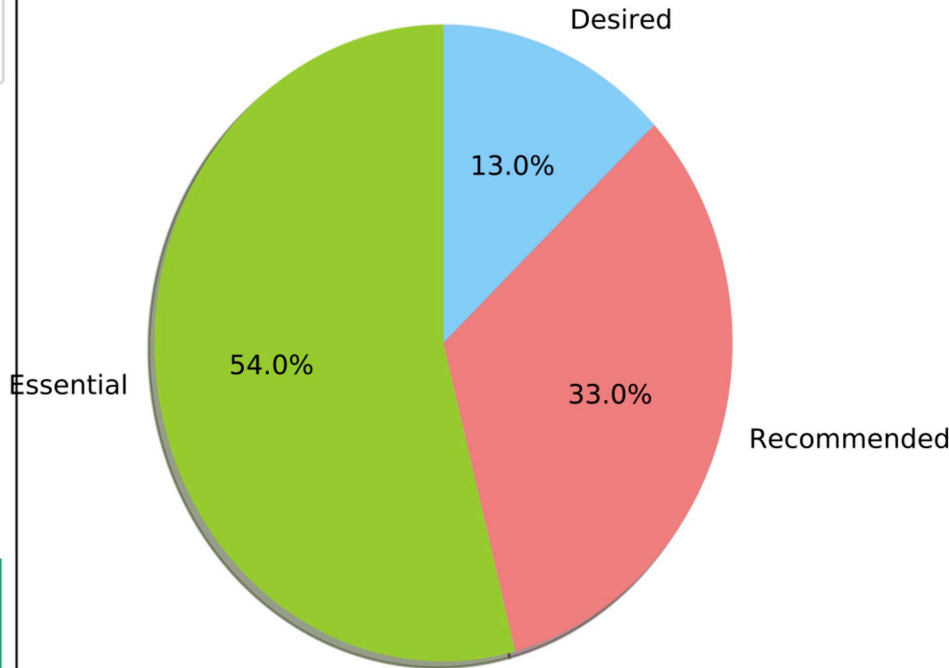
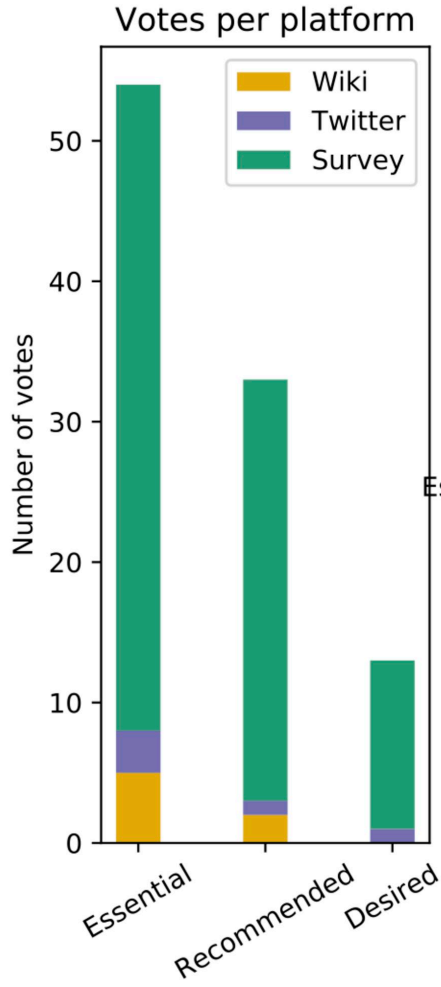


Figure 5.

Community Standards for Paleoclimate Data and Metadata

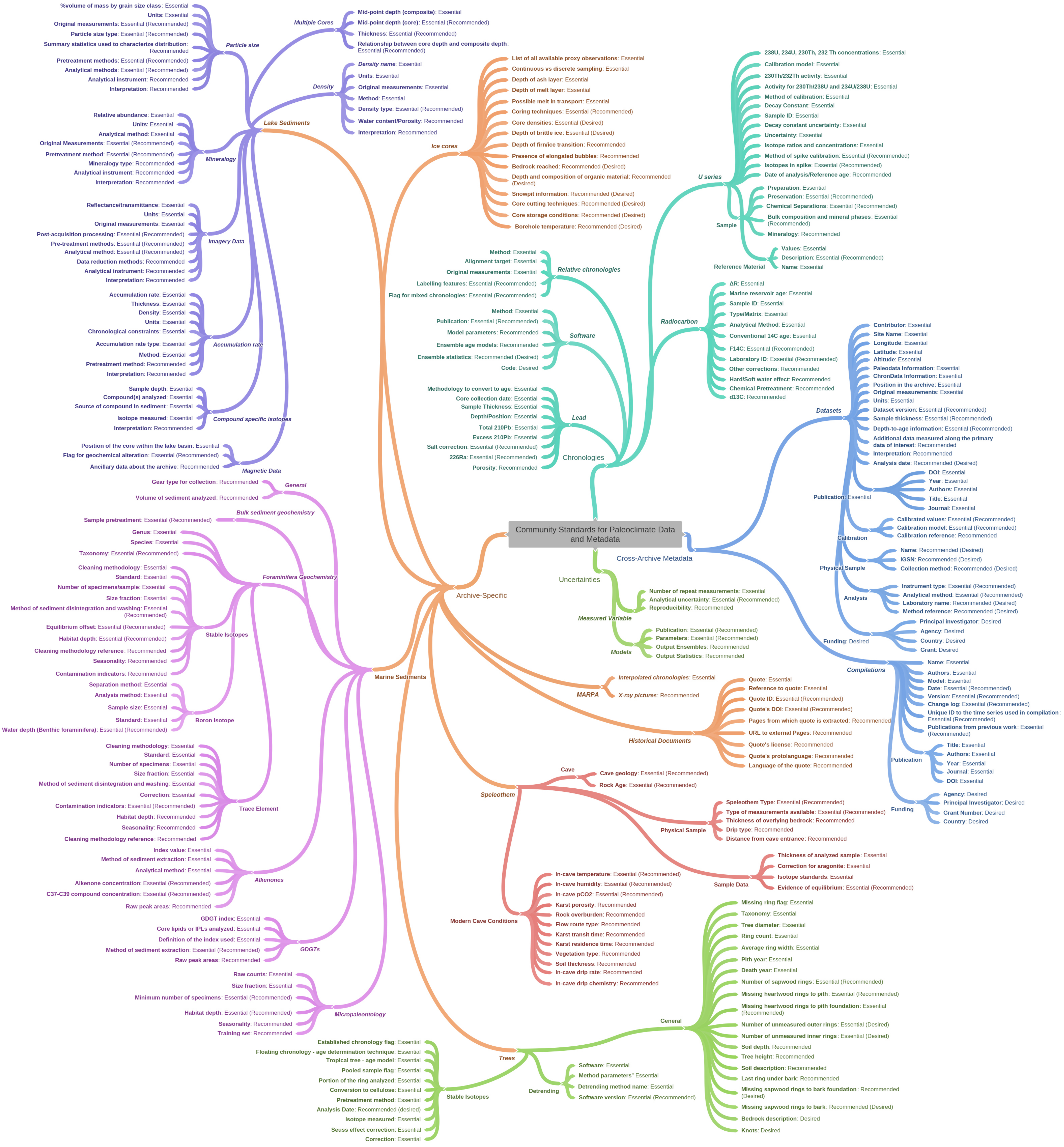
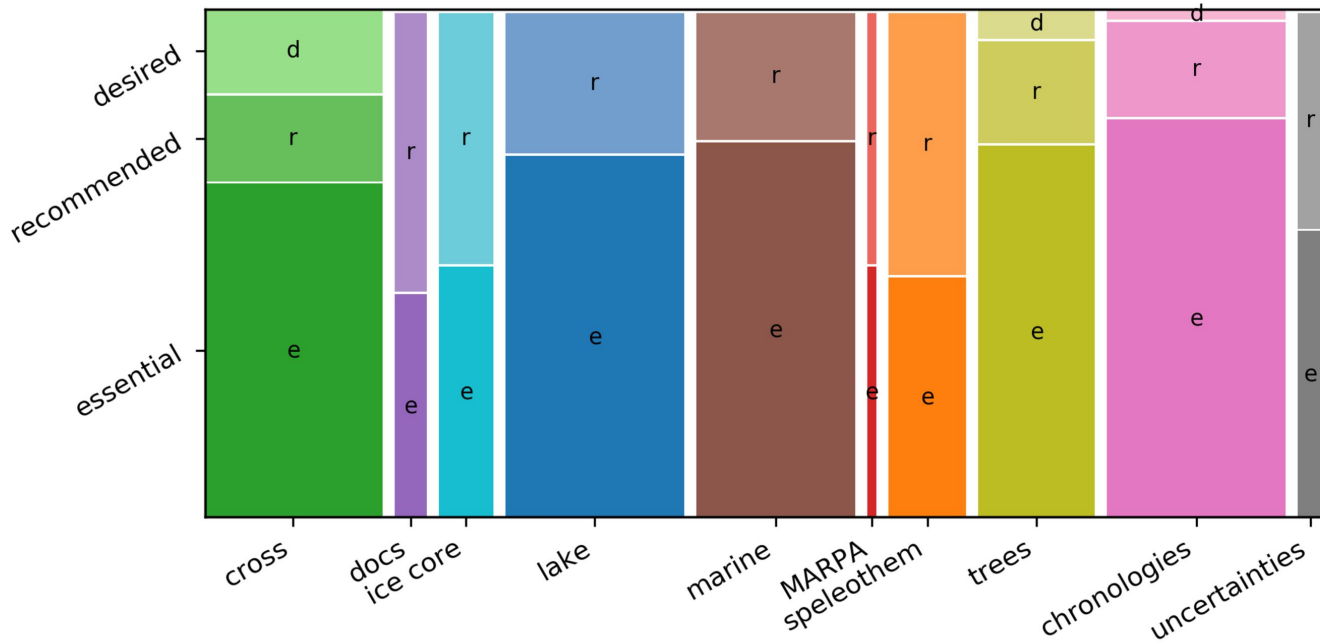


Figure 6.

a. Recommendation for new datasets



b. Recommendation for legacy datasets

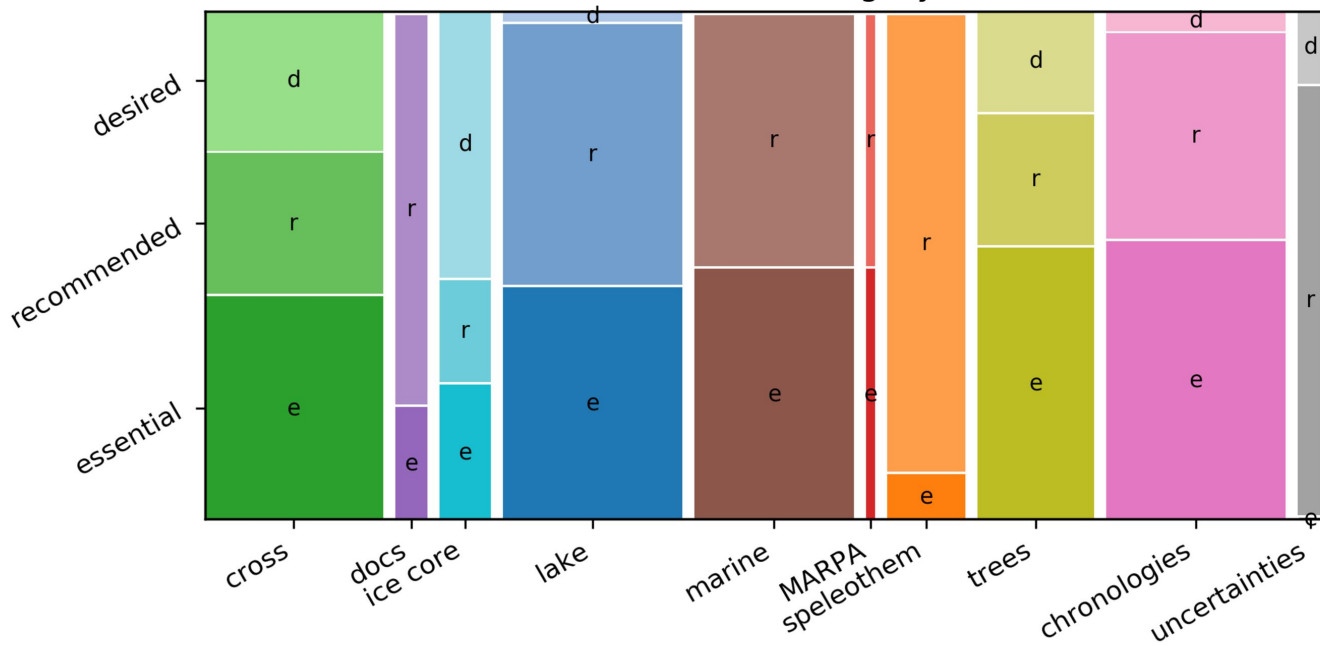


Figure 7.

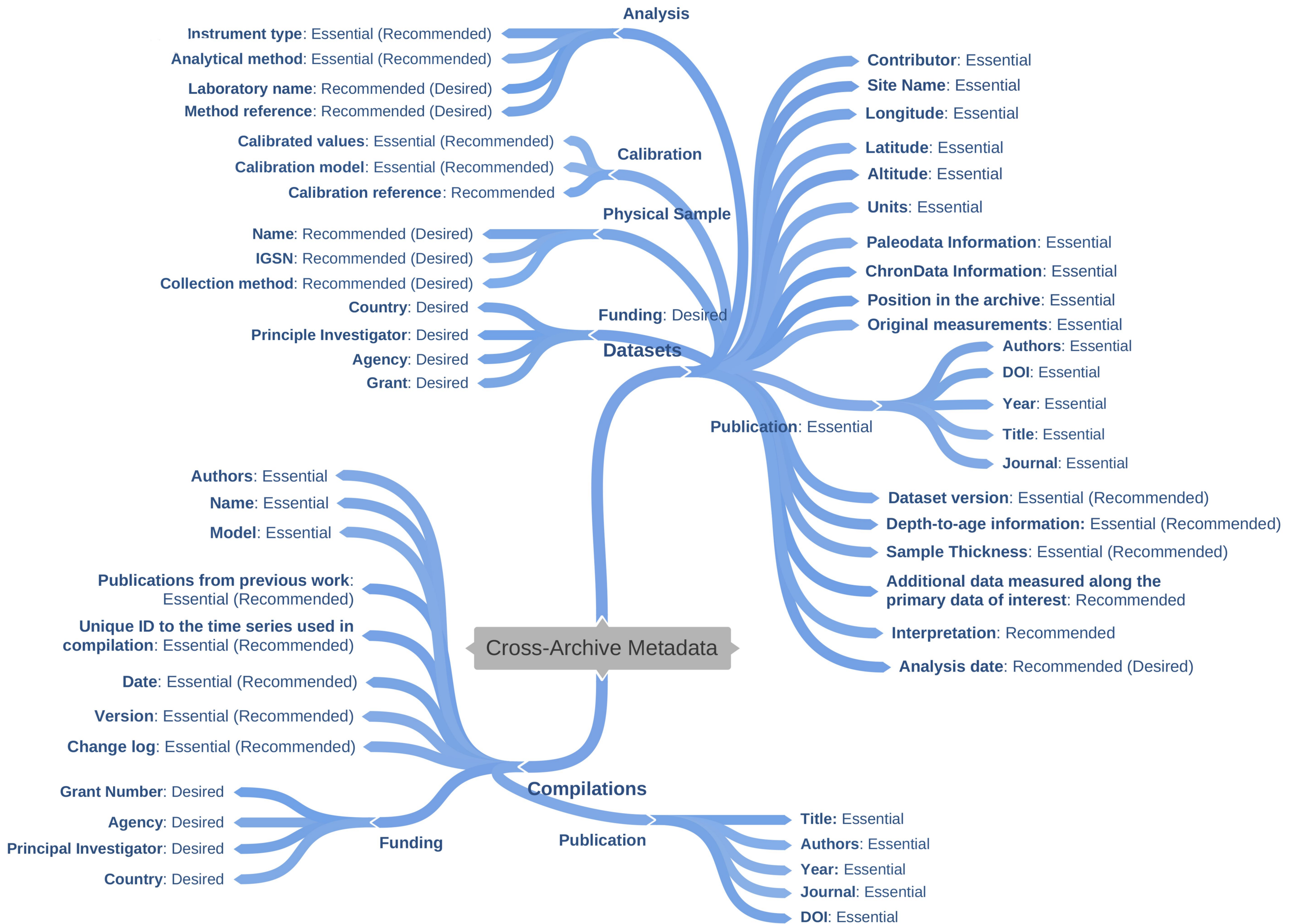


Figure 8.

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.

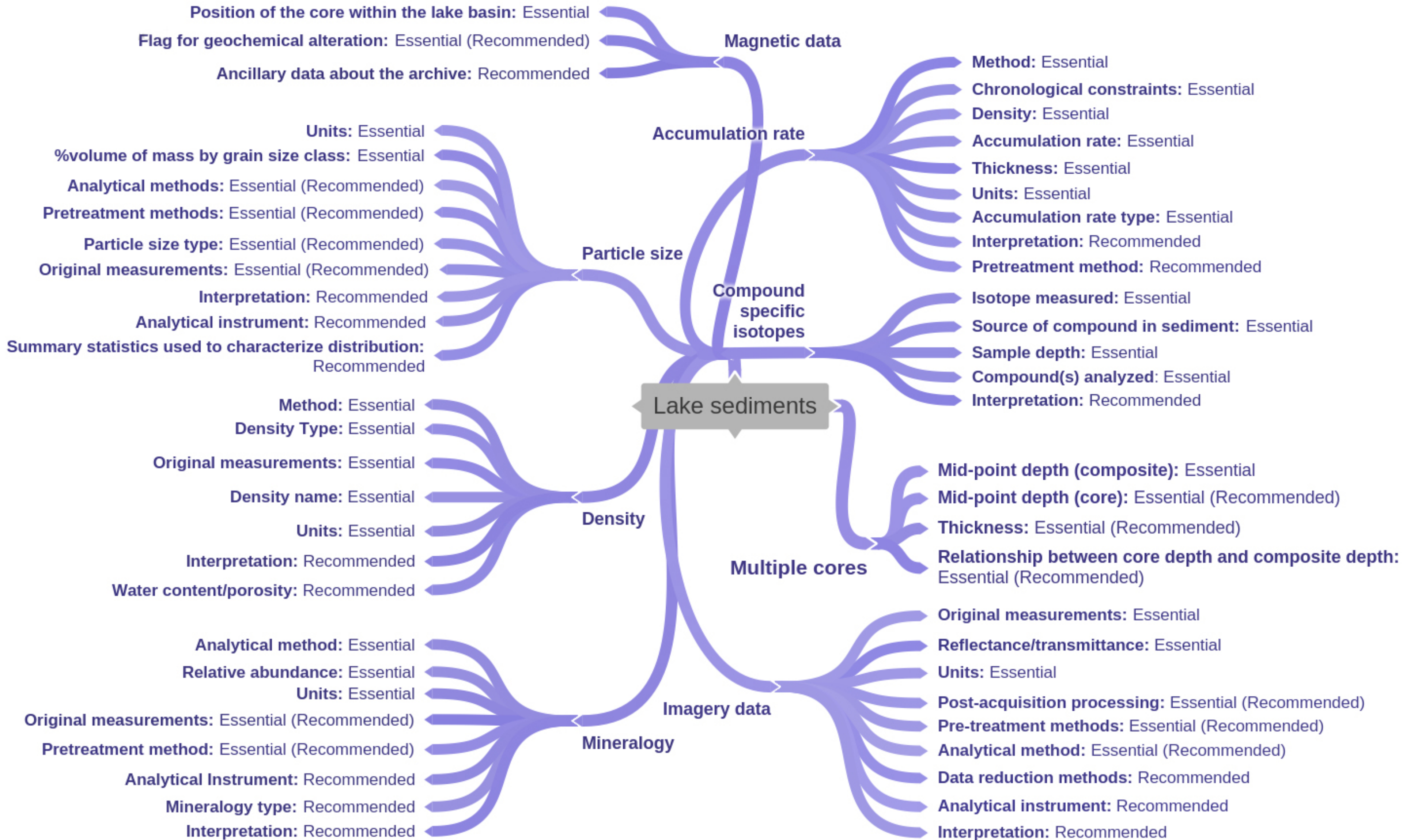


Figure 10.

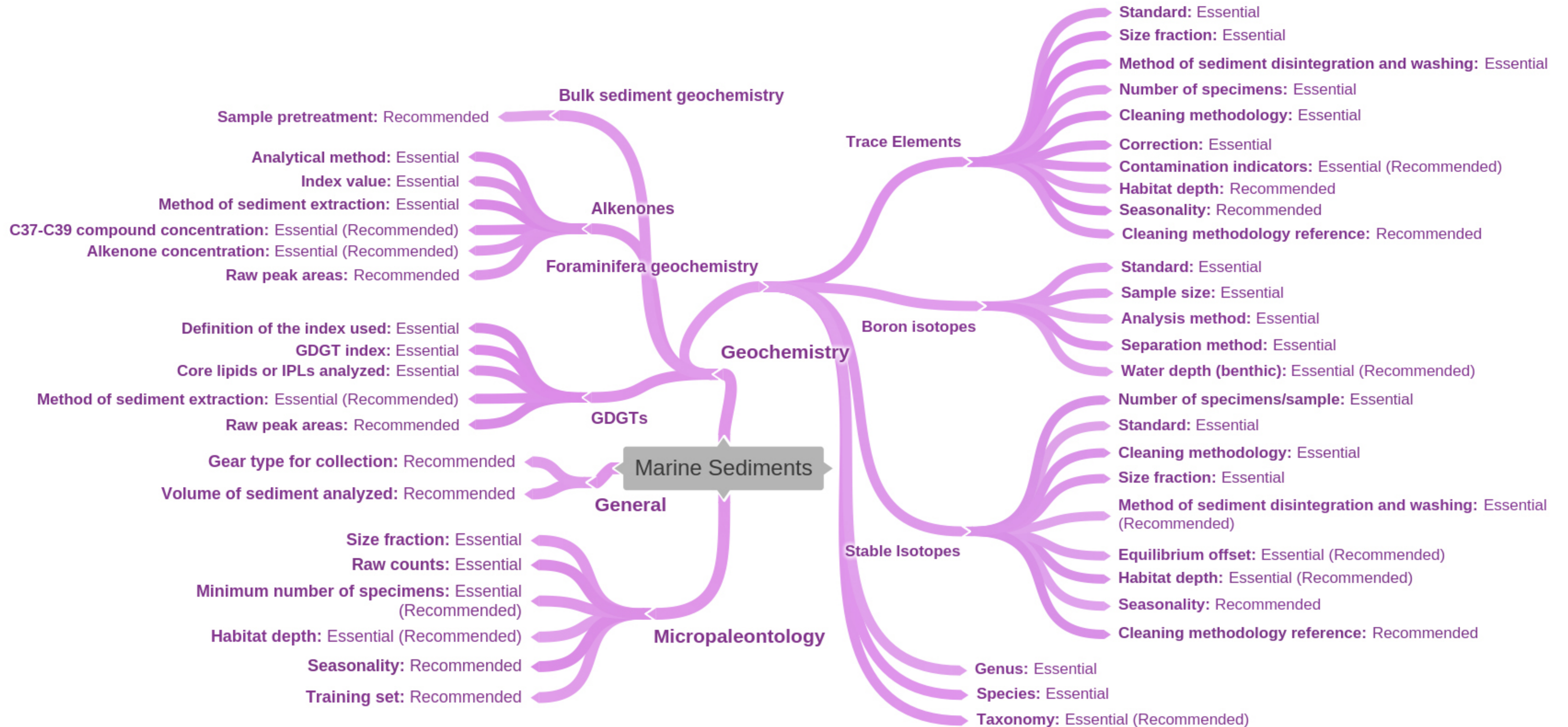


Figure 11.

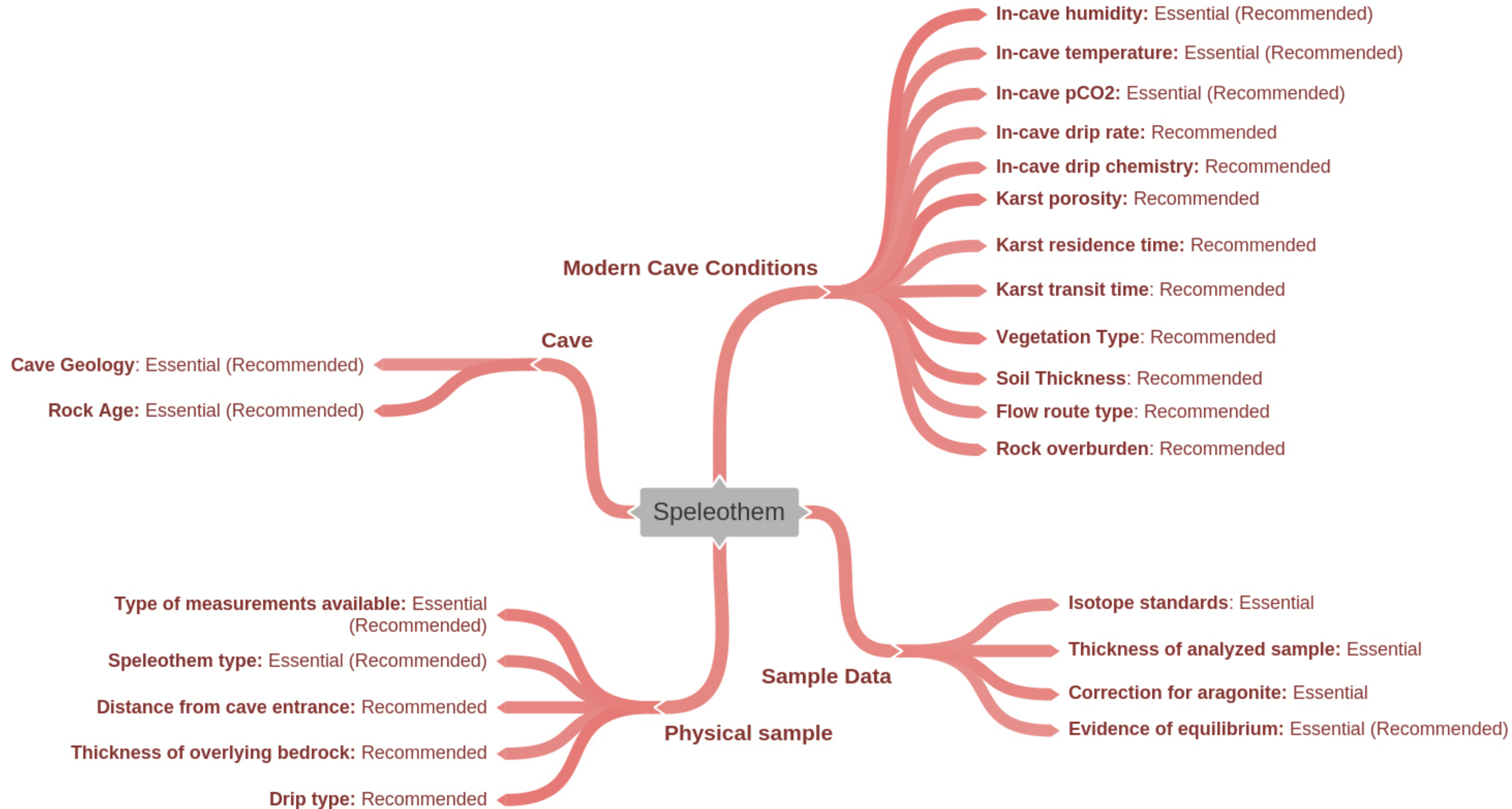


Figure 12.

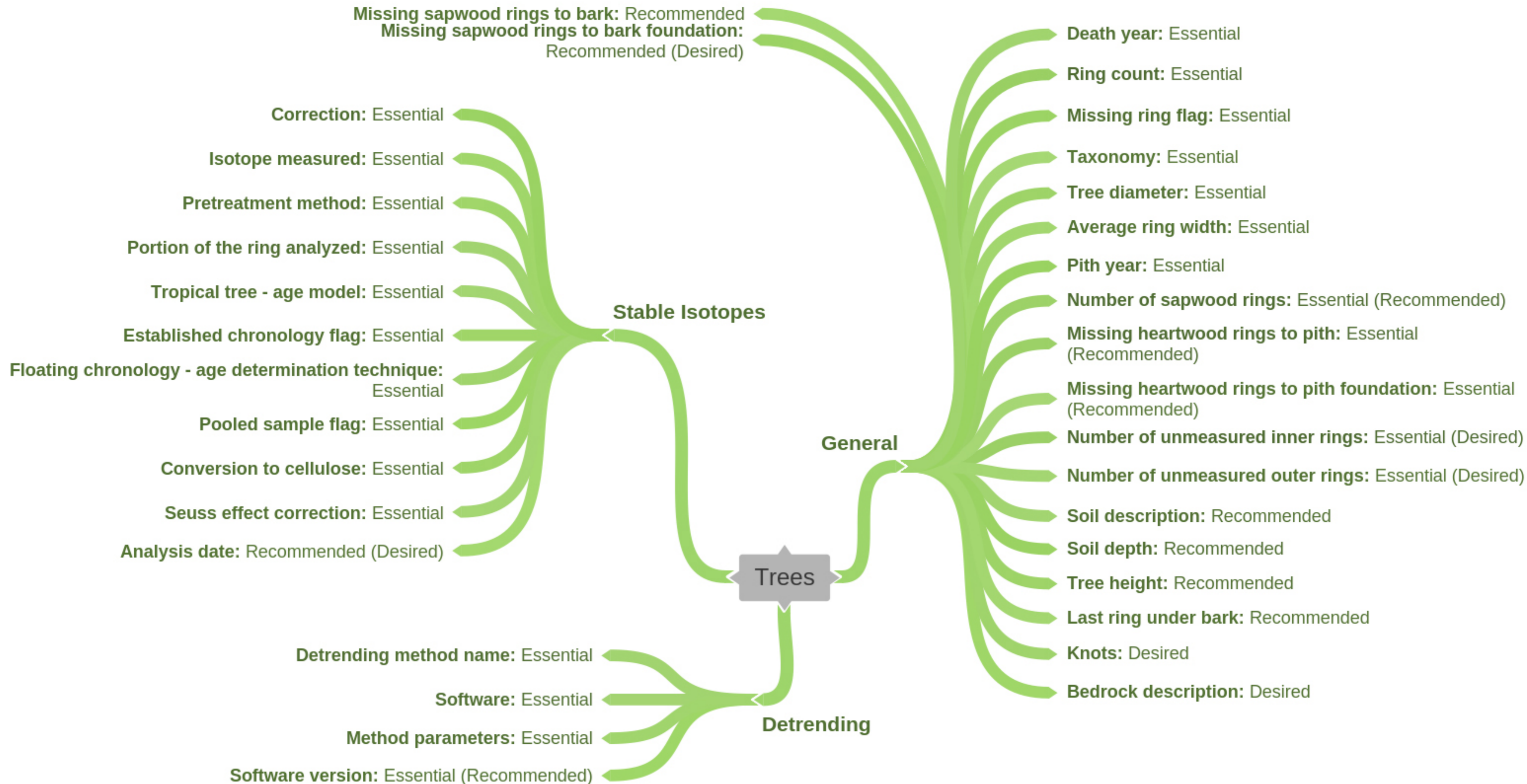


Figure 13.

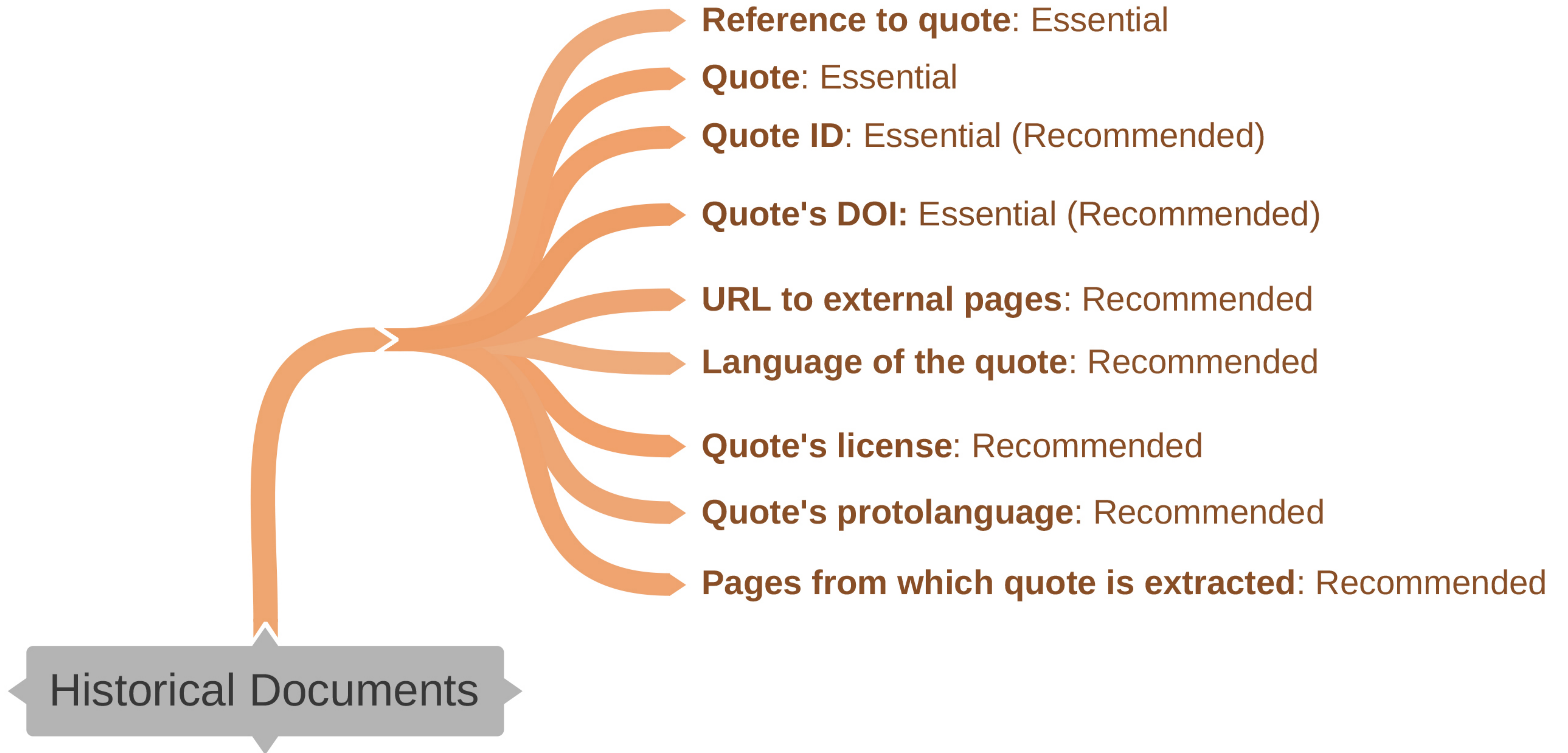


Figure 14.

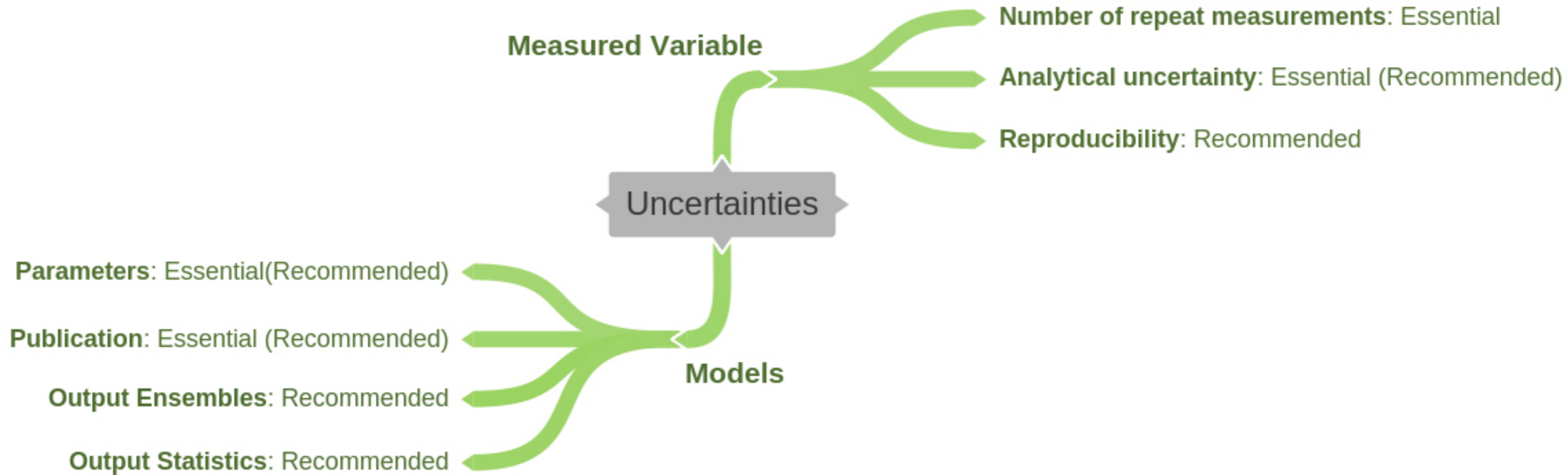


Figure 15.

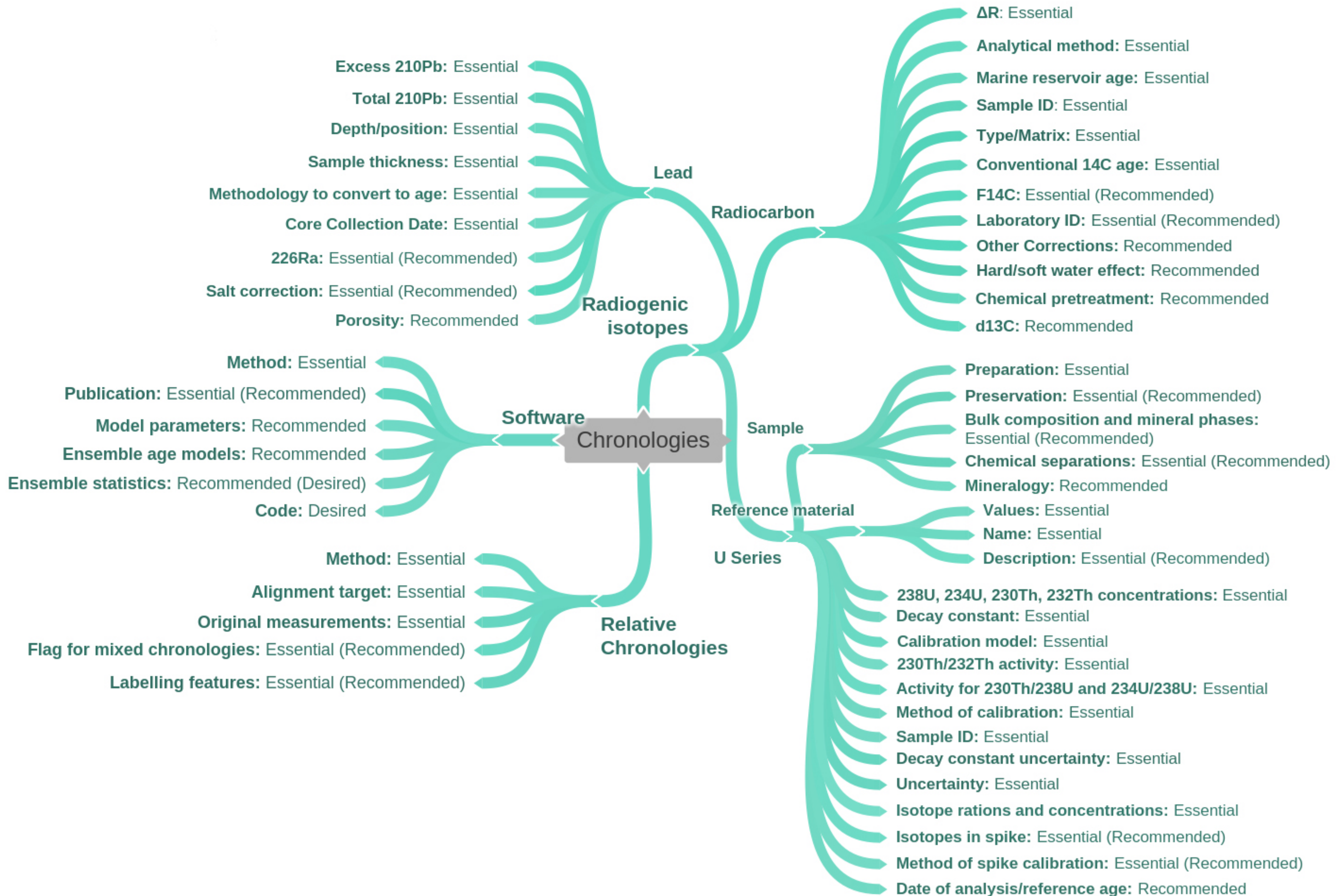
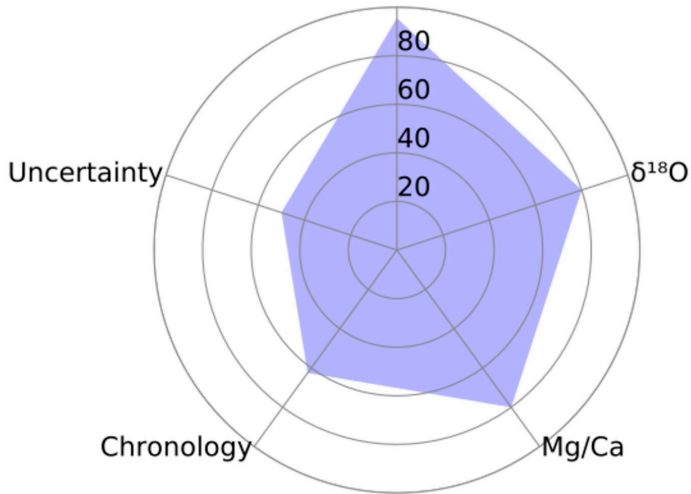


Figure 16.

a. Essential

Cross-Archive Metadata



b. Recommended

Cross-Archive Metadata

