The Iso2k database: a global compilation of paleo-$\delta^{18}O$ and $\delta^{2}H$ records to aid understanding of Common Era climate


1Department of Earth and Planetary Sciences, Washington University, Saint Louis, Missouri 63108, USA
2School of Earth and Sustainability, Northern Arizona University, Flagstaff, Arizona 86011, USA
3Institute of Ecology and Geography, Siberian Federal University, Krasnoyarsk, 660041, Russian Federation
4Department of Forest Dynamics, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, 8903, Switzerland
5School of Archaeology, Geography & Environmental Sciences, Russell Building, University of Reading, Whiteknights, Reading, Berkshire, RG6 6DR, United Kingdom
6EPOC Laboratory, University of Bordeaux, Bordeaux, 33615, France
7Department of Geography and Anthropology, Coastal Studies Institute, Louisiana State University, Baton Rouge, Louisiana 70803, USA
8NSSTL Environment, ANSTO, Sydney, NSW 2234, Australia
9School of Geography, University of Nottingham, Nottingham, NG7 2RD, UK
10MARUM Center for Marine Environmental Sciences, Bremen University, 28359 Bremen, Germany
11Aix Marseille University, CNRS, IRD, INRAE, Coll France, CEREGE, Aix-en-Provence, 13545, France
12Earth and Climate Science, Indian Institute of Science Education and Research, Pune, Maharashtra, 411008, India
13Department of Environmental Chemistry, Spanish Council for Scientific Research (CSIC), Institute of Environmental Assessment and Water Research (IDAEA), Barcelona, 08034, Spain
14Polar Terrestrial Environmental Systems and PALICE Helmholtz Young Investigator Group, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, 14473 Potsdam, Germany
15L-IPSL, CEA-CNRS-UVSQ-Université Paris Saclay, Laboratoire des Sciences du Climat et de L’Environnement, Gif-sur-Yvette, 91191, France
16Institute for Geophysics, University of Texas at Austin, Austin, Texas 78758, USA
17School of Earth and Atmospheric Science, Georgia Institute of Technology, Atlanta, Georgia 30332, USA
18Department of Geology, University at Buffalo, Buffalo, New York 14260, USA
19Department of Geosciences, University of Arizona, Tucson, Arizona 85719, USA
20Department of Earth Sciences, The University of Adelaide, Adelaide, SA 5005, Australia
21Research School of Earth Sciences and Centre of Excellence for Climate Extremes, Australian National University, Canberra, ACT 2601, Australia
22Department of Earth, Ocean, and Atmospheric Sciences, Florida State University, Tallahassee, Florida 32306, USA

Published by Copernicus Publications.
Abstract. Reconstructions of global hydroclimate during the Common Era (CE; the past ∼ 2000 years) are important for providing context for current and future global environmental change. Stable isotope ratios in water are quantitative indicators of hydroclimate on regional to global scales, and these signals are encoded in a wide range of natural geologic archives. Here we present the Iso2k database, a global compilation of previously published datasets from a variety of natural archives that record the stable oxygen (\(\delta^{18}O\)) or hydrogen (\(\delta^2H\)) isotopic compositions of environmental waters, which reflect hydroclimate changes over the CE. The Iso2k database contains 759 isotope records from the terrestrial and marine realms, including glacier and ground ice (210); speleothems (68); corals, sclerosponges, and mollusks (143); wood (81); lake sediments and other terrestrial sediments (e.g., loess) (158); and marine sediments (99). Individual datasets have temporal resolutions ranging from sub-annual to centennial and include chronological data where available. A fundamental feature of the database is its comprehensive metadata, which will assist both experts and nonexperts in the interpretation of each record and in data synthesis. Key metadata fields have standardized vocabularies to facilitate comparisons across diverse archives and with climate-model-simulated fields. This is the first global-scale collection of water isotope proxy records from multiple types of geological and biological archives. It is suitable for evaluating hydroclimate processes through time and space using large-scale synthesis, model–data intercomparison and (paleo)data assimilation. The Iso2k database is available for download at https://doi.org/10.25921/57j8-vs18 (Konecky and McKay, 2020) and is also accessible via the NOAA/WDS Paleo Data landing page: https://www.ncdc.noaa.gov/paleo/study/29593 (last access: 30 July 2020).

1 Introduction

1.1 Progress and challenges in the synthesis of Common Era hydroclimate

The past ∼ 2000 years, otherwise known as the Common Era (CE), are an important research target for contextualizing modern climate change. Decades of paleoclimate research have yielded numerous records spanning all or part of this time period, making it sufficiently data-rich to assess the range of natural (internal and forced) climate variability prior to the Industrial Revolution. These records are also used in conjunction with climate model simulations to detect and attribute anthropogenic climate change. Over the past several years, large-scale data synthesis efforts within the international paleoclimate community have produced important constraints on regional to global surface air and ocean temperature patterns during the CE (McGregor et al., 2015; McKay and Kaufman, 2014; PAGES 2k Consortium, 2013, 2017, 2019; Tierney et al., 2015). However, progress on the synthesis of hydroclimate patterns has been limited (PAGES Hydro2k Consortium, 2017), despite the societal relevance of the changing water cycle (e.g., Kelley et al., 2015). The water cycle is a far more complex target than surface air and ocean temperature, and different proxy systems track different aspects of the water cycle in different ways (PAGES Hydro2k Consortium, 2017). For example, annual precipitation amount at any given location on the Earth’s surface is gov-
erned not just by atmospheric processes that deliver moisture to the region but also by topography, varying characteristics of storms and associated clouds, dynamics of the seasonal cycle, and variations in the contribution of extreme precipitation events to the water budget (Bowen et al., 2019).

Individual paleoclimatic proxy types are often sensitive to multiple aspects of the water cycle that can be difficult to disentangle, making it challenging to directly compare among proxy types. For example, precipitation amount in the Arctic could be inferred from two common precipitation proxies: grain size from lake sediments and accumulation rates from ice cores. Grain size fluctuations in lake sediments can track extreme precipitation and runoff events, but inter-lake comparison requires knowledge of lake morphometry and competing moisture source regions (Conroy et al., 2008; Kiefer and Karamperidou, 2019; Rodysill et al., 2019). Comparison of sedimentary grain size to snow accumulation rates would be uninformative without understanding how annual precipitation and dry season ablation, which both affect accumulation rates, are related to moisture delivery from extreme precipitation events (Hurley et al., 2016; Thompson et al., 1986). Snow accumulation rates can be strongly affected by air temperature, whereas grain size is generally not. Thus, although comparison of such heterogenous hydroclimatic proxies is certainly possible, the lack of a common environmental signal to serve as a reconstruction target has been a major hindrance to the global reconstruction of hydroclimatic variables. These challenges have been further exacerbated by archive- and record-specific standards for data formatting, sampling resolution, metadata availability, and public archiving. These limitations may be addressed by creating a metadata-rich, multi-proxy, and multi-archive database of hydrological proxies united through standardized formatting and a common environmental signal: water isotopes.

1.2 The potential for a network of paleo-water isotope records to track past hydroclimate variations

In order to address these challenges, we focus here on the stable oxygen ($\delta^{18}O$) and hydrogen ($\delta^2H$) isotopic compositions of environmental waters such as precipitation, seawater, lake water, and soil and groundwater (Fig. 1). The stable isotopic compositions of such waters (here collectively referred to as “water isotopes”) have long been used as integrative tracers of the modern water cycle (e.g., Bowen et al., 2019; Galewsky et al., 2016; Gat, 2010; Rozanski et al., 1993). The rare heavy isotopologues of water (e.g., $^1\text{H}_2^{18}\text{O}$, $^1\text{H}_2^{2}\text{H}^{16}\text{O}$) fractionate from their lighter, more common counterpart ($^1\text{H}_2^{16}\text{O}$) during evaporation, condensation, and other phase changes, capturing an integrative history of parcels of water as they move through and among oceans, atmosphere, and land (Fig. 1). Global databases of isotopic measurements of modern precipitation (IAEA/WMO, 2019), rivers (Halder et al., 2015), seawater (LeGrande and Schmidt, 2006), and water vapor (Galewsky et al., 2016) have contributed consider-ably to our understanding of the contemporary water cycle on scales from microscales (e.g., cloud microphysics) (Kurita et al., 2011), to mesoscales (e.g., hurricane dynamics) (Good et al., 2014; Kurita et al., 2011), and to global scales (e.g., residence time of atmospheric moisture) (Aggarwal et al., 2012). More recently, spaceborne measurements of $^1\text{H}_2^{18}\text{O}/^1\text{H}_2^{16}\text{O}$ in multiple levels in the atmosphere have identified the critical role of poorly observed processes such as tropical rain re-evaporation (Aggarwal et al., 2012; Worden et al., 2007) and forest–atmosphere feedbacks (Wright et al., 2017). Together with climate and Earth system model simulations, which increasingly incorporate sophisticated water isotope tracers into their hydrologic schemes (Brady et al., 2019; Haese et al., 2013), water isotopes offer observational constraints on processes that are otherwise difficult to identify or constrain (Brady et al., 2019; Nusbaumer et al., 2017).

In the paleoclimate realm, hydroclimate proxy records using water isotopes are commonly obtained from a variety of natural archives, including glaciers, ground ice, cave formations, corals, spherulites, mollusk shells, tree wood, lake sediments, and marine sediments. Of all of the proxy types that are used to reconstruct past hydroclimate changes, water isotopes are arguably the most common and certainly the most widely distributed geographically. A global, spatially distributed network of water isotope proxy records therefore has the potential to capture features of large-scale circulation patterns while minimizing site-specific influences from individual locations (Evans et al., 2013). Paired with an understanding of water cycle processes from modern observations and isotope-enabled model simulations, reconstructions of paleo-$\delta^{18}O$ and $\delta^2H$ from these archives can provide
critical information about water vapor source and air mass transport history, precipitation amount and other characteristics, glacial ice volume changes, and temperature, prior to the beginning of instrumental climate observations (Bowen et al., 2019; Dayem et al., 2010; Galewsky et al., 2016; Konecky et al., 2019b). Further, proxy system models (Evans et al., 2013) are available for most water isotope proxies, facilitating direct comparison with paleoclimate model output and thus an improved understanding of the climate dynamics responsible for observed (spatial and temporal) water isotope variability (Dee et al., 2015, 2018; Dolman and Laepple, 2018; Jones and Dee, 2018; Konecky et al., 2019a; Thompson et al., 2011).

One of the obstacles to synthesizing hydroclimate-sensitive paleoclimate records has been a lack of standardized metadata at the proxy system level that systematically encodes the important variables that are necessary for integrating records into a multi-proxy synthesis and interpreting the results. Although the paleoclimate community is in the process of defining and adopting metadata conventions (Khider et al., 2019), the “bare minimum” current standards (e.g., ISO 19115 for geographic metadata) used by World Data System (WDS) repositories (e.g., NOAA Paleoclimatology, PANGAEA) are insufficient for characterizing water isotope proxy systems in a way that can be reliably applied to large-scale paleo-hydroclimate syntheses. One key example of this challenge is the temperature dependence of O- and H-isotopic fractionation, which has frequently been exploited to reconstruct past temperature changes in locations where air or water temperature exerts first-order influence on isotope ratios in precipitation and/or seawater (Kilbourne et al., 2008; Meyer et al., 2015; Porter et al., 2014). Yet in most places, the influence of temperature on isotopic fractionation is only one of many factors that influence the δ^{18}O and δ^{2}H of precipitation (Liu et al., 2012; Thomas et al., 2018) and seawater (Conroy et al., 2017; Partin et al., 2012; Russon et al., 2013). A network of water isotope records will inevitably contain information about air and water temperature but also other key hydroclimatic variables such as atmospheric moisture source changes and surface water evaporation. In order to tap the full potential of water isotope proxy records in a large-scale synthesis, the metadata associated with such records must be sufficient to capture at least a bare minimum of the complexity of the environmental signals that the records contain.

Additional metadata challenges have hindered progress in paleo-water isotope synthesis thus far. Most published datasets shared outside WDS repositories follow nonuniform metadata standards or contain minimal metadata. Datasets are often catalogued using different conventions (often at the authors’ discretion), stored in varying formats (e.g., text, CSV, PDF), and uploaded to different public or private (i.e., behind journal paywalls) repositories. Furthermore, datasets are frequently archived without the raw chronological information that would be required to propagate age uncertainties if desired. These challenges are common to any paleoclimate synthesis effort and are not unique to water isotopes (Atsavawaranunt et al., 2018; Emile-Geay and Eshleman, 2013; PAGES 2k Consortium, 2017), but they exacerbate the challenge of hydroclimate-specific metadata needs.

1.3 The PAGES Iso2k database

Here we introduce the Past Global Changes (PAGES) Iso2k database, a collection of 759 water isotope proxy records (i.e., individual time series) from 506 sites (geographic locations) covering all or part of the CE. The database has been assembled by the PAGES Iso2k Project (hereafter “Iso2k”). The Iso2k database contains δ^{18}O and δ^{2}H-based paleoclimate records from 10 different archives: glacier and ground ice (210 records); speleothems (68 records); corals, scleropogones, and mollusks (143 records); wood (81 records); terrestrial and lake sediments (158 records); and marine sediments (99 records). Of these, 606 records are considered to be the primary time series for each site (Fig. 2) (see Sect. 2.4 and Table S1 in the Supplement). To address the complexity of environmental signals preserved in these proxy records, the database contains detailed metadata about each record’s isotope systematics and proxy system context, as well as details about the original authors’ climatic interpretation, chronological and analytical uncertainties, and other information required for robust data synthesis and interpretation. Iso2k has developed a uniform framework suitable for all proxy archives in the database. The architecture of the Iso2k database therefore provides a scalable foundation on which future multi-proxy hydroclimatic databases can be built, for example, by incorporating non-isotopic proxy records, such as the grain size and ice accumulation example in Sect. 1.1.

The Iso2k database is the latest in a series of community-led paleoclimate data synthesis efforts endorsed by PAGES (Atsavawaranunt et al., 2018; Kaufman et al., 2020; McGregor et al., 2015; McKay and Kaufman, 2014; PAGES 2k Consortium, 2013, 2017; Tierney et al., 2015). The main distinguishing feature of the Iso2k database is that it is not organized around one archive type, climate variable, or region; rather, it contains a systematic representation of the suite of environmental signals preserved in the water isotopic composition of diverse paleoclimate archives, with no a priori assumptions about the underlying climatic interpretation of those signals. This novel approach yields a database that is flexible enough to evaluate many different environmental parameters and processes during the CE, depending on investigator interest. The Iso2k database also contains even more comprehensive metadata descriptions than previous PAGES compilations (e.g., PAGES 2k Consortium, 2017). Database users can therefore filter for and process only the records required for their research question of interest.

This data descriptor presents version 1.0.0 of the PAGES Iso2k database. We describe the collaborative process of assembling the database (including quality control and valida-
Figure 2. The Iso2k database version 1.0.0. (a) Spatial distribution of “primary time series” records in the Iso2k database. Symbols represent records from different archives. (b) Availability of records in the Iso2k database over time during the past 2000 years.

2 Methods

2.1 Collaborative model

Iso2k is a contribution to Phase 3 of the PAGES 2k Network (PAGES 2k Network Coordinators, 2017). Calls for participation in Iso2k were widely distributed, ensuring a representative cross section of scientists from various disciplines (Konecky et al., 2017, 2018, 2015; Partin et al., 2015). Iso2k built on the successes and challenges of previous PAGES 2k projects (Anchukaitis and McKay, 2014; Kaufman, 2014; PAGES 2k Consortium, 2017; PAGES Hydro2k Consortium, 2017) when deciding on the selection criteria (i.e., requirements for inclusion of records) and metadata fields necessary to make the database suitable for a wide range of applications. Most work was done remotely via teleconferences, with one in-person meeting at the 2017 PAGES Open Science Meeting in Zaragoza, Spain.

The workload for assembling the data and metadata was subdivided among working groups, representing one of the following archive types: marine sediment, marine carbonates (corals, mollusks, sclerosponges), glacier ice, ground ice, lake sediments, speleothems, and wood. This archive-based approach ensured that data were collated by researchers with...
an in-depth, process-based understanding of each proxy system.

2.2 Data aggregation and formatting

The database comprises publicly available water isotope proxy records that span all or part of the CE and meet the criteria outlined in Sect. 2.3. The database was compiled in two main stages. During the first stage, the archive teams obtained records, entered data, and compiled the extensive metadata outlined in Sect. 4. During the second stage, the data and metadata were extensively quality-controlled following the procedure outlined in Sect. 2.4.

We used a variety of sources to identify records for inclusion in the database. We first extracted records that met our selection criteria (described in Sect. 2.3.1) from existing data compilations, including the PAGES 2k temperature database (PAGES 2k Consortium, 2017), the Arctic Holocene Transitions database (Sundqvist et al., 2014), and the SISAL database (Atsawawaranunt et al., 2018). Archive teams then searched the literature and online data repositories (WDS-NOAA and PANGAEA) for additional suitable datasets. For records that had been published but that had not previously been made available in an online public repository (referred to as “dark data”), datasets were digitized from publication tables, appendices, and supplementary materials. Datasets that were not available in their original publications were requested from the authors by email. If two or more email requests went unanswered, the dataset was deemed not publicly available and therefore did not meet that criterion for inclusion in this database. Of the 606 primary time series in the database, more than 20% (128 records) are dark datasets that were added by Iso2k members and are now available in a public, online, machine-readable format for the first time.

The vast majority of those datasets were from wood or from lake and terrestrial sediments (58 and 52, respectively), with an additional 14 from glacier and ground ice, 2 from marine sediments, and 2 from corals.

In addition to isotopic datasets, raw age control data (e.g., 14C ages) were obtained for records where age–depth modeling is required (i.e., non-annually resolved records). Many isotopic datasets that were available through data repositories did not contain raw age control data, in which case we followed the dark data procedure described previously to obtain appropriate chronological data from the authors. For dark age control data, authors were emailed with a request for the data and a spreadsheet template in which chronological information could be added. Age control data from authors who did not respond to these requests could not be added to the database. Again, the majority of “dark” age control data added to the Iso2k database was from the lake sediments archive (over 40 age control datasets are now publicly available for the first time).

Metadata (Sect. 4) were obtained from the data source, extracted from the original publication, or requested from the original data generators (again following the dark data procedure above). We note that even for datasets that were previously publicly available, the Iso2k database has expanded on these data by adding chronological data and compiling an extended suite of metadata not previously available in a consolidated format.

2.3 Record selection criteria

Records were screened by their respective archive teams to ensure that criteria for inclusion in the database were met. Criteria for inclusion in the database were formulated to optimize spatiotemporal coverage of the data, with the goal of building a comprehensive database of water isotope records that can be subsampled as needed to address diverse scientific questions. The selection criteria for data records to be included in the Iso2k database are as follows.

2.3.1 Record resolution and duration

The duration and temporal resolution of records included in the Iso2k database varies by archive type. For ~annually or ~sub-annually banded archives (corals, shells, sclerosponges, tree wood, varved lake and marine sediments, and glacier ice), the minimum record duration for inclusion in the database is 30 years. For all other archives (speleothems, non-varved lake and marine sediments), records must have a minimum duration of 200 years and contain at least five data points during the CE.

2.3.2 Chronological constraints

The PAGES 2k temperature database (PAGES 2k Consortium, 2017) was used as a guide for minimum chronological control criteria. Records from annually banded archives must be either cross-dated or layer-counted; records from non-annually banded archives must have at least one age control point near both the oldest and youngest portions of the record, with one additional age control point somewhere near the middle required for records longer than 1000 years.

2.3.3 Peer review and public availability

To qualify for inclusion in the database, isotope records must be published in a peer-reviewed journal (i.e., not university-published theses and dissertations). Records included in version 1.0 of the database had to be published and publicly available (see definition in Sect. 2.2) before 4 May 2018.

2.3.4 Ancillary data

In some cases, paired geochemical measurements are also included in the Iso2k database to complement interpretation of the isotopic data, such as paired trace elemental measurements (e.g., Sr/Ca or Mg/Ca) that accompany some carbonate δ18O records from corals, sclerosponges, and planktonic
foraminifera, or δ¹³C data that accompany some carbonate records. Derived isotopic data for deuterium excess (d-excess) are also included for glacier and ground ice, where paired measurements of δ¹⁸O or δ²H allowed the original authors to calculate this additional hydroclimatic indicator. Similarly, derived values for the δ¹⁸O of seawater are available for coral and marine sediment records in cases where an independent temperature reconstruction was available for the same archive (e.g., Sr/Ca for corals and Mg/Ca for planktonic foraminifera). Where the paired carbonate δ¹⁸O and Sr/Ca or Mg/Ca records can be used to infer the δ¹⁸O of seawater (Cahyarini et al., 2008; Elderfield and Ganssen, 2000; Gagan et al., 1998), both time series (δ¹⁸O measured directly on carbonate and δ¹⁸O seawater calculated from paired records) and ancillary, non-isotopic geochemical records are included in the database (Sect. 4).

2.4 Quality control procedure

Records considered to be a primary time series for their respective sites (Sect. 4; Table 6) were quality-controlled to the highest degree possible, as described below. Primary time series were judged to be the one or two time series upon which the original authors based their main climatic interpretations. For archives such as corals and speleothems, the primary time series are typically a composite of multiple records from a site or the latest of a series of modified records from a site, whereas for other archives the primary time series is one deemed to have the most robust climatic signal (e.g., for lake sediments, a biomarker of terrestrial vs. mixed terrestrial and aquatic origin). Non-primary time series were quality-controlled as much as possible and are included because they may contain valuable information for database users. Both data and required metadata fields were screened for accuracy and completeness by one or more project members, with the initials of the project member performing the final quality control (QC) check included in the Iso2k_QC_certification metadata field. Metadata fields that required standardized or controlled vocabularies were double-checked to ensure those terms were adhered to (Sect. 4). During the quality control certification process, project members used a web-based data viewer (lipdverse.org) and other visualization tools to display the raw data and metadata.

Each metadata field in the database (Tables 1–7) has a quality control certification “Level” from 1 to 3, defined as follows.

- **Level 1** fields are required metadata for inclusion in the Iso2k database. These fields are generalizable enough to be suitable for all archive types, and they are recommended as primary fields for filtering, sorting, and querying records in the database. Level 1 required fields were subject to the highest QC standard. They follow standardized Iso2k vocabularies, where appropriate (Table 7); geographical data were checked against maps, and interpretation fields were checked against the original publication. Examples of Level 1 metadata include geographical (ISO 19115) and publication information (DOI), and the minimum required subset of isotope and proxy system interpretation metadata fields (see Sect. 4).

- **Level 2** fields are highly useful (but not required) metadata fields in the Iso2k database. They may be used as secondary fields for further filtering, sorting, and querying records in the database; these fields may be particularly useful for certain archives or to refine interpretations after an analysis has been performed. Examples of Level 2 fields include species name (marine and lake sediments and corals) and compound chain length for compound-specific δ²H measurements (lake sediments). Terminology was standardized only where necessary and appropriate. In other cases, these fields contain freeform text with direct quotes from the original publications. During the QC certification process these fields were checked against the original publication for clarity and consistency.

- **Level 3** fields may be useful to some users of the Iso2k database but are not generally recommended as fields for filtering and sorting records in the database. Level 3 fields are not entered as standardized vocabularies and the information is sometimes not available in the original publications (thus, these fields are blank for many records). Examples of Level 3 fields include information pertaining to the integration time of a proxy sensor.

- **Automatic fields** are the automatically generated fields that were computed directly from the data records following QC certification. Fields use standardized vocabularies and units. Examples include binary fields for whether the dataset contains raw chronological control data.

Ancillary data are not quality-controlled but are included in LiPD format for reference.

3 Contents of the Iso2k database

3.1 Archive types within the Iso2k database

The Iso2k database contains data from a variety of geological and biological archives. Following proxy system terminology (Evans et al., 2013), each archive has one or more sensors that directly sense and incorporate environmental signals, i.e., the δ¹⁸O and δ²H of environmental waters, into their structures. Over time these sensors then form, are deposited into, or are otherwise imprinted upon an archive that is then subsampled and subjected to isotopic measurements or observations. In this section, we describe the key characteristics of the archives and sensors that are important for the
interpretation of the paleohydrological signals that they preserve.

3.1.1 Corals, sclerosponges, and mollusks

Corals, sclerosponges, and mollusks (predominantly bivalves and gastropods) form hard body parts of calcium carbonate (aragonite or calcite) that record the conditions of the aquatic environment in which they live (see Black et al., 2019; Corrèze, 2006; Druffel, 1997; Evans et al., 2013; Lough, 2010; Sadler et al., 2014; Surge and Schöne, 2005). Further, except for sclerosponges (which are dated using U/Th geochronology), these aquatic carbonates contain annual banding structures, enabling precise chronology development. Reef-building corals represent the bulk of annually resolved marine archives included in the Iso2k database. These corals are distributed in warm shallow waters throughout the tropical oceans, whereas sclerosponges (i.e., coralline sponges or Demospongiae) and mollusks are found worldwide, the latter in both estuarine and freshwater environments. Micro-sampling and laser ablation technologies allow for sub-annual to annual sampling resolution in corals, mollusks, and sclerosponges for elemental (e.g., Sr/Ca, Mg/Ca) and isotopic analysis ($\delta^{18}O$ and $\delta^{13}C$). When living samples are collected in modern waters, they contain environmental archives of the recent past (decades to several centuries), whereas dead, fossil, and archeological material can be radiometrically dated to provide windows of past isotopic variability, some of which have been cross-dated with modern records (Black et al., 2019, and references therein).

The $\delta^{18}O$ signal in these archives represents a combination of linear, temperature-dependent isotopic fractionation, as well as changes in the isotopic composition of the surrounding water ($\delta^{18}O_w$) (Grottoli and Eakin, 2007; Rosenheim et al., 2005). In some regions, the temperature component dominates the $\delta^{18}O$ signal, whereas in other regions $\delta^{18}O_w$ variability is the primary driver of the $\delta^{18}O$ variability and reflects hydrological and/or oceanographic processes such as vertical and horizontal advection or the freshwater end-member (Conroy et al., 2017; Russon et al., 2013; Stevenson et al., 2018). In some ocean settings, the close coupling between ocean and atmosphere variability leads to co-occurring cool and dry (or warm and wet) anomalies that produce complementary isotopic anomalies (Carilli et al., 2014; Russon et al., 2013; Stevenson et al., 2015, 2018) (e.g., ENSO variability; Cobb et al., 2003). In estuarine or freshwater settings, mollusk $\delta^{18}O$ values are closely linked to the local precipitation–evaporation budget (Azzoug et al., 2012; Carré et al., 2019). Coral $\delta^{18}O$ and $\delta^{13}C$ contain a vital effect and coral $\delta^{18}O$ is offset from $\delta^{18}O_w$, whereas mollusk and sclerosponge $\delta^{18}O$ is generally precipitated in equilibrium with environmental water. Some coral $\delta^{18}O$ records in the Iso2k database have had their mean $\delta^{18}O$ removed by the original authors for comparison and cross-dating with other coral records, and this is noted in the metadata.

3.1.2 Glacier ice

Climate records from glacier ice are found primarily at high latitudes (Antarctica, Arctic) and high elevations (e.g., Andes, Himalayas) (Eichler et al., 2009; Meese et al., 1994). Glacier ice is formed from the accumulation of snow, which over time compacts into a section of chronologically continuous layered ice. Cores drilled through layers of glacier ice preserve sub-annually to centennially resolved climate information, with resolution varying among records due to snow accumulation rates and laboratory sampling and analysis methods (Rasmussen et al., 2014). Ice cores are dated through a variety of methods; annual layer counting and alignment to volcanic horizons are the most common approaches for records spanning the CE (Sigl et al., 2014). This database contains records of $\delta^{18}O$, $\delta^2H$, and/or d-excess of glacier ice. These proxies reflect the isotopic composition of precipitation (snowfall and ice), which is highly correlated to local temperature but additionally reflects changes in moisture source and condensation processes (Goursaud et al., 2019). Physical processes such as isotopic diffusion in the firn, melt and infiltration, and compaction of ice layers generally smooths the seasonal to interannual signal of climate variability in glacier ice, and the potential influence of these processes is site specific.

3.1.3 Ground ice (wedge ice and syngenetic pore ice)

Ground ice includes all types of ice found in permafrost; wedge ice and syngenetic pore ice hold the largest potential for paleoclimate reconstructions (Opel et al., 2018; Porter et al., 2016; Porter and Opel, 2020). Ice wedges in permafrost landscapes form via repetitive thermal contraction cracking in winter and infilling of frost cracks mostly by snowmelt in spring (with potential minor contribution of snow and/or depth hoar). The integrated isotopic composition of the previous winter’s snow pack is transferred into a single ice vein without additional isotopic fractionation due to rapid freezing in the permafrost. Thus, ice wedges preserve precipitation of the meteorological winter and spring, with $\delta^{18}O$ and $\delta^2H$ commonly interpreted as proxies for local air temperature (Meyer et al., 2015). Ice-wedge records are temporally constrained by radiocarbon dating of macrofossils or dissolved organic carbon in the ice. Conversely, pore ice in syngenetic permafrost integrates precipitation that reaches the maximum thaw depth in late summer. The pore ice seasonality is a function of the local precipitation climatology and residence time of active-layer pore waters, and pore ice is enriched in heavy isotopes relative to the initial pore waters due to equilibrium fractionation during freezing (O’Neil, 1968). Because syngenetic pore ice formed within accumulating surface sediments, its age can be modeled based on a radiometrically constrained sediment age–depth profile. Syngenetic pore ice can be cored and subsampled in the same way as glacier ice (Porter et al., 2019).
3.1.4 Lake sediments

Lake sediments may provide long and continuous records of past environmental change (Dee et al., 2018; Mills et al., 2017) and preserve a number of sensors for oxygen and hydrogen isotopes (e.g., Leng and Marshall, 2004). Carbonate minerals – precipitated inorganically from lake waters or in the shells of aquatic invertebrates – have been used as sensors for the isotopic composition of lake water (e.g., Hodell et al., 2001; Morrill, 2004; Von Grafenstein et al., 1998). Additional proxies analyzed with increasing frequency include biogenic silica, mostly from diatoms; (e.g., Chaplin et al., 2016; Swann et al., 2018), cellulose (Heyng et al., 2014), chitinous invertebrate remains (Van Hardenbroek et al., 2018), and lipids (Konecky et al., 2019a; Sachse et al., 2012). Of these proxies, the oxygen isotope composition of carbonates and silicates is subject to temperature-dependent isotope fractionation during mineralization, whereas the isotopic composition of organic materials is generally not influenced by temperature (Rozanski et al., 2010). The compound-specific hydrogen isotopic composition of a lipid reflects the environment in which the organism producing the lipid grew. Lipids produced by aquatic macrophytes or algae reflect the isotopic composition of the lake water, whereas lipids produced by terrestrial plants reflect the isotopic composition of soil or leaf water (which is, in many cases, highly influenced by the isotopic composition of precipitation). Both types of lipids are preserved in lake sediments (Castañeda and Schouten, 2011; Rach et al., 2017; Thomas et al., 2016).

For sensors that record the $\delta^{18}O$ or $\delta^2H$ of lake water, the climatic or hydrological change recorded in $\delta^{18}O$ or $\delta^2H$ depends primarily on the degree to which evaporation influences the lake’s hydrological balance relative to other factors (Gibson et al., 2016; Morrill, 2004). In turn, the effect of evaporation on lake water isotopes largely depends on the residence time of water within the lake system, and the degree of hydrological “closure” of the lake. In open lake systems – which often have surface water inflows and outflows, with a resulting short water residence time – lake waters often reflect the isotopic values of the inflowing waters, which itself generally approximates a (sometimes) lagged signal of the weighted mean of the isotopic composition of local precipitation (Jones et al., 2016; Tyler et al., 2007). In hydrologically closed lakes – often without surface outflows and where more water leaves the system through evaporation – the initial isotopic composition of inflowing waters is altered due to this evaporation, with the $\delta^{18}O$ or $\delta^2H$ of water increasing with increasing evaporation (Dean et al., 2015; Leng and Marshall, 2004).

3.1.5 Wood

The wood in tree rings (tree-ring cellulose) is one of the few terrestrial proxy archives that can be directly constrained to calendar years (McCarroll and Loader, 2004; Schweingruber, 2012). Information about climatic and environmental changes on seasonal-to-annual timescales is recorded in tree-ring cellulose $\delta^{18}O$. The $\delta^{18}O$ of tree-ring cellulose is influenced by (i) the $\delta^{18}O$ of source waters, (ii) factors influencing $\delta^{18}O$ of the leaf water, and (iii) a fractionation factor related to the isotopic exchange of carbonyl oxygen of cellulose intermediates with cellular waters. This fractionation is derived from enriched leaf water and unaltered xylem or source waters and results in an overall $\sim 27\%$ enrichment of cellulose $\delta^{18}O$ relative to cellular waters (Barbour et al., 2004; Gessler et al., 2014; Roden et al., 2000). This fractionation is regarded as a constant in mechanistic models (e.g., Cernusak et al., 2005; Roden et al., 2000), such that cellulose $\delta^{18}O$ variability mainly reflects the $\delta^{18}O$ of source water and leaf waters. The $\delta^{18}O$ of the source water is closely related to the $\delta^{18}O$ of precipitation-derived soil water (Bowen et al., 2019).

As such, the primary climatic signal that controls $\delta^{18}O$ of tree-ring cellulose varies by location, depending on the climatic signals controlling precipitation $\delta^{18}O$ (Sect. 1.2). For example, tree cellulose $\delta^{18}O$ records have been interpreted to reflect temperature at midlatitude to high-latitude sites (e.g., Churakova (Sidorova) et al., 2019; Porter et al., 2014; Saurer et al., 2002; Sidorova et al., 2012) and precipitation amount in tropical or monsoonal sites (Brienen et al., 2013; Managave et al., 2011). As the $\delta^{18}O$ of the soil water is also affected by evaporation of the soil water, precipitation minus evaporation $(P - E)$ influences $\delta^{18}O$ tree cellulose (Sano et al., 2012; Xu et al., 2018). The extent of evaporative enrichment of the source water in $^{18}O$ in the leaf (and hence $\delta^{18}O$ of the leaf water and tree cellulose) is controlled by the water vapor pressure deficit between the leaf intercellular space and the ambient atmosphere, as well as leaf physiological traits (Kahmen et al., 2011; Szejner et al., 2016).

3.1.6 Speleothems

Speleothems are secondary cave deposits that form when water percolates through carbonate bedrock. Both atmospheric CO$_2$ and CO$_2$ generated by plant root respiration and organic matter decomposition are dissolved into rainwater as it percolates through the soil, producing carbonic acid that rapidly dissociates to produce weakly acidic water. As this acidic water percolates through the bedrock, it dissolves carbonate until the water becomes supersaturated with respect to calcium and bicarbonate (Fairchild and Baker, 2012). When the percolating waters emerge in a cave, CO$_2$ degassing from the drip water to the cave atmosphere induces CaCO$_3$ precipitation, resulting in the formation of stalagmites and stalactites (Atkinson et al., 1978) that preserve the $\delta^{18}O$ signal of the waters that have percolated through from the surface (Lachniet, 2009). The $\delta^{18}O$ of the deposited carbonate therefore reflects the $\delta^{18}O$ of soil and groundwater that it infiltrates, which is strongly influenced by the $\delta^{18}O$ of precipitation but with additional influences of aquifer mixing.
times, seasonality of infiltration, and in some cases extreme events (Moerman et al., 2014; Taylor et al., 2013). Processes within the karst and cave, such as calcite precipitation prior to speleothem deposition and/or kinetic isotope effects, can alter the δ¹⁸O of the deposited carbonate. Although there are hydroclimatic limits on speleothem growth, speleothem distribution is largely constrained by the presence of carbonate bedrock (Fairchild and Baker, 2012). Speleothems form in a wide range of hydroclimate conditions, from extremely cold climates in Siberia to arid regions in the Middle East and Australia. The temporal resolution of speleothem paleoclimate series ranges from sub-anual to centennial, and primarily depends on the karst and cave environment. Due to the high precision of uranium series dating, speleothems provide opportunities to determine the timing of regional hydrological response to global events and links to external forcing mechanisms (e.g., insolation changes) (Fischer, 2016). The different types of measurements made on speleothems— including δ¹⁸O, δ¹³C, and various trace elements—and their fluid inclusions can be used to reconstruct past changes in the hydrological cycle.

3.1.7 Marine sediments

Marine sediments contain two types of sensors that have widely been used for measuring water isotope variability: planktonic foraminifera and biomarkers. Planktonic foraminifera are unicellular zooplankton living in the upper hundreds of meters of the ocean. They build a calcite skeleton, which is preserved in the sediment. The δ¹⁸O of planktonic foraminifera calcite reflects a spatially (and temporally) variable combination of temperature and δ¹⁸Osw (Urey, 1948) and to a lesser degree the seawater carbonate ion concentration as well (Spero et al., 1997), although changes in the latter parameter are likely negligible during the CE. The temperature effect on the δ¹⁸O of foraminifera calcite is systematic, i.e., the δ¹⁸Osw can be reconstructed using (species-specific) paleotemperature equations in conjunction with an independent estimate of calcification temperature based on Mg/Ca (Elderdfield and Ganssen, 2000). Planktonic foraminifera have a short life cycle (about a month) and species-specific seasonal and depth habitat preferences (Jonkers and Kučera, 2015; Meiland et al., 2019), such that any planktonic foraminifera record bears an imprint of the ecology of the sensor (Jonkers and Kučera, 2017).

Biomarkers in marine sediments are lipids synthesized either by marine photoautotrophs, which track past changes in surface seawater isotopic values, or from vascular plants, which track soil water isotopic values on an adjacent land mass (Sachse et al., 2012). Biomarkers are strongly affected by isotopic fractionation during lipid biosynthesis, and that fractionation is often assumed to be constant (Sachse et al., 2012). However, as for planktonic foraminifera, biomarker δ²H values are also affected by a combination of environmental parameters. The δ²H values of C₃₇ alkenones (synthesized by coccolithophorids) are impacted by fractionation that changes with salinity and growth rates (Schouten et al., 2006), which can mask changes in the δ²H of seawater. The sources of leaf waxes are terrestrial plants, and the processes affecting leaf waxes in marine sediments are the same as in lake sediments but generally have longer associated time lags between the sensor recording the δ²H of soil water and ultimate deposition in the marine sediment archive.

4 Description of Iso2k metadata fields

The Iso2k database contains over 180 metadata fields. The 55 main fields are described in Tables 1–6; 23 of these were strictly quality-controlled following the Level 1 definition in Sect. 2.4. Entries for some required metadata fields were standardized with controlled vocabulary to allow users to easily query the database for records based on archive type, isotopic ratio (O or H), waters from which the isotope ratios are derived, materials on which the isotope ratios were measured, or the environmental parameter that controls isotopic variability (Fig. 1). Metadata fields describe the primary isotopic variable being inferred, i.e., the “isotope interpretation” (e.g., the δ²H of precipitation); the water from which it was inferred, i.e., “inferred material” (e.g., soil water); the material that was actually measured, i.e., “measured material” (e.g., long-chain n-alkane components of leaf waxes); and information about the original climate interpretation. Distinctions between the archive type (Fig. 2), inferred material (Fig. 3), and the isotope interpretation (Fig. 4) allow for advanced analyses and straightforward data–model comparisons using the database. These metadata interpretation fields were derived from interpretations reported in the original publications. Below and in Tables 1–6, we describe key metadata fields in the database, including all Level 1 and Level 2 fields (see Sect. 2.4 for a description of levels). Table 7 provides standardized vocabularies and common terminologies. Table 8 provides selected chronological control metadata. Table S1 gives key metadata for each primary time series (Sect. 2.4), including all Level 1 fields and selected additional Level 2 fields, and references to original publications (citations also listed in Tables S2 and S3).

4.1 Entity metadata

The entity metadata fields provide basic information for each record, including the isotope measured, the archive type, location (longitude, latitude, and elevation), start and end dates of each record, and both the DOI and citation for the original publication. Entries for the archiveType, paleoData_variableName, and paleoData_units metadata fields are standardized (Table 7) across all archive types to facilitate easy querying and analyses. Each record is assigned a unique LiPD identifier, and all isotope records are assigned a unique Iso2k identifier. The alphanumeric Iso2k identifiers contain 11 characters and digits as follows: archive type (2 char-
4.2 Paleodata metadata

The paleodata metadata fields provide information for each proxy record; a detailed description of key paleodata metadata fields is provided in Table 2. Measured and derived water isotope time series are identified using the paleoData_variableType and paleoData_description fields and should not be confused with the isotope interpretation metadata fields (Sect. 4.3), which more broadly refer to the way each proxy record is interpreted (e.g., speleothem carbonate interpreted as a proxy for the $\delta^{18}$O of precipitation). The variable description (paleoData_description) is the general category of material that was measured for its isotopic ratio (e.g., carbonate or terrestrial biomarker). Further details...
Table 1. Key entity metadata. Bold text indicates Level 1 or required fields in the database; italics are references to other metadata or variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Name of field in database</th>
<th>Additional description</th>
<th>QC level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archive type</td>
<td>archiveType</td>
<td>Type of proxy archive (Tables 2 and 7).</td>
<td>1</td>
</tr>
<tr>
<td>Latitude</td>
<td>geo_latitude</td>
<td>Site latitude in decimal degrees (−90 to +90).</td>
<td>1</td>
</tr>
<tr>
<td>Longitude</td>
<td>geo_longitude</td>
<td>Longitude in decimal degrees (−180 to +180).</td>
<td>1</td>
</tr>
<tr>
<td>Elevation</td>
<td>geo_elevation</td>
<td>Site elevation in meters relative to mean sea level (− below sea level, + above sea level).</td>
<td>1</td>
</tr>
<tr>
<td>Site name</td>
<td>geo_siteName</td>
<td>Name of the site and locality of nearest geopolitical center or municipality if applicable (i.e., islands retain their names).</td>
<td>1</td>
</tr>
<tr>
<td>Dataset ID</td>
<td>dataSetName</td>
<td>Iso2k-specific identifier assigned to all isotope records from a given site and publication.</td>
<td>1</td>
</tr>
<tr>
<td>Unique record ID</td>
<td>paleoData_iso2kUI</td>
<td>Unique Iso2k identifier assigned to each isotope record to distinguish among records when more than one record exists in the original publication.</td>
<td>1</td>
</tr>
<tr>
<td>LiPD ID</td>
<td>paleoData_TSid</td>
<td>Unique LiPD file identifier for each time series in the database.</td>
<td>1</td>
</tr>
<tr>
<td>Variable name</td>
<td>paleoData_variableName</td>
<td>Variable measured (e.g., δ¹⁸O, δ²H). See Table 2 and Table 7 for more metadata.</td>
<td>1</td>
</tr>
<tr>
<td>Variable units</td>
<td>paleoData_units</td>
<td>Units for paleoData_variableName (e.g., permil). See Table 2 and Table 7 for more metadata.</td>
<td>1</td>
</tr>
<tr>
<td>LiPD link</td>
<td>lipdverseLink</td>
<td>Link to LiPDverse webpage.</td>
<td>1</td>
</tr>
<tr>
<td>Maximum year</td>
<td>maxYear</td>
<td>Maximum (most recent) year of each isotope record in calendar year (CE). See Table 8 for more chronology metadata.</td>
<td>auto</td>
</tr>
<tr>
<td>Minimum year</td>
<td>minYear</td>
<td>Minimum (earliest) date of each isotope record in calendar year (CE). See Table 8 for more chronology metadata.</td>
<td>auto</td>
</tr>
<tr>
<td>Publication DOI</td>
<td>pub1_doi</td>
<td>Digital object identifier for the first publication presenting the isotope record.</td>
<td>1</td>
</tr>
<tr>
<td>Publication citation</td>
<td>pub1_citation</td>
<td>Citation for the first publication presenting the isotope record.</td>
<td>3</td>
</tr>
<tr>
<td>Dataset DOI</td>
<td>datasetDOI</td>
<td>Digital object identifier for dataset assigned by original authors if available.</td>
<td>3</td>
</tr>
<tr>
<td>Dataset URL</td>
<td>paleoData_WDSPaleoUrl</td>
<td>URL linking back to records obtained from the NOAA NCEI data repository.</td>
<td>3</td>
</tr>
</tbody>
</table>

are given by measurementMaterial, which is a more specific description of what was measured (e.g., coral, glacier ice, lake sediment), and measurementMaterialDetail, which provides further specificity of the measurementMaterial, such as mineral, species, or compound. In contrast, the inferredMaterial field indicates the environmental source waters whose isotope variability is inferred (e.g., precipitation, lake water, groundwater) (Fig. 1). The environmental source waters in the inferredMaterial field are not meant to be highly specific (e.g., intracellular leaf water) but are instead broad pools of environmental waters that have direct analogs or counterparts in climate models.

4.3 Isotope interpretation metadata

The isotope interpretation metadata fields compile critical information about environmental variables that influence isotopic variability within each record (Table 3). These fields indicate the environmental variable thought to exert dominant
Table 2. Key paleodata metadata. Bold text indicates Level 1 or required fields in the database. italics are references to other metadata or variable; italics are references to other metadata or variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Name of field in database</th>
<th>Description</th>
<th>QC level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable description</td>
<td>paleoData_description</td>
<td>Human-readable description of paleodata_variableName (e.g., carbonate, δ¹⁸O of glacier ice).</td>
<td>1</td>
</tr>
<tr>
<td>Measurement material</td>
<td>paleoData_measurementMaterial</td>
<td>Type of material in which paleodata_variableName was measured (e.g., coral, cellulose, biomarkers).</td>
<td>1</td>
</tr>
<tr>
<td>Measurement material detail</td>
<td>paleoData_measurementMaterialDetail</td>
<td>Free-form text with additional information about paleoData_measurementMaterial.</td>
<td>2</td>
</tr>
<tr>
<td>Inferred material</td>
<td>paleoData_inferredMaterial</td>
<td>Source water whose isotope variability is inferred (e.g., surface seawater, lake water, precipitation). See Table 7.</td>
<td>1</td>
</tr>
<tr>
<td>Inferred material group</td>
<td>paleoData_inferredMaterialGroup</td>
<td>Super-group of inferred material; see Table 7 for controlled vocabulary.</td>
<td>1</td>
</tr>
<tr>
<td>Archive genus</td>
<td>paleoData_archiveGenus</td>
<td>Genus name of the archive, if available.</td>
<td>3</td>
</tr>
<tr>
<td>Archive species</td>
<td>paleoData_archiveSpecies</td>
<td>Species name of the archive, if available.</td>
<td>3</td>
</tr>
<tr>
<td>Values (data field)</td>
<td>paleoData_values</td>
<td>Field containing isotope time series or other measurements for each paleorecord.</td>
<td>3</td>
</tr>
<tr>
<td>Analytical uncertainty</td>
<td>paleoData_uncertaintyAnalytical</td>
<td>Analytical uncertainty in the measured variable when provided by the original publication, based on long-term precision of an internal standard of known value.</td>
<td>3</td>
</tr>
<tr>
<td>Analytical reproducibility</td>
<td>paleoData_uncertaintyReproducibility</td>
<td>Analytical reproducibility in the measured variable when provided by the original publication, based on repeat measurements of replicate samples, transects, or cores from the same site.</td>
<td>3</td>
</tr>
<tr>
<td>Equilibrium evidence</td>
<td>paleoData_equilibriumEvidence</td>
<td>Indicates whether equilibrium conditions were present when the archive formed.</td>
<td>2</td>
</tr>
<tr>
<td>Variable type</td>
<td>paleoData_variableType</td>
<td>Indicates whether the isotope value was measured directly, temporally interpolated (e.g., from age tie points for annually banded archives), or inferred (e.g., seawater isotopic variability, inferred from paired δ¹⁸O and Sr/Ca or δ¹⁸O and Mg/Ca in marine sediments). This information is also incorporated into paleoData_description.</td>
<td>3</td>
</tr>
</tbody>
</table>

control on isotopic variability of the inferred environmental source waters (inferredMaterial) of each record, the mathematical relationship between the isotope interpretation variable and the isotope record, and the season(s) during which this interpretation applies. All isotope interpretation fields in the database are prefaced by isotopeInterpretation. The isotopeInterpretation1_variable field lists the primary driver of isotopic variability in the environmental source waters according to the original publications, for example air temperature or relative humidity (Table 7). For records where multiple variables can explain some fraction of the variability, the isotopeInterpretation2 and isotopeInterpretation3 fields are also populated. The isotopeInterpretation1_direction is a field that gives the sign (positive or negative) of the relation-
ship between the isotope measurements and the environmental variable.

The isotopeInterpretation1_variableGroup field is a simplified super-grouping of terms in the isotopeInterpretation1_variable field in order to facilitate comparisons across different archives and realms, with three options (temperature; isotopic composition of precipitation, i.e., “P_isotope”; or effective moisture). Controlled vocabulary for metadata fields isotopeInterpretation1_variable and isotopeInterpretation1_variableGroup are standardized across all archive types (Table 7).

The isotope interpretation metadata fields reflect the isotope systematics of the environmental source waters and as such are distinct from the climatic inferences that one can make from a proxy record (Sect. 4.4). In some publications, this distinction is explicitly spelled out. For example, the cave drip water that becomes incorporated into the δ18O of speleothem carbonate in Borneo reflects the δ18O of water mixed throughout an aquifer system over many months, which ultimately reflects a smoothed version of precipitation δ18O (Moeran et al., 2014). In that case, the inferredMaterial is soil and groundwater and the isotopeInterpretation1_variable is δ18O_precipitation (“P_isotope”). Separately, δ18O_precipitation at that same site reflects multiple hydroclimatic processes, such as moisture transport and precipitation amount, that lend it a regional imprint of the El Niño–Southern Oscillation (ENSO) (Moeran et al., 2013), and so the climate interpretation of speleothem δ18O is related to ENSO, which would be described separately in the climate interpretation fields (Sect. 4.4). In many publications, the isotope systematics of the environmental source waters and the climate interpretation are stated implicitly rather than explicitly (e.g., by stating that the δ18O of speleothem carbonate reflects monsoon intensity or by stating that it reflects local precipitation amount via the amount effect; Dansgaard, 1964). In these cases, the isotopeInterpretation1_variable is still “P_isotope” and information about the climatic interpretation is included in the climate interpretation fields. These distinctions are critical for facilitating comparisons with isotope-enabled climate models, where complex and nonstationary climate–isotope relationships can be examined directly.

For isotopeInterpretation1_seasonality, some proxy sensors and/or archives are interpreted to record a seasonally biased signal, whereas others may record climate at an annual or sub-annual resolution (e.g., corals, some speleothems, sclerosponges, mollusks, wood). If the record is interpreted to be biased towards a specific season, the calendar months corresponding to that season — generally given as the first letter of each month, unless clarification is necessary — are recorded in the metadata field (e.g., MAM, DJFM, Jan). If the record represents an approximately mean annual signal, “annual” is recorded in the seasonality field. For coral records, if the record has sub-annual resolution (e.g., sampled at monthly or bimonthly intervals), but the overall record is not biased to any particular season, “sub-annual” is recorded in the metadata field.

4.4 Climate interpretation metadata

In contrast to the isotope interpretation (Table 3), climate interpretation metadata (Table 4) represent the original authors’ expert judgment about the primary climatic controls on the isotope ratios at their study site. Climate interpretation metadata specify either climatic variables (e.g., temperature, precipitation amount) or processes (e.g., the Pacific Decadal Oscillation, Asian monsoon intensity) that the authors interpreted to influence the isotopic composition of the proxy record, and as such they are neither standardized nor quality-controlled. These metadata are included as useful background information but should not serve as a primary filter for users of the Iso2k database. A user might filter records based on the isotope interpretation field, then check the climate interpretation field for a qualitative understanding of which climatic processes may be important for the filtered set of records. For records where the isotopeInterpretation2 and isotopeInterpretation3 metadata are populated (Table 3), the corresponding climateInterpretation2 and climateInterpretation3 metadata may also be provided.

4.5 Queryable and standardized metadata

To make the database more user-friendly and queryable, some metadata fields contain logical flags (e.g., 0 or 1, true or false), cross-links (e.g., to a corresponding record ID in another PAGES 2k database), or geographic labels (e.g., continental or ocean basin) that allow for easy sorting (Table 6). For example, if a record was included in the PAGES 2k temperature database and reconstructions (Abram et al., 2016; Kaufman, 2014; PAGES 2k Consortium, 2017; Stenni et al., 2017; Tierney et al., 2015), that record is cross-linked to its associated PAGES 2k ID wherever possible, permitting easy database query and analysis of records in only one database and those common to both databases. Approximately 15% of the records in the Iso2k database were also incorporated into other PAGES 2k compilations, with the most overlap occurring in coral records and high-latitude ice cores. For these records, the extensive metadata can be used to facilitate deeper analyses of the hydroclimatic signals contained in these mainly temperature-dominated isotopic records. For example, with coral δ18O records, many of which are included in both the PAGES 2k temperature and Iso2k databases, the isotope interpretation fields denote the relative influence of δ18Osw vs. temperature on the isotopic variability of the coral carbonate skeleton.

4.6 Chronological control data

Chronological or depth–age metadata provide essential information for isotope records across all archive types, including

Earth Syst. Sci. Data, 12, 2261–2288, 2020

https://doi.org/10.5194/essd-12-2261-2020
Table 3. Key isotope interpretation metadata. Bold text indicates Level 1 or required fields in the database.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Name of field in database</th>
<th>Description</th>
<th>QC level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary isotope interpretation</td>
<td>isotopeInterpretation1_variable</td>
<td>Variable that controls isotopic variability within the record (e.g., “Temperature_air”, “d18O seawater”). See Table 7.</td>
<td>1</td>
</tr>
<tr>
<td>Direction of relationship</td>
<td>isotopeInterpretation1_direction</td>
<td>Sign (“positive” or “negative”) of the relationship between the isotope values and the isotope interpretation variable. For example, a record with a temperature interpretation may have a decrease in $\delta^{18}$O that corresponds to an increase in temperature.</td>
<td>1</td>
</tr>
<tr>
<td>Interpretation group</td>
<td>isotopeInterpretation1_variableGroup</td>
<td>Super-group of isotope interpretations (one of temperature, effective moisture, or precipitation isotope ratio). See Table 7.</td>
<td>1</td>
</tr>
<tr>
<td>Mathematical relation</td>
<td>isotopeInterpretation1_mathematicRelation</td>
<td>Type of relationship between isotope and climate variable (“linear” or “nonlinear”).</td>
<td>2</td>
</tr>
<tr>
<td>Seasonality</td>
<td>isotopeInterpretation1_seasonality</td>
<td>The calendar months the isotope interpretation applies to is given as first initial of the months or as “annual” or “sub-annual” where applicable (e.g., corals, speleothems).</td>
<td>2</td>
</tr>
<tr>
<td>Basis</td>
<td>isotopeInterpretation1_basis</td>
<td>Basis for the isotope interpretation of each record as stated in the original publication (text or citation may be given).</td>
<td>2</td>
</tr>
<tr>
<td>Coefficient</td>
<td>isotopeInterpretation1_coefficient</td>
<td>Numerical coefficient with interpretation variable.</td>
<td>2</td>
</tr>
<tr>
<td>Fraction</td>
<td>isotopeInterpretation1_fraction</td>
<td>Fraction of variance explained by given climate variable.</td>
<td>2</td>
</tr>
</tbody>
</table>

an age model and the average temporal resolution for each isotope record. For non-annually banded records, age–depth models and radiometric dating information (Table 8) are included where available to facilitate independent age modeling. This information is stored in “chronData” tables that are linked to the measured data (“paleoData”) tables. If a record has raw chronology data in the database (e.g., radiometric age determinations), hasChron is set to 1; otherwise this parameter is 0. Similarly, if sample depth data are available (e.g., core depth), hasPaleoDepth is set to 1.

To support the information implicit within each record’s age–depth model, chronological metadata are provided for all individual age constraints (when available) and these metadata are summarized in Table 8. If available, sample information (thickness and labID) is provided for all age constraints. Each age constraint that is not in radiocarbon years has age in calendar years before 1950 CE and ageUncertainty. Radiocarbon age constraints have age14C in radiocarbon years before 1950 CE and age14Cuncertainty. The materialDated, reservoirAge14C, and reservoirAge14Cuncertainty are also provided for radiocarbon age constraints to allow users to derive their own age–depth models if desired. For radiocarbon ages, we also provide fractionModern, fractionModernUncertainty, delta13C (of the material that was radiocarbon dated), and delta13Cuncertainty when available.

Several lake and marine sediment archives contain measurements of radiogenic isotopes – $^{210}$Pb, $^{137}$Cs, and/or $^{239+240}$Pu – to constrain the age of the sediment at and near the surface or core top. Where applicable, we provide the
Table 4. Key climate interpretation metadata.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Name of field in database</th>
<th>Description</th>
<th>QC level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary climate interpretation</td>
<td>climateInterpretation1_variable</td>
<td>Climate variables interpreted in each record (queryable freeform text with quotes from original publications; e.g., “salinity”, “temperature”).</td>
<td>2</td>
</tr>
<tr>
<td>Primary climate interpretation</td>
<td>climateInterpretation1_variableDetail</td>
<td>Provides more information about the climate variable (e.g., sea surface for temperature or salinity).</td>
<td>2</td>
</tr>
<tr>
<td>Climate interpretation direction</td>
<td>climateInterpretation1_direction</td>
<td>Sign (“positive” or “negative”) of the relationship between the isotope ratios and climate variable. For example, a record with a temperature interpretation may have a decrease in $\delta^{18}O$ that corresponds to an increase in temperature.</td>
<td>2</td>
</tr>
<tr>
<td>Climate interpretation basis</td>
<td>climateInterpretation1_basis</td>
<td>Basis for climate interpretation of each record as stated in the original publication.</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 5. Key depth–age metadata. *Bold text indicates Level 1 (required) fields in the database.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Name of field in database</th>
<th>Description</th>
<th>QC level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year (data field)</td>
<td>year</td>
<td>Field containing year data (units are CE) for the paleorecord.</td>
<td>1</td>
</tr>
<tr>
<td>Year units</td>
<td>yearUnits</td>
<td>Units of year data (CE).</td>
<td>1</td>
</tr>
<tr>
<td>Depth (data field)</td>
<td>depth</td>
<td>Depth in archive (e.g., in sediment core, stalagmite).</td>
<td>2</td>
</tr>
<tr>
<td>Depth units</td>
<td>depthUnits</td>
<td>Units of depth measurements.</td>
<td>2</td>
</tr>
<tr>
<td>Chronological integration time</td>
<td>paleoData_chronologyIntegrationTime</td>
<td>Average temporal resolution of each record in years per measurement.</td>
<td>3</td>
</tr>
<tr>
<td>Chronological integration time units</td>
<td>paleoData_chronologyIntegrationTimeUnits</td>
<td>Units for the field.</td>
<td>3</td>
</tr>
</tbody>
</table>

isotope activity and the activityUncertainty. For $^{210}$Pb measurements, the supportedActivity field is Y if the activity is supported by $^{210}$Pb production in the surrounding matrix and N if the activity is not supported. The x210PbModel describes the type of model used to determine the age based on the radiogenic isotope measurements.

For carbonate systems such as speleothems and corals, U/Th dating is often used. Where available, chronological tables in the database contain information about the $^{238}$U and $^{232}$Th content ($U_{238}$, $Th_{232}$), the $^{230}$Th/$^{238}$U activity ratio ($Th_{230}/U_{238}$Activity), $\delta^{234}$U ($d^{234}U$), and their uncertainties ($U_{Thactivity_error}$ and $d^{234}U_{error}$). Fields such as the initial $^{234}$U/$^{238}$U activity ($dU_{234}$initial) and $^{230}$Th/$^{232}$Th activity ratios ($Th_{230}/Th_{232}ratio$) are also included for correcting ages for the initial $^{234}$U/$^{238}$U activity and detrital thorium contamination, respectively.

The useInAgeModel is a binary field where Y indicates that age constraint was used in the published age model and N indicates that age constraint was not used in the published age model.

The amount and type of uncertainty in each chronology are provided in paleoData_chronologyIntegrationTimeUncertainty and paleoData_chronologyIntegrationTimeUncertaintyType, respectively, while paleoData_chronologyIntegrationTimeBasis outlines how the chronology was constructed. By contrast, the paleoData_sensorIntegrationTime, paleoData_sensorIntegrationTimeBasis, paleoData_sensorIntegrationTimeUncertainty, paleo-
Table 6. Selected queryable metadata. *Bold text indicates Level 1 or required fields in the database.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Name of field in database</th>
<th>Description</th>
<th>QC level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Has chronology?</td>
<td>hasChron</td>
<td>Indicates whether chronology data for the isotope record are available in the database.</td>
<td>auto</td>
</tr>
<tr>
<td>Record included in previous PAGES 2k compilation?</td>
<td>paleoData_inCompilation</td>
<td>Indicates whether the record was used in earlier PAGES 2k databases.</td>
<td>2</td>
</tr>
<tr>
<td>Ocean2k ID</td>
<td>paleoData_ocean2kID</td>
<td>Ocean2k unique ID for records included in both databases.</td>
<td>2</td>
</tr>
<tr>
<td>PAGES 2k Dataset ID</td>
<td>paleoData_pages2kID</td>
<td>PAGES 2k temperature dataset ID for records included in both databases.</td>
<td>2</td>
</tr>
<tr>
<td>QC certification</td>
<td>paleoData_iso2kCertification*</td>
<td>Initials of Iso2k Project Member that quality-controlled the record.</td>
<td>1</td>
</tr>
<tr>
<td>Iso2k primary time series for dataset</td>
<td>paleoData_iso2kPrimaryTimeseries*</td>
<td>For sites with multiple time series (e.g., caves with multiple stalagmites and a final composite), this time series should be primarily used (“TRUE” or “FALSE”).</td>
<td>1</td>
</tr>
<tr>
<td>PAGES 2k region</td>
<td>geo_pages2kRegion</td>
<td>The continental (e.g., “SAm” for South America) or ocean (i.e., Ocean) regions corresponding to the PAGES 2k or Ocean2k temperature reconstructions for the records included in those data compilations.</td>
<td>3</td>
</tr>
<tr>
<td>Ocean region</td>
<td>geo_ocean</td>
<td>The ocean region (e.g., Pacific) corresponding to the record site.</td>
<td>3</td>
</tr>
</tbody>
</table>

Data_sensorIntegrationTimeUncertaintyType, and paleoData_sensorIntegrationTimeUnits fields – where available – describe the amount of time over which a sample integrates isotopic values.

5 Key characteristics of Iso2k data records

5.1 Spatial, temporal, archival, and isotopic characteristics of data coverage

The Iso2k database contains 759 stable isotope ($\delta^{18}O$, $\delta^2H$, d-excess) records from 506 unique sites. There are 10 archive types, including 143 records from annually banded skeletal carbonate marine archives, i.e., corals ($n = 137$), sclerosponges ($n = 4$), and mollusks ($n = 2$); 210 from glacier ice ($n = 206$) and ground ice ($n = 4$); 158 from lake or terrestrial sediments; 99 from marine sediments; 68 from speleothems; and 81 from wood (Fig. 2a). A total of 87% of the 759 stable isotope records in the database are $\delta^{18}O$, and 13% are $\delta^2H$, with 12 sites (~2%) having records of both isotope systems (derived from the same sensor in ice cores or different sensors in lake sediments). In addition to the 759 stable isotope records, the database contains 255 records containing ancillary data (e.g., $\delta^{13}C$, Mg/Ca, Sr/Ca). Of the 759 records, 606 are considered “primary” $\delta^{18}O$, $\delta^2H$, d-excess time series (Fig. 2, Table S1 in the Supplement, and Sect. 2.4), including 101 records from annually banded skeletal carbonate marine archives, i.e., corals ($n = 95$), sclerosponges ($n = 4$), and mollusks ($n = 2$); 170 from glacier ice ($n = 166$) and ground ice ($n = 4$); 114 from lake or terrestrial sediments; 95 from marine sediments; 47 from speleothems; and 79 from wood.

Spatial coverage of the sites in the database is global, but most sites are from low latitudes and Northern Hemisphere midlatitudes (Figs. 2a and 4b). Data availability is low for most of the Southern Hemisphere, with the exception of glacier ice records from Antarctica (Fig. 4b). The temporal coverage increases from about 250 proxy time series near the year 0 CE to more than 400 time series at the beginning of the 20th century (Fig. 2b). The average length and resolution of each $\delta^{18}O$ time series vary considerably and are archive dependent. Banded, biologically derived archives (corals, sclerosponges, mollusks, and wood) offer the highest resolution (monthly to seasonal) and a temporal extent of between 24 to 375 years for corals and 38 to 1030 years for...
Table 7. Standardized controlled vocabulary options for metadata fields in the Iso2k database (standardized labels show labels used in Iso2k Database, and parentheses expand any abbreviations).

<table>
<thead>
<tr>
<th>Metadata field</th>
<th>Standardized labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>archiveType</td>
<td>coral, glacier_ice, ground_ice, lake_sediment, marine_sediment, mollusk_shells, terrestrial_sediment, speleothem, sclerosponge, wood</td>
</tr>
<tr>
<td>paleoData_variableName</td>
<td>d2H, d18O</td>
</tr>
<tr>
<td>paleoData_units</td>
<td>permil, zscore, permil_anomaly (specify relative to), PC (principal component)</td>
</tr>
<tr>
<td>isotopeInterpretation1_direction</td>
<td>positive, negative</td>
</tr>
<tr>
<td>isotopeInterpretation1_variable</td>
<td>T_water, d18O_seawater, P_E (precipitation/evaporation), I_E (input/evaporation), P_isotope, T_air, relative humidity, Veg (vegetation dynamics), ET (evapotranspiration)</td>
</tr>
<tr>
<td>isotopeInterpretation1_variableGroup</td>
<td>– Temperature (comprising T_water, T_air)</td>
</tr>
<tr>
<td></td>
<td>– EffectiveMoisture (comprising d18O_seawater, P_E, I_E, relative humidity, Veg, ET)</td>
</tr>
<tr>
<td></td>
<td>– P_isotope</td>
</tr>
<tr>
<td>isotopeInterpretation1_inferredMaterial</td>
<td>Surface seawater, subsurface seawater, precipitation, lake water, soil water, lagoon water, groundwater</td>
</tr>
<tr>
<td>paleoData_inferredMaterialGroup</td>
<td>– Surface water (comprising surface seawater, lake water, lagoon water, subsurface seawater)</td>
</tr>
<tr>
<td></td>
<td>– Precipitation</td>
</tr>
<tr>
<td></td>
<td>– Soil and/or leaf water (comprising soil water, groundwater)</td>
</tr>
<tr>
<td>paleoData_measurementMaterial</td>
<td>Coral, mollusk, ostracod, gastropod, glacier ice, aquatic or terrestrial biomarkers (n-alkane, n-alkanoic acid, dinosterol, botryococcene), planktonic foraminifera, cellulose, carbonate, or bulk carbonate</td>
</tr>
</tbody>
</table>

Tree records (time span is the 2.5%–97.5% quantiles). Layer-counted archives such as glacier ice generally offer annual resolution and a time span between 41 and 1979 years. Other archives have lower resolution but provide more continuous coverage across the CE. The median resolution of records is 12 years per sample for speleothems, 25 years per sample for lake sediments, 28 years per sample for marine sediments, and 97 years per sample for ground ice, and the median time span of records in these archives is >1200 years. These lower-resolution time series almost exclusively make up the records in the database prior to ~1700 CE, preventing the characteristic drop in coverage in older time periods observed in and described by other PAGES 2k compilations (PAGES 2k Consortium, 2013).

The records in the Iso2k Database capture many aspects of hydroclimate (Fig. 4). The first-order interpretation (isotopeInterpretation1_variable) for 44% of the δ18O and δ2H records in the database is “P_isotope”, meaning that δ18O and δ2H of the inferred material (ice, soil water, seawater, etc.) is primarily driven by the δ18O and δ2H of precipitation. The first-order interpretation for 26% of the records in the database is “T_water” or “T_air”, meaning that the temperature of water or air is the primary driver of δ18O and δ2H of the inferred material. Finally, 24% of records in the database are primarily driven by some aspect of evaporation or evapotranspiration, collectively referred to as “effective moisture” in the isotopeInterpretation1_variableGroup category. This category includes “d18O_seawater” (driven by ocean circulation and by precipitation/evaporation at the sea surface), “ET” (evapotranspiration), “I_E” (infiltration/evaporation), and “P_E” (precipitation/evaporation) entries for isotopeInterpretation1_variable.

5.2 Validation

There is currently no existing observational dataset of isotope ratios in all major pools of the water cycle that can serve as a true validation of the Iso2k database. However, the vast majority of ice records in the Iso2k database have an inferred material of “precipitation” and a first-order isotope interpretation of “P_isotope”. For these records, the δ18O averaged for the 20th century (all data points after 1900 CE) provides a reasonable match with the observed annual average δ18O of precipitation from the Global Network of Isotopes in Precipitation (GNIP) (Terzer et al., 2013) (Fig. 5). This provides confidence that the isotopic data contained in the Iso2k database can reasonably be used for analyses such as the calculation of latitudinal gradients in δ18O over the CE, even...
Table 8. Key chronological metadata.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Name of field in database</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>age</td>
<td>Age in calendar years before 1950 CE (after any dating-technique-specific corrections have been applied).</td>
</tr>
<tr>
<td>Age uncertainty</td>
<td>ageUncertainty</td>
<td>1 SD (standard deviation) uncertainty of calendar age.</td>
</tr>
<tr>
<td>Radiocarbon age</td>
<td>age14C</td>
<td>Age in radiocarbon years before 1950 CE.</td>
</tr>
<tr>
<td>Radiocarbon age uncertainty</td>
<td>age14Cuncertainty</td>
<td>1 SD uncertainty of radiocarbon age in years.</td>
</tr>
<tr>
<td>Fraction modern $^{14}$C activity</td>
<td>fractionModern</td>
<td>Fraction of modern radiocarbon activity.</td>
</tr>
<tr>
<td>Fraction modern $^{14}$C activity uncertainty</td>
<td>fractionModernUncertainty</td>
<td>1 SD uncertainty of fraction of modern radiocarbon activity.</td>
</tr>
<tr>
<td>$^{13}$C</td>
<td>delta13C</td>
<td>$^{13}$C of material analyzed for radiocarbon.</td>
</tr>
<tr>
<td>$^{13}$C uncertainty</td>
<td>delta13Cuncertainty</td>
<td>1 SD uncertainty of $^{13}$C of material analyzed for radiocarbon.</td>
</tr>
<tr>
<td>Thickness</td>
<td>thickness</td>
<td>Thickness of the layer analyzed for the age constraint.</td>
</tr>
<tr>
<td>Lab identifier</td>
<td>labID</td>
<td>Unique identifier provided by lab where age analysis was conducted.</td>
</tr>
<tr>
<td>Material dated</td>
<td>materialDated</td>
<td>For radiocarbon age constraints, the material dated.</td>
</tr>
<tr>
<td>Activity</td>
<td>activity</td>
<td>$^{210}$Pb, $^{239+240}$Pu, or $^{137}$Cs activity.</td>
</tr>
<tr>
<td>Activity uncertainty</td>
<td>activityUncertainty</td>
<td>$^{210}$Pb, $^{239+240}$Pu, or $^{137}$Cs activity uncertainty.</td>
</tr>
<tr>
<td>Supported activity</td>
<td>supportedActivity</td>
<td>“Y” if $^{210}$Pb activity is supported; “N” if $^{210}$Pb activity is unsupported.</td>
</tr>
<tr>
<td>$^{210}$Pb model</td>
<td>x210PbModel</td>
<td>Model used to convert $^{210}$Pb activity to age (e.g., constant rate of supply).</td>
</tr>
<tr>
<td>$^{14}$C reservoir age</td>
<td>reservoirAge14C</td>
<td>$^{14}$C reservoir age.</td>
</tr>
<tr>
<td>$^{14}$C reservoir age uncertainty</td>
<td>reservoirAge14CUncertainty</td>
<td>$^{14}$C reservoir age uncertainty.</td>
</tr>
<tr>
<td>U/Th depth</td>
<td>depthUTh</td>
<td>Midpoint depth of the subsample drilled for U/Th age.</td>
</tr>
<tr>
<td>U/Th sample ID</td>
<td>sampleID</td>
<td>Sample ID for the U/Th age measured.</td>
</tr>
<tr>
<td>U/Th sample weight</td>
<td>weight</td>
<td>Weight of powder analyzed for U/Th age in mg.</td>
</tr>
<tr>
<td>$^{238}$U content</td>
<td>U238</td>
<td>$^{238}$U content of the subsample in ppb.</td>
</tr>
<tr>
<td>$^{238}$U error</td>
<td>U238_error</td>
<td>Analytical uncertainty of $^{238}$U in ppb.</td>
</tr>
<tr>
<td>$^{232}$Th content</td>
<td>Th232</td>
<td>$^{232}$Th content of the subsample in ppt.</td>
</tr>
<tr>
<td>$^{232}$Th error</td>
<td>Th232_error</td>
<td>Analytical uncertainty of $^{232}$Th in ppt.</td>
</tr>
<tr>
<td>$^{234}$U ratio</td>
<td>d234U</td>
<td>$^{234}$U ratio measured in the subsample.</td>
</tr>
<tr>
<td>$^{234}$U error</td>
<td>d234U_error</td>
<td>Analytical uncertainty of $^{234}$U.</td>
</tr>
<tr>
<td>$^{230}$Th/$^{238}$U activity</td>
<td>Th230_U238activity</td>
<td>$[^{230}\text{Th}/^{238}\text{U}]$ activity measured in the subsample.</td>
</tr>
<tr>
<td>$^{230}$Th/$^{238}$U activity error</td>
<td>U_Thactivity_error</td>
<td>Analytical uncertainty of $^{230}$Th/$^{238}$U activity.</td>
</tr>
<tr>
<td>$^{230}$Th/$^{232}$Th ratio</td>
<td>Th230_Th232ratio</td>
<td>$[^{230}\text{Th}/^{232}\text{Th}]$ ratio in the subsample in ppm.</td>
</tr>
<tr>
<td>$^{230}$Th/$^{232}$Th ratio error</td>
<td>Thratio_error</td>
<td>Analytical uncertainty of $^{230}$Th/$^{232}$Th ratio in ppm.</td>
</tr>
</tbody>
</table>
Table 8. Continued.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Name of field in database</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncorrected U/Th age</td>
<td>AgeUncorrected</td>
<td>Uncorrected U/Th age of the subsample in years ago.</td>
</tr>
<tr>
<td>Uncorrected U/Th age uncertainty</td>
<td>AgeUncorr_error</td>
<td>Analytical uncertainty of uncorrected Age in years.</td>
</tr>
<tr>
<td>Corrected U/Th age uncertainty</td>
<td>AgeCorr_error</td>
<td>Uncertainty of corrected age (includes Th correction) in years.</td>
</tr>
<tr>
<td>Initial (\delta^{234})U</td>
<td>dU234initial</td>
<td>Calculated initial (\delta^{234})U ratio in the subsample.</td>
</tr>
<tr>
<td>Initial (\delta^{234})U error</td>
<td>dU234i_error</td>
<td>Analytical uncertainty of calculated (\delta^{234})U initial.</td>
</tr>
<tr>
<td>Use in age model?</td>
<td>useInAgeModel</td>
<td>“Y” if this age constraint was used in the published age model and “N” if it was not.</td>
</tr>
</tbody>
</table>

Figure 5. Average \(\delta^{18}\)O from glacier and ground ice records in the Iso2k database (symbols), calculated as the average value since 1900 CE, compared with mean annual \(\delta^{18}\)O from the Global Network of Isotopes in Precipitation (GNIP) (shading) (Terzer et al., 2013). Antarctica is excluded from this map due to the scarcity of GNIP stations.

before accounting for seasonal biases and other transformations within the proxy system. We note that while other proxy data types such as speleothems and leaf wax biomarkers are sensitive to P_isotope (and isotopeinterpretation1_variable for many of these records is listed as “P_isotope”; Fig. 4), their most direct inferred materials are meteoric waters such as soil water or groundwater rather than precipitation; further, water isotope values are fractionated by proxy sensors, such that they are not as directly comparable to the GNIP database.

6 Usage notes

6.1 General applications

The Iso2k database is the most comprehensive database of paleo-water isotope records to date for the CE. For the first time, this database allows for the investigation of spatial and temporal hydroclimate variability from regional to global scales across multiple proxy systems. Using the “inferred material” metadata, the database can be directly compared with the output of climate models, allowing investigation of the water cycle in far greater depth than was previously possible.

Alongside the data itself, the detailed “isotope interpretation” metadata fields are the foundation of this database. These fields allow users to understand the processes reflected in the isotope data and filter the database according to particular scientific questions. For example, a user may be interested in the temporal variability of isotope records driven primarily by changes in effective moisture, and the Iso2k standardized vocabulary means that it is straightforward to filter for these records. Note that, for many records in the database, isotopic variability is affected by more than one variable and that these secondary influences may not be trivial when conducting meta-analyses. Although only isotopeinterpretation1 fields have been quality-controlled to the highest level, the
subsequent isotope interpretation fields also contain well-curated information that is important for data interpretation.

6.2 Example workflow for filtering and querying data records

Records in the Iso2k database are provided as published (i.e., not recalibrated or validated). This preserves the large amount of information contained within water isotope proxy measurements that would be lost if condensed to reconstruct discrete variables. Rather, we leave it to the database users to filter and assess records as needed.

The MATLAB and R serializations contain three variables: “D”, “TS”, and “sTS”. The variable D includes site-level data for each dataset structured in the LiPD format. Datasets in D often contain multiple variables (e.g., stable isotope, ancillary, and chronological data), and represent how LiPD data appear when loaded into the initial environment. For most users, however, a “flattened” version of the database is more useful. We have provided this as the TS variable, where each entry contains an individual time series and its associated metadata. A slightly modified version of TS is included with R and MATLAB, called sTS, which is identical to TS except that the interpretation fields are split by scope (“isotope” or “climate”) in order to simplify querying, which may be preferable for some users. The Python serialization contains only D and TS because tools to split by scope were unavailable.

For initial querying of the database, in nearly all cases, we recommend first filtering by the following variables:

1. `variableName` = “d18O” or “d2H” (excludes any non-isotopic data);
2. `paleoData_units` = “permil” (excludes records published as z scores or anomalies);
3. `paleoData_iso2kPrimaryTimeseries` = “TRUE” (includes only primary time series for each site).

Additional filtering of records should be performed using Level 1 or Level 2 fields; e.g.,

- `isotopeInterpretation1_variable` = `P_isotope` (includes only records where the first-order control of isotopic variability is the isotopic composition of precipitation);
- `paleoData_description` = “carbonate” or “terrestrial biomarker” or “tree ring cellulose” (to extract terrestrial archives sensitive to `P_isotope` aside from ice cores);
- `paleoData_inferredMaterial` = “groundwater” or “soil water” or “lake water” (accomplishes similar results to the above).

Additional filtering of records may be useful with other Level 2 fields; e.g.,

- `climateInterpretation1_variable` = contains “P” or “Precipitation_amount” or “P_amount” (to extract only records where authors’ primary climatic interpretation was based on the amount effect).

The sample R, MATLAB, and Python codes provided with this dataset (in the Supplement) provide a similar example to users.

6.3 Database updates, versioning scheme, and submission of new or updated datasets

This publication marks Version 1.0.0 of the Iso2k database. Following publication, the database will continue to evolve, as new datasets are added (both new studies and previous records that have been missed) and existing data or metadata are extended or, as necessary, corrected. Readers who know of missing datasets are asked to submit them directly through http://lipd.net/playground (last access: 30 July 2020). Database users who find errors in individual datasets can submit proposed edits using the “Edit LiPD file” function at http://lipdverse.org/iso2k/current_version/ (last access: 30 July 2020), or they can use the “Report an issue” option for errors that apply to multiple datasets. More detailed instructions for dataset submission and a link to a LiPD entry template hosted through http://lipd.net/playground (last access: 30 July 2020) will be added to the WDS-NOAA landing page (https://www.ncdc.noaa.gov/paleo/study/29593, last access: 30 July 2020) when they become available.

As the database updates, it will be versioned following the scheme used by other PAGES data collections (Kaufman et al., 2020; McKay and Kaufman, 2014; PAGES 2k Consortium, 2013, 2017), with the following format: X1.X2.X3, where X1, X2, and X3 are incrementing integers. When X1 increases, X2 and X3 reset to zero. When X2 increases, X3 resets to zero. X1 represents the number of publications describing the database. X2 increments each time the set of records in the database changes (addition or removal of a dataset). X3 increments when the data or metadata within the dataset change, but the set of records remains the same. Upon updates, extensions, or corrections to the database, rather than issuing errata to this publication, changes will be included in subsequent versions of the database and updated and described through the online data repository.

6.4 Citation

This Iso2k data descriptor should be cited when the database is used in whole or in part, including its metadata fields, for subsequent studies. We encourage users of the database to not only cite the Iso2k data product but also the original publications of the underlying primary data (Tables S2 and S3). Citation of both the Iso2k data product and the underlying studies is particularly encouraged when analyses make explicit use of individual records or small subsets of records.
The Iso2k database also provides an unprecedented direct comparison for state-of-the-art water-isotope-enabled climate models. Many data–model comparison efforts compile water isotope proxy records from a range of multiple variables in a single proxy record, but this global compilation of water isotope proxy records from the same components of the water cycle circumvents these uncertainties, providing a more direct comparison of proxies and model simulations in the same units. Model validation on this relatively level playing field will improve estimates of climate models’ ability to simulate changes in hydroclimate on long timescales. For those archives that further filter the isotopic signal, proxy system models can aid data model comparison (Dee et al., 2015, 2018; Jones and Dee, 2018). Therefore, the Iso2k database will not only enable global-scale comparisons with isotope-enabled climate models but may also serve as an input database for paleoclimate data assimilation reconstructions such as the Last Millennium Reanalysis (Hakim et al., 2016; Steiger et al., 2014) and the Paleo Hydrodynamics Data Assimilation (Steiger et al., 2018).

**8 Conclusions and anticipated applications of the Iso2k database**

The global extent, quantity, and quality of metadata included in the Iso2k database allow examination of the multiple variables that impact water isotopes, including moisture source and transport history, temperature, and precipitation amount. These multivariate controls mean that water isotopes contain a wealth of information about climate. Importantly, water isotope signals contained in proxy archives can be modified by local environmental processes, such as evaporation, biosynthetic fractionation, bioturbation in sediments, or diffusion. These archive- or proxy-specific transformations therefore additionally allow for reconstruction of water balance ($P - E$), different forms of drought (e.g., meteorological, hydrological, or soil moisture), and relative humidity (Rach et al., 2017). It is difficult to tease apart the effects of multiple variables in a single proxy record, but this global compilation of water isotope proxy records from a range of archives will help to overcome this barrier, facilitating extraction of common signals from the noise of individual proxies and providing insights into different aspects of the hydrological cycle at a range of spatial and temporal scales.

The Iso2k database also provides an unprecedented direct comparison for state-of-the-art water-isotope-enabled climate models. Many data–model comparison efforts compare climate model variables such as temperature and precipitation to paleoclimate data; the latter is often a complex and nonlinear signal integration of multiple climate influences, and uncertainties arise from the assumptions that must be made (Dee et al., 2016; Evans et al., 2013). Comparing water isotope fields from climate model outputs to isotope proxy records of the same components of the water cycle circumvents these uncertainties, providing a more direct comparison of proxies and model simulations in the same units. Model validation on this relatively level playing field will improve estimates of climate models’ ability to simulate changes in hydroclimate on long timescales. For those archives that further filter the isotopic signal, proxy system models can aid data model comparison (Dee et al., 2015, 2018; Jones and Dee, 2018). Therefore, the Iso2k database will not only enable global-scale comparisons with isotope-enabled climate models but may also serve as an input database for paleoclimate data assimilation reconstructions such as the Last Millennium Reanalysis (Hakim et al., 2016; Steiger et al., 2014) and the Paleo Hydrodynamics Data Assimilation (Steiger et al., 2018).

**Supplement.** The supplement related to this article is available online at: https://doi.org/10.5194/essd-12-2261-2020-supplement.

**Team list.** The “Iso2k Project members” group author includes: Kerstin Braun (Institute of Human Origins, Arizona State University, Tempe, Arizona 85287, USA), Matthieu Carré (LOCEAN Laboratory, Sorbonne Universités (UPMC)-CNRS-IRD-MNHN, Paris, 75005, France), Alessandro Incarbona (Department of Earth and Marine Sciences, Palermo University, Palermo, 90134, Italy), Nikita Kaushal (Department of Earth Sciences, University of Oxford, Oxford, Oxfordshire, OX1 3AN, UK), Robert M. Klaebe (Department of Earth Sciences, The University of Adelaide, Adelaide, SA 5005, Australia), Hannah R. Kolus (School of Earth and Sustainability, Northern Arizona University, Flagstaff, Arizona 86011, USA), P. Graham Mortyn (ICTA and Dept. of Geography, Universitat Autonoma de Barcelona (UAB), Bellaterra, 08193, Spain), Andrew D. Moy (Australian Antarctic Division, Kingston, TAS 7050, Australia), Heidi A. Roop (Climate Impacts Group, University of Washington, Seattle, Washington 98195, USA), Marie-Alexandrine Sicre (LOCEAN Laboratory, Sorbonne Universités (UPMC)-CNRS-IRD-MNHN, Paris, 75005, France), and Kei Yoshimura (Institute of Industrial Science, The University of Tokyo, Kashiwa, Chiba, 2778574, Japan).

**Author contributions.** Author contributions are as follows. BLK directed the Iso2k Project. NPM built and managed the Iso2k database. BLK, NPM, OVC, EPD, KLD, DSK, GL, SRM, BM, AJO, JWP, HRS, EKT, DMT, JJT, OC, and MAC designed the database (including development of metadata fields, data selection criteria). OVC, LCB, EPD, MDJ, LJ, DSK, GL, SRM, BM, TO, AJO, JWP, HRS, EKT, DMT, JJT, DVD, and TJP coordinated an archive team. BLK, NPM, OVC, LCB, EPD, KLD, GMF, MJF,
MDJ, LJ, DSK, GL, SRM, BM, TO, AJO, JWP, HRS, EKT, DMT, JTT, AA, OC, MD, DVD, TJP, KB, MC, NK, RMK, HRK, PGM, ADM, HAR, and MAS assembled or entered datasets and/or metadata into database. BLK, NPM, OVC, LCB, EPD, KLD, GMF, MJF, MDJ, LJ, DSK, GL, SRM, BM, TO, AJO, JWP, HRS, EKT, DMT, JTT, AA, JLC, MAC, DVD, ZK, TJP, MC, AI, PGM, ADM, and MAS performed quality control, term standardization, database cleaning, and/or QC certification. BLK, OVC, LCB, EPD, KLD, MJF, MDJ, LJ, DSK, GL, SRM, BM, TO, AJO, JWP, HRS, EKT, DMT, JTT, AA, JLC, MAC, DVD, ZK, TJP, PGM, ADM, and MAS located missing isotopic and/or chronological datasets. BLK, NPM, GMF, MJF, and MDJ analyzed data and generated figures for this paper. BLK, NPM, OVC, LCB, KLD, GMF, MJF, MDJ, LJ, DSK, GL, BM, TO, HRS, EKT, DMT, JTT, NJA, AA, JLC, SGD, ZK, TJP, SS, and MC wrote the manuscript text. BM, NJA, and LvG coordinated with the broader 2k Network. SS and SGD helped align metadata with model comparison needs. KY provided initial project guidance.

Competing interests. The authors declare that they have no conflict of interest.

Special issue statement. This article is part of the special issue “Paleoclimate data synthesis and analysis of associated uncertainty (BG/CP/ESSD inter-journal SI)”. It is not associated with a conference.

Acknowledgements. We gratefully acknowledge Helen Xiu from the Washington University in St. Louis for the illustration in Fig. 1. Iso2k is a contribution to Phase 3 of the PAGES 2k Network. We thank Chief Editor Kirsten Elger and Topical Editor Attila Demény for the handling of this paper. We also thank two anonymous referees and the editors Johannes Wagner and Attila Demény for providing helpful feedback on the paper.

We gratefully acknowledge the authors who sent in published datasets or raw chronological data that had not previously been posted to public repositories: Ramadan Abu-Zied (King Abdulaziz University, Saudi Arabia), Bernard Aichner (University of Potsdam, Germany), Jessica Baker (University of Leeds, UK), Phil Barker (Lancaster University, UK), Max Berkelhammer (University of Illinois at Chicago, USA), Pascal Bohleber (Ca’ Foscari University of Venice, Italy), Thibault Bollé (Laboratoire des Sciences du Climat et de l’Environnement, France), Annette Bolton (Nanyang Technological University, Singapore), Roel Brienen (University of Leeds, UK), Yuda Cahyarin (Indonesian Institute of Sciences, Indonesia), Todd Dawson (University of California, Berkeley, USA), Peter Douglas (McGill University, Canada), Warren Eastwood (University of Birmingham, UK), Nathalie Goodkin (Nanyang Technological University, Singapore), Chis Gouramanis (National University of Singapore, Singapore), Jussi GriESSinger (Friedrich-Alexander-University Erlangen-Nürnberg, Germany), Dan Hammarlund (Lund University, Sweden), Yuxin He (Zhongshan University, China), Maia Heikilä (University of Helsinki, Finland), Andrew Henderson (Newcastle University, UK), David Hodell (University of Cambridge, UK), Jonathan Holmes (University College London, UK), Sally Horn (University of Tennessee Knoxville, USA), James Johnstone (University of California, Berkeley, USA), Vivienne Jones (University College London, UK), Oliver Konter (Johannes Gutenberg University, Germany), Anna Kozachek (Arctic and Antarctic Research Institute and Russian Academy of Sciences, Russian Federation), Jack Lacey (British Geological Survey, UK), Henry Lamb (Aberystwyth University, UK), Chad Lane (University of North Carolina Wilmington, USA), Yanbin Lei (Chinese Academy of Sciences, China), Xiaohua Li (University of Science and Technology of China, China), Yi Lin (National Taiwan University, Republic of China), Neil Loader (Swansea University, UK), Yanbin Lu (Nanyang Technological University, Singapore), Steve Lund (University of Southern California, USA), Christoph Mayr (Friedrich-Alexander-Universität, Germany), Stefan Mischke (University of Iceland, Iceland), Mario Morellón Marteles (University of Cantabria, Spain), Sujata Murty (WHOI), Pratigya Polissar (Lamont-Doherty Earth Observatory of Columbia University, USA), Celia Martin Puertas (GFZ, Germany), David Reynolds (University of Arizona, USA), Donald Rodbell (Union College, USA), Michael Rosenmeier (University of Florida, USA), Ninis Rosqvist (Stockholm University, Sweden), James Russell (Brown University, USA), Dave Ryves (Loughborough University, UK), Masaki Sano (Waseda University, Japan), Matthias Saurer (Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Switzerland), Mark D. Shapley (University of Minnesota, USA), Chuan-Chou (River) Shen (National Taiwan University, Republic of China), Rolf Siegwolf (Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Switzerland), Lora R. Stevens (California State University – Long Beach, USA), Linda Taft (University of Bonn, Germany), Robert Thunell (University of South Carolina, USA), Françoise Vimeux (Laboratoire des Sciences du Climat et de l’Environnement, France), Victor I. Voronin (Siberian Institute of Plant Physiology and Biochemistry, Russian Federation), Brent B. Wolfe (Wilfrid Laurier University, Canada), Chenxi Xu (Chinese Academy of Sciences, China), Giles Young (Swansea University, UK), Zicheng Yu (Lehigh University, USA), Fasong Yuan (Cleveland State University, USA), Mike Evans (University of South Carolina, USA), Sujata Murty (University of California, Berkeley, USA), Rolf Siegwolf (Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Switzerland), Lora R. Stevens (California State University – Long Beach, USA), Amanda Ward (University of Colorado, USA), and Greg Skrzypek (University of Western Australia, Australia) for providing input during the early stages of the project.

Financial support. PAGES received support from the Swiss Academy of Sciences, the US National Science Foundation, and the Chinese Academy of Sciences. Support for this work includes NSF-AGS no. 1805141 to Bronwen L. Konecky and Samantha L. Stevenson and NSF-AGS PRF no. 1433408 to Bronwen L. Konecky. Lukas Jonkers was funded through PalMod, the German paleocli-
mate modeling initiative. PalMod is part of the Research for Sustainable Development initiative funded by the German Federal Ministry of Education and Research (BMBF).

Review statement. This paper was edited by Attila Demény and reviewed by two anonymous referees.

References


Cahyarini, S. Y., Pfeiffer, M., Timm, O., Dullo, W.-C., and Schönberg, D. G.: Reconstructing seawater $\delta^{18}$O from paired coral $\delta^{18}$O and Sr/Ca ratios: Methods, error analysis and problems, with examples from Tahiti (French Polynesia) and Timor (Indonesia), Geochim. Cosmochim. Ac., 72, 2841–2853, 2008.


iso2k1_0_0: LiPDiverse page for Iso2k Current Version, available at: http://lidpdiverse.iso2k/current_version/, last access: 30 July 2020.


Kelley, C. P., Mohtadi, S., Cave, M. A., Seager, R., and Kushnir, Y.: Climate change in the Fertile Crescent and implications of the


Thompson, L. G., Mosley-Thompson, E., Dansgaard, W., and Grootes, P. M.: The Little Ice Age as recorded in the stratigraphy of the tropical Quelccaya ice cap, Science, 234, 361–364, 1986.


