Response to editor and reviewers for

"Technical note: A global database of the stable isotopic ratios of meteoric and terrestrial waters"

by Annie L. Putman and Gabriel J. Bowen

The authors thank Dr. Tian and two anonymous reviewers for their helpful comments. A summary of comments is presented below along with the author's response in blue italics. All page and line numbers refer to the marked up version of the text.

Handling editor, Dr. Fuqiang Tian:

This manuscript introduced a database of isotopes in water. This database is very important because it is better than other existed databases on the aspect of spatial coverage, number of samples and water type. However, because of the importance of this database, we need to think carefully over the quality of data based on enough information. By now, I think the information provided in the manuscript is not enough and some issues should be clarified in more details.

The authors thank Dr. Tian their helpful comments. We have worked to address the concerns of the editor, and hope our adjustments clarify the important points of the manuscript. Page and line numbers refer to the marked up version of the text.

1) The description of method is not enough.
    a. How did you check the data contributed by scientists?
    Within the text we state that we assume that data from peer-reviewed publications are quality checked as part of the publication and peer review process (P 5, L 16-17). However, we do check these data for 'reasonable isotope analysis values, dates and times, and geographic locations. Periodically, we manually check the whole database to ensure latitudes between -90° and 90° and longitudes between -180°and 180°. We check for reasonable water isotope analysis values by reference to the Global Meteoric Water Line, which describes the expected linear correlation between $d^{18}O$ and $d^{2}H$, combined with the expected ranges of each isotope in natural samples of various types.' (P 5, L 17-21). Finally, we caution database users to refer to the original publication/dataset in the case that any data 'appear inconsistent with expected values.' (P 5, L 22-23).
    b. A large number of datasets are available from papers, but they may suffer from a wide range of metadata completeness as you say. How did you solve this problem and integrate the data with different quality together? Or did you remove the data without complete information?
    We chose to include metadata in forms that are flexible and generalizable (Figure 1, panel a). For example, our collection date field includes both date and time, as well as time zone, so can handle sub-daily to multiyear collections. We have added text to clarify these issues, particularly with respect to the most critical metadata. (P 5, L 33 – P 6, L 4)

c. Digitizing data and reconstructing metadata from figures are required in some cases. What kinds of figure have you reconstructed data from? And how about the software you used and its accuracy? Please clarify.

Digitizing data for geographic information may mean approximating sample points from a figure in the manuscript overlaid on a google earth map. In other cases, the table including the data may come from a scanned pdf (in the case of older manuscripts) so the digitizing simply means transcribing the contents of the pdf table into an excel spreadsheet. This is typically done manually, although may be done in some cases aided by the pdf to excel spreadsheet conversion tool available from Adobe Acrobat Pro. We have added text for clarification. (P 4, L 10 – 14).

2) The database is not displayed sufficiently in the manuscript.

a. Only the codes for processing are available (some of the link are not accessible, e.g., the xlsx format files in the 'reports_example_files' folder), but the database itself is not accessible for reader. Considering the database is not published yet, at least some small parts of the database could be provided for example, so that we can easily understand what the database looks like and how to use it.

This is a misunderstanding, due in part because we did not provide the link to the database in the text. Apologies for the oversight. The database is indeed published, and is accessible at:
http://wateriso.utah.edu/waterisotopes/pages/spatial_db/SPATIAL_DB.html
At the request of the reviewers, we have also added this link in the introduction of the wiDB, in the Introduction section as well as in the section titled "Usage notes and further comments". (P 3, L 4 and P 6, L 8)

b. Some important properties of the database are not displayed.

i. The authors mentioned some of the data are not available, and only public data are permitted downloads. How many are they?

These details, included as number and percentage of samples is now included in the Database records section. (P 4, L 28-30)

ii. Please clarify how many data are from scientists' contribution and peer-reviewed paper.

74.2% of samples in the database come from agency or network datasets. Interested readers can use the table in the supplemental information to sort and sum the data in different ways. (P 4, L 30 -32).

iii. Please display the spatial distribution of different sample types, because it can be quite different according to some existed database.

While we like the idea of being able to explicitly show the different sample types on a map, the map would likely contain too many points to be interpretable. Instead, we suggest that readers who are interested in the distributions of different samples types visit the web portal and use the filtering tools provided there. To clarify this point, we have added this additional information to the section "Database records" (P 5, L 1-2).

iv. Again, did you remove the data without complete metadata? If not, the data will have different levels of quality. Please show the number of data with different quality. Also, please show the number of the data you constructed from figures.
As explained in the response above, we do tolerate data with some missing information. However, the most important information (geographic, water type, collection time) is nearly always present (99% of the time) if we preserve data. This detail is included in the Technical validation section. (P 6, L 3-4).

Reviewer 1:

1) Only the importance of isotope research and the data collected at present are explained in the abstract. I suggest that the author add the purpose and significance of this technique to the summary section in a short language.
We have added "The motivation to develop this database comes from water isotope ratios utility in identifying sources, transport pathways, and phase-change processes within the water cycle, which can be used for hydrologic, forensic, ecologic, and hydroclimatic investigations." to the conclusion. (P 7, L 4-6)

2) I recommend that the Database records for individual projects could be presented in Table, and it also shows the collection status of water samples on the global map (like Fig.2), and clearly labels specific sites and projects
We have replaced the list of projects in the supplemental with an excel spreadsheet which contains the same information but is more easily searchable and sortable, which is included as a supplement. However, for the best opportunities to search the available data, we recommend users visit the database portal.

   While we like the idea of a static map of all sites and projects, the amount of data that we'd need to display would render the map difficult for a reader to interpret. Thus, we suggest that readers interested in more detail about any project visit the wiDB portal where the data may be sorted by project, type, and time range, among other parameters. We have added this additional information to the section "Database records" (P 5, L 1-2).

3) Can the explanation of the method be more detailed and specific? Can you make a clear illustration of the three ways of introducing new data to the database? I suggest that the author make a simple sketch so that everyone can understand it.
We have created a flow chart, which has been added as panel (b) to Figure 1. This panel is referenced in the text (P 3, L 30), and will assist readers in understanding the sources of data, and how they are passed to the wiDB.

4) Figure 3 is not clear. I recommend that the author redraw Figure 3. It's better to distinguish the color clear.

We have changed the color scheme and distribution of the different sample types to aid in the interpretation of Figure 3 by the reader.

Reviewer 2:

1) Describe the sample types more detail. For example, considering the phase of precipitation (rain, snow or mixed), the temperature of spring (hot or cold), the types or depths of groundwater (confined or unconfined, and shallow or deep).
   We have added the sentence "The samples table also includes pertinent information for specific water types, like the phase of precipitation (solid, liquid), and depth of groundwater sampling." to the Methods section. (P3, L 23-25)

2) Give the homepage of the wiDB web interface in the manuscript.
   The URL of the wiDB is now included with the introduction of the wiDB, in the Introduction section as well as in the section titled "Usage notes and further comments". (P 3, L 4 and P 6, L 8)

List of relevant changes:

Abstract (P 1, L5 -6) and Database records (P 4, L 22 – 30): All statistics of data included in the database (number of samples, sites, projects and water isotope analyses) were updated from the April 2019 numbers to the July 2019 numbers.

P 3, L 4 and P 6, L 8: Included the wiDB URL

P 3, L 23 – 25: Further explanation of specific metadata included in *Samples* table.

P 3, L 30: Reference to new panel of Figure 1, which shows the three ways data was introduced into the database.

P 4, L 10 -14: More detail was included about how data was digitized and how sample locations were reconstructed if no geographic data was included.

P 4, L 28 – P 5, L 2: More descriptive information about the database was added.

P 6, L 1 – 4: Further information about handling of missing data, and specification of critical metadata.

P 7, L 4 – 6: Reiteration of motivation for development of database.

Technical Note: A global database of the stable isotopic ratios of meteoric and terrestrial waters

Annie L. Putman[1,2] and Gabriel J. Bowen[1,2]

[1]University of Utah Geology & Geophysics, University of Utah, 383 F.A. Sutton Bldg. 115 S. 1460 E., Salt Lake City, UT 84112-0102, US

[2]Global Change & Sustainability Center, University of Utah, 234 F.A. Sutton Bldg. 115 S. 1460 E., Salt Lake City, UT 84112-0102, US

Correspondence: Annie Putman (putmanannie@gmail.com)

Abstract. The hydrogen and oxygen stable isotope ratios of water have been used to identify sources, transport pathways, and phase-change processes within the water cycle, supporting hydrologic, forensic, ecologic, and hydroclimatic investigations. Here, we introduce an unique, open-access, global database of stable water isotope ratios ($\delta^{18}O$, $\delta^{17}O$ and $\delta^{2}H$) from various waters types. This database facilitates data preservation, supports standardized metadata collection, and decreases the time investment for metanalytic research and reference dataset discovery. As of April July 2019, the database includes 227,699 samples from 51,321 231,586 samples from 52,210 sites, associated with 207 218 projects, spanning 1949 through 2019. Key information stored includes the hydrogen and oxygen isotope ratios, water type, collection date and time, site location, and project information. To promote rapid data discovery and collaboration, the database exposes metadata and data owner contact information embargoed data, but only permits downloads of public data. The database is supported by two companion apps, one for processing and upload of analytical data from laboratories and the other an iOS application that supports digital collection of sample metadata.

1 Introduction

Stable isotopes of hydrogen and oxygen in terrestrial and plant waters have proven useful for addressing questions of ecohydrologic connectivity, residence time changes, and fluxes between the atmospheric and continental branches of the hydrologic cycle (e.g., Jasechko et al., 2013; Brooks et al., 2010; Ala-aho et al., 2018; Bowen et al., 2018). For example, terrestrial waters, like lakes, rivers, soil waters, and groundwaters, exhibit complex connectivity to precipitation and among terrestrial water pools (Good et al., 2015), and natural variation in water isotopes, if well documented through reference data, can be used to link sampled water to its sources. In atmospheric branch of the hydrologic cycle, water isotopes have been used to identify continental and oceanic sources of vapor and the lateral and vertical transport of water by circulation systems (e.g., Aemisegger et al., 2015; Fiorella et al., 2015; Cai and Tian, 2016). Data documenting both the continental and atmospheric domains inform each other and research into forensic questions like food or water provenance (Oerter et al., 2017; Jameel et al., 2018) and ecohydrologic questions like plant water use (Oerter et al., 2019).

Oxygen has three major isotopes, ^{16}O (99.757%), ^{17}O (0.038%) and ^{18}O (0.205%), and hydrogen has two stable isotopes ^{1}H (99.9885%) and ^{2}H (0.0115%). The isotopologues most often measured in water are $^{1}H_2^{16}O$, $^{1}H^{2}H^{16}O$, and $^{1}H_2^{18}O$, with fewer $^{1}H_2^{17}O$ measurements. Isotopologue abundance is reported as the Vienna Standard Mean Ocean Water (VSMOW)-normalized heavy-to-light isotope ratio ($\delta = \frac{R_{SA} - R_{VSMOW}}{R_{VSMOW}}$, where $R = \frac{^{18}O}{^{16}O}$, $\frac{^{17}O}{^{16}O}$ or $\frac{^{2}H}{^{1}H}$), in ‰ (Coplen et al., 1996).

Variability in $\delta^{18}O$, $\delta^{2}H$, and $\delta^{17}O$ arise from isotope fractionation during water cycle processes. The term fractionation refers to the sorting of heavy atom (^{18}O, ^{17}O or ^{2}H)-carrying water molecules from the water molecules comprised of only light atoms (^{16}O and ^{1}H), and occurs during phase changes (e.g., evaporation, condensation, and deposition) and across humidity gradients (e.g., vapor diffusion from a saturated water surface into dry air above (Craig and Gordon, 1965)). Fractionation factors, which quantify the strength of sorting, are controlled by both temperature and humidity, where the isotopic sorting effect increases at cooler temperatures and for strong vapor pressure gradients.

The majority of the meridional and altitudinal variation in δ values of observed meteoric waters (precipitation) arises from variation in the extent of Rayeligh distillation: the progressive rainout of heavy isotopologues during the evolution of a precipitating airmass (Gat, 1996). The spatial and temporal variability in precipitation δ values arising from water cycle processes imprints on terrestrial and ecological water pools. Most hydrologic processes (e.g, infiltration, evapotranspiration) reflect mixing of different source waters as opposed to fractionation. Thus, precipitation isotope ratios provide a framework for interpreting variability observed in terrestrial and ecologic water pools. For example, precipitation and terrestrial waters can be used together in mixing models to estimate groundwater provenance (Jasechko et al., 2014) and lake water recharge season (Bowen et al., 2018). Likewise, water isotopes can trace the seasonality of precipitation utilized by trees and plants (Brooks et al., 2010) and the origin of food or drinking water (Oerter et al., 2017; Jameel et al., 2018). Large regional collections of precipitation data have been interpreted in terms of hydroclimatic variability (Liu et al., 2010).

Pioneering research in the areas of climate, hydrology, ecology and forensics that use water isotopes have relied upon large continental or global datasets (Dansgaard, 1964; Rozanski et al., 1992; Bowen and Wilkinson, 2002; Liu et al., 2010), or have assimilated multiple datasets (Masson-Delmotte et al., 2008; Jasechko et al., 2013; Li and Garzione, 2017). These projects highlight the types of research that might be supported by an organized, publicly available archive of oceanic, meteoric, and terrestrial water isotope datasets. Furthermore, such an archive would allow researchers to identify data availability and the extent of previous work in their research area and provide a centralized location for data archiving to satisfy the requirements of funding agencies. A publicly available database might also be used as an educational tool for instruction or class projects (Oerter et al., 2017).

Currently, there are four publicly available water isotope databases. The Stable Water Vapour Isotope Database (SWVID) (Wei et al., 2019), which stores timeseries data for water vapor sampled at several dozen sites worldwide, is hosted by Yale University and supported by the U. S. National Science Foundation. The Global Seawater Database (Schmidt et al., 1999), is a periodically-updated collection of ocean water isotope data. The International Atomic Energy Agency manages two databases that largely serve to distribute data from their long-standing isotope monitoring programs, the Global Network of Isotopes in Precipitation (GNIP) and the Global Network of Isotopes in Rivers (GNIR) archives (IAEA/WMO, 2019).

In this paper we describe the Waterisotopes Database, the 'wiDB', a relational database that archives stable isotope data for a wide range of environmental waters sampled at sub-daily to multiyear temporal scales . (http://wateriso.utah.edu/waterisotopes/pages/spatia The database was originally developed as a private resource supporting the development of gridded precipitation isotope data products (Bowen and Wilkinson, 2002; Bowen et al., 2005), and primarily contained measurements from the PI's lab and precipitation isotope data from GNIP. Over time, it has grown in scope through contributions by the community (e.g., Mayer, 2016; Csank, 2017; Nelson, 2018; Thomas, 2018), and assimilation of literature data (e.g., Xie et al., 2011; Yao et al., 2013; Oshun et al., 2016). About two years ago we began developing open access protocols for the wiDB and publicizing it as a community archive. The current goal for the project is to develop a community repository providing a comprehensive compilation of stable isotope water data via a platform supporting and promoting Findable, Accessible, Interoperable, Reusable (FAIR) data management practices (Wilkinson et al., 2016). Our hope is that the wiDB will both facilitate and improve data preservation within the water isotope research community, in part through the structured archive process, and in part through standardizing the metadata recorded alongside samples in future studies. Web and programmatic exposure of metadata for all wiDB data will support data discovery and sharing within the community, hopefully expanding the accessibility and scope of water isotope-enabled science.

2 Methods

The wiDB is a relational MySQL database hosted by the University of Utah Center for High Performance Computing, and available without authentication via a web search interface and custom APIs. A public API is in development. The wiDB hosts a wide variety of environmental water types, including *Precipitation*, *Rime*, *Lake*, *River_or_stream*, *Ocean*, *Ground*, *Soil*, *Stem*, *Cave_drip*, *Mine*, *Spring*, *Tap*, *Bottled*, *Sprinkler*, *Canal*, *Snow_pit*, *Firn_core*, *Ice_core*, *Cloud_or_fog* and *Vapor*. The wiDB has five tables, *Water_Isotope_Data*, *Samples*, *Sites*, *Projects*, and *Climate_Data* (Figure 1 (a)). The *Samples* table contains the unique *Sample_ID*, the sample start and collection dates, and the water type. The samples table also includes pertinent information for specific water types, like the phase of precipitation (solid, liquid), and depth of groundwater sampling. The *Samples* table is linked to the *Sites* table, which stores site metadata like latitude, longitude, and country, The *Water_Isotope_Data* table, which contains analytical results and metadata, the *Climate_Data* table, which stores basic climate information like temperature and precipitation amount, and the *Projects* table, which records the dataset contributor's contact and data citation information as well as a public/private flag that is used to restrict access to data not (currently) intended for distribution.

We introduce new data to the database in three ways (Figure 1 (b)). First, data may be introduced via the calibration and storage procedures associated with in-house water isotope analyses. Data from all water isotope analysis performed at the University of Utah Stable Isotopes Facility for Environmental Research (SIRFER) lab are automatically stored in the wiDB *Water_Isotope_Data* table as a component of the lab's processing and calibration routines. Researchers analyzing their samples in the SIRFER facility are encouraged to provide metadata to accompany the submission of their samples, which are then uploaded along with the analytical data, making those results discoverable, and (if redistributable) usable by others. Although this protocol is currently used only by SIRFIR, we are eager to work with other labs to test and more widely implement similar measurement-to-archive data protocols. Second, datasets may be contributed by scientists wishing to archive their data to satisfy grant funding requirements or because they support open data initiatives. These datasets are formatted for upload (the template excel file provided to data contributors, 'WI_Template.xlsx' is provided in the supplementary information), checked, and pushed to the database using an R script. Third, a large number of datasets are available as part of peer-reviewed papers or technical reports. Various such published datasets have been formatted for upload by authors or by other members of the community and uploaded to the database. In some cases, this requires digitizing data (e.g., Jacob and Sonntag, 1991; Scholl et al., 2014) ~~,~~ and reconstructing metadata from figures (e.g., Yi et al., 2008). Digitizing data is accomplished by manual transcription of table data from a PDF to excel spreadsheet, sometimes aided by the PDF to excel conversion tool available from Adobe Acrobat Pro DC. Reconstruction of sampling points may be accomplished by uploading or projecting a sampling locations figure into Google Earth and extracting the geographic information. The use of this method is always noted in the sample site metadata.

In an effort to promote standardization of metadata, and to streamline its collection, we have developed and released the wiSamples iOS app. Field metadata can be collected using the app and exported as *.csv files that use the wiDB structure. Thus, sample metadata captured with the app can be directly imported into the database. The app leverages capabilities like GPS, clock and time zone, and reverse gocoding to autopopulate metadata fields using standardized formats. It also queries the wiDB via an API to display the distribution of existing sampling sites and allows association of new samples with these sites, if appropriate.

3 Database ~~Records~~records

As of ~~April~~ July 2019, the *Water_Isotope_Data* table contains ~~248,857~~ 251,481 water isotope analyses, ~~32,027~~ 33,320 of which are in-house analyses from the SIRFIR lab. All analyses are from laser instruments (e.g., Picarro or LGR) or isotope ratio mass spectrometers. The water isotope analyses correspond to ~~227,699~~ 231,586 entries in the *Samples* table, distributed among samples types, as shown in Table 1. There are more analyses than samples because some samples are analyzed multiple times, or because some water isotope analyses do not have a matching sample entry in the case that water was analyzed but no metadata were provided. The samples come from ~~51,321~~ 52,210 sites. Among the ~~207~~ 218 projects currently in the database, there are ~~185~~ 198 publicly available datasets ~~, 13 proprietary datasets , and 10~~ (116,330 samples, 50.2% of all samples), 11 proprietary datasets (14,812 samples, 6.4% of all samples), and 9 datasets that are publicly available elsewhere but for which redistribution is not permitted (100,444 samples, 43.4% of all samples) (e.g., IAEA/WMO, 2015). All ~~projects associated with the database are described individually~~ wiDB projects are described in the supplemental information~~.~~ , and include categorization into network, agency, published and unpublished sub-categories. The majority (74.2%) of samples come from network and agency datasets. Further interactive exploration of the wiDB is possible through the web portal. Here samples can be filtered by project, type, time period, geographic location and/or isotope.

The countries with the most samples in the database include the United States, China, Canada, and Germany (Figure 2). In general, these countries have both long-term monitoring sites and one or more large-scale spatially distributed sampling programs, like national studies characterizing the spatial distribution of lake, river, or tap waters.

Aside from *Ice_core* and *Firn_core* samples, the earliest sampling date recorded in the Samples table is in June of 1949, and the most recent is January 2019. In general, most of the samples collected prior to 2000 are *Ocean*, *Precipitation*, *Ground* or *Ice_core* type samples (Figure 3). In part, the temporal bias may reflect the evolution of the science, as a wider range of water types were measured during the expansion of stable isotope applications to ecohydrologic and forensic studies. However, a large amount of data, particularly for groundwater, was collected during earlier timeperiods that hasn't been published, publicly released, or pulled into the wiDB yet. This demonstrates the opportunity for continued work and community involvement in improving this resource.

4 Technical Validation

The technical validation methods applied vary for data assimilated in different ways. Samples analyzed at the SIRFIR lab are subject to a set of standard laboratory quality control checks using automated scripts and manual screening before being imported to the database. Samples imported from peer-reviewed publications and other databases are assumed to be quality controlled as part of the analysis and publication effort. However, these datasets are checked during organization and after upload for reasonable isotope analysis values, dates and times, and geographic locations. Periodically, we manually check the whole database to ensure latitudes between -90° and 90° and longitudes between -180°and 180°. We check for reasonable water isotope analysis values by reference to the Global Meteoric Water Line, which describes the expected linear correlation between $\delta^{18}O$ and $\delta^{2}H$, combined with the expected ranges of each isotope in natural samples of various types. If a value is updated from a prior version, the update is noted in the comments column of the table. Nonetheless, users are advised to refer to original references to cross-check any data that may appear inconsistent with expected values. To support tracing of potential errors, we attempt to record as much site, sample, and analytical metadata as is available and make an effort to provide raw (non-averaged) water isotope analysis values when available.

Among datasets published with manuscript tables or supplements, we have encountered a wide range of metadata completeness. Issues include, but are not limited to: missing latitudes and longitudes, geographic data reported in difficult to universalize units (e.g., Township and range system (e.g., Williams and Rodoni, 1997) or localized coordinate systems without necessary reference points), missing sampling start or end data, and/or only including processed data (e.g., precipitation weighted monthly or annual averages) as opposed to raw data. In some cases this may reflect changes in technology (e.g., use of maps vs. handheld GPS units) or lack of community guidelines for metadata completeness. However, in other cases this may reflect an author's desire to keep certain aspects of a dataset proprietary. In either case, metadata incompleteness reduces the utility of datasets for both metanalyses and as contextual information for related studies. ~~For this reason, we~~ To address this issue, we have designed the database fields to be flexible to data with varying resolution or missing data. For example, the *Collection_date* field (Figure 1) is a 'datetime' data type, so can handle samples with sub-daily to multi-year time integrations. Likewise, almost all fields can be left blank. Nonetheless, we are able to include critical metadata, like *Collection_Date*, *Type*, *Latitude* and *Longitude* with 98.7% of our samples. However, we suggest that the water isotope community adopt a systematic standard for completeness in reporting of metadata associated with datasets. This practice will ensure the utility and longevity of our datasets.

5 Usage notes and further comments

The database is accessible using a website interface that includes a zoomable-scrollable map showing site locations . (http://wateriso.utah.ed
The sites are clickable, displaying the sample type(s), number of $\delta^{18}O$ or δ^2H analyses ($\delta^{17}O$ are not included yet on the web interface as there are so few data in the wiDB), and range of sample collection dates associated with the site. As well, a link to the data provenance information, including contact names and citations, is provided. A html-based form can be used to search the database using spatial, temporal, sample type, analyte, and project fields. In the future, wiDB access will be provided via a documented, public API supporting programmatic search and download of data.

In the browser portal interface, all proprietary datasets (either restricted due to the data policy of the contributor (e.g., IAEA/WMO, 2015) or because the authors require a data embargo prior to publication) are present on the map, and all information besides the water isotope analyses can be downloaded. This allows users to 1) discover when and where data have been collected, even they are not available for direct download and 2) obtain the data owner's contact information, so a user may contact the owner to request access to the dataset. This solution represents a potential incentive for early data archival. It allows data producers to deposit data early in its lifecycle (e.g., immediately after analysis) without compromising their priority access, while also providing exposure for their work that might lead to new collaborations with the potential data-users and advance the timeline for data discovery and reuse by those users. While allowing and supporting proprietary data is not ideal from the standpoint of FAIR data management practices, the goal reflected in the wiDB design is to recognize that multiple perspectives on timescales for data release persist in the community and offer a middle-ground solution that ensures data archival, but respects the desires of providers for priority use during an embargo period. By exposing metadata, including sampling location, time, and water type, coupled with data owner contact information we hope to promote openness in the community, drive creative research project design, and facilitate collaboration among researchers.

Since the development of the online database, a private API has been used to pull precipitation data from the database, process them to a common time resolution, and use the resulting dataset as the basis for the Isoscapes Modeling, Analysis and Prediction (IsoMAP) (Bowen et al., 2019) tool, a web-based platform for development of derived, gridded, spatiotemporal isotope data productes (isoscapes). This API has access to all of the data within the database, including private data, but does not expose any of the data directly to user download. This represents yet another way in which this water isotope database supports and furthers the knowledge base of the broader community of ecologists, forensic scientists, hydrologists and atmospheric scientists who use stable water isotopes.

6 Conclusions

In this technical note, we present the global Waterisotopes Database (wiDB) , which we have designed as a community repository that supports FAIR data management practice and provides a comprehensive compilation of stable isotope water data. The motivation to develop this database comes from water isotope ratios utility in identifying sources, transport pathways, and phase-change processes within the water cycle, which can be used for hydrologic, forensic, ecologic, and hydroclimatic investigations. Key information stored includes the hydrogen and oxygen isotope ratios, water type, collection date and time, site location, and project information. As of April 2019, the database holds 207 projects totaling 227,699 samples from 51,321 sites, spanning the years 1949 through 2019. We hope that the wiDB will improve data preservation within the water isotope research community, in part through the structured archive process, and in part through standardizing the metadata recorded alongside samples in future studies. In support of those goals, we have developed two associated applications, the first for streamlining the analysis to archiving pipeline, and the second for standardizing and digitizing the sample metadata collection processes. The wiDB web interface is designed to expose metadata for all wiDB data, thereby supporting data discovery and sharing within the community. By introducing and documenting this new set of resources, we hope to expand the accessibility and scope of water isotope-enabled science.

Code and data availability. The R scripts used for SIRFIR data processing and various modes of data upload were developed in-house by the authors and are available via GitHub (specifically 'CRDS_liquid_1.R', 'CRDS_liquid_2.R', 'CRDS_liquid_3.R', 'Upload_metadata.r', and 'Metadata_functions.r', in https://github.com/SPATIAL-Lab/CRDS-processing). All projects associated with the database are described in a spreadsheed included with supplemental information, organized by Project ID. All datasets in the list are available from the wiDB unless otherwise noted. If projects are not redistributable, we comment on where data may be accessed. Finally, code to recreate figures 2 and 3 is available from https://github.com/putmanannie/wiDBViz/upload/hess-2019-173

Author contributions. ALP co-wrote the paper, contributed and implemented management ideas, and organized data. GJB co-wrote the paper, initiated the database, developed upload applications, and organized data.

Competing interests. The authors declare no competing financial interests.

Acknowledgements. This research was supported by NSF grant 1241286, 1565128 and 1759730, and the University of Utah Graduate Research Fellowship. The authors thank Chonghuan Xia for her work on the APIs for wiDB web interface, Karan Sequiera for his design of the web portal, and the Univerity of Utah Center for High Performance Computing for their support of the wiDB. The authors thank Rich

Fiorella for his contributions to database discussions and efforts in digitizing and uploading datasets. Finally, the authors thank all of the researchers and organizations who have contributed data.

References

Aemisegger, F., Spiegel, J. K., Pfahl, S., Sodemann, H., Eugster, W., and Wernli, H.: Isotope meteorology of cold front passages: A case study combining observations and modeling, Geophysical Research Letters, 42, 5652–5660, https://doi.org/10.1002/2015GL063988, 2015.

Ala-aho, P., Soulsby, C., Pokrovsky, O., Kirpotin, S., Karlsson, J., Serikova, S., Vorobyev, S., Manasypov, R., Loiko, S., and Tetzlaff, D.: Using stable isotopes to assess surface water source dynamics and hydrological connectivity in a high-latitude wetland and permafrost influenced landscape, Journal of Hydrology, 556, 279 – 293, https://doi.org/https://doi.org/10.1016/j.jhydrol.2017.11.024, 2018.

Bowen, G. and Wilkinson, B.: Spatial distribution of δ^{18}O in meteoric precipitation, Geology, 30, 315–318, https://doi.org/10.1130/0091-7613(2002)030<0315:SDOOIM>2.0.CO;2, 2002.

Bowen, G. J., Wassenaar, L. I., and Hobson, K. A.: Global application of stable hydrogen and oxygen isotopes to wildlife forensics., Oecologia, 143, 337–348, https://doi.org/10.1007/s00442-004-1813-y, 2005.

Bowen, G. J., Putman, A., Brooks, J. R., Bowling, D. R., Oerter, E. J., and Good, S. P.: Inferring the source of evaporated waters using stable H and O isotopes, Oecologia, 187, 1025–1039, https://doi.org/10.1007/s00442-018-4192-5, 2018.

Bowen, G. J., West, J., Miller, C. C., Zhao, L., and T., Z.: IsoMAP: Isoscapes Modeling, Analysis and Prediction (version 1.0). The IsoMAP Project, http://isomap.org, 2019.

Brooks, J. R., Barnard, H. R., Coulombe, R., and McDonnell, J. J.: Ecohydrologic separation of water between trees and streams in a Mediterranean climate, Nature Geoscience, 3, 100, https://doi.org/10.1038/NGEO722, 2010.

Cai, Z. and Tian, L.: Atmospheric Controls on Seasonal and Interannual Variations in the Precipitation Isotope in the East Asian Monsoon Region, Journal of Climate, 29, 1339–1352, https://doi.org/10.1175/JCLI-D-15-0363.1, 2016.

Coplen, T. B., De Biévre, P., Krouse, H. R., Vocke Jr., R. D., Gröning, M., and Rozanski, K.: Ratios for light-element isotopes standardized for better interlaboratory comparison, Eos, Transactions American Geophysical Union, 77, 255–255, https://doi.org/10.1029/96EO00182, 1996.

Craig, H. and Gordon, L. I.: Deuterium and oxygen 18 variations in the ocean and marine atmosphere, in: Stable Isotopes in Oceanographic Studies and Paleotemperatures, p. 9, 1965.

Csank, A.: Private Communication, 2017.

Dansgaard, W.: Stable isotopes in precipitation, Tellus, 16, 436–468, https://doi.org/10.1111/j.2153-3490.1964.tb00181.x, 1964.

Fiorella, R. P., Poulsen, C. J., Zolá, R. S. P., Barnes, J. B., Tabor, C. R., and Ehlers, T. A.: Spatiotemporal variability of modern precipitation δ^{18}O in the central Andes and implications for paleoclimate and paleoaltimetry estimates, Journal of Geophysical Research: Atmospheres, 120, 4630–4656, https://doi.org/10.1002/2014JD022893, 2015.

Gat, J. R.: Oxygen and hydrogen isotopes in the hydrologic cycle, Annual Review of Earth and Planetary Sciences, 24, 225–262, https://doi.org/10.1146/annurev.earth.24.1.225, 1996.

Good, S. P., Noone, D., and Bowen, G.: Hydrologic connectivity constrains partitioning of global terrestrial water fluxes, Science, 349, 175–177, https://doi.org/10.1126/science.aaa5931, 2015.

IAEA/WMO: Global Network of Isotopes in Precipitation: The GNIP Database, accessed 12-2015, 2015.

IAEA/WMO: Global Network of Isotopes in Rivers, http://www-naweb.iaea.org/napc/ih/IHS_resources_gnir.html, 2019.

Jacob, H. and Sonntag, C.: An 8-year record of the seasonal variation of ^{2}H and ^{18}O in atmospheric water vapour and precipitation at Heidelberg, Germany, Tellus B, 43, 291–300, https://doi.org/10.1034/j.1600-0889.1991.t01-2-00003.x, 1991.

Jameel, Y., Brewer, S., Fiorella, R. P., Tipple, B. J., Terry, S., and Bowen, G. J.: Isotopic reconnaissance of urban water supply system dynamics, Hydrology and Earth System Sciences, 22, 6109–6125, https://doi.org/10.5194/hess-22-6109-2018, https://www.hydrol-earth-syst-sci.net/22/6109/2018/, 2018.

Jasechko, S., Sharp, Z. D., Gibson, J. J., Birks, S. J., Yi, Y., and Fawcett, P. J.: Terrestrial water fluxes dominated by transpiration, Nature, 496, 347, https://doi.org/10.1038/nature11983, 2013.

Jasechko, S., Birks, S. J., Gleeson, T., Wada, Y., Fawcett, P. J., Sharp, Z. D., McDonnell, J. J., and Welker, J. M.: The pronounced seasonality of global groundwater recharge, Water Resources Research, 50, 8845–8867, https://doi.org/10.1002/2014WR015809, 2014.

Li, L. and Garzione, C. N.: Spatial distribution and controlling factors of stable isotopes in meteoric waters on the Tibetan Plateau: Implications for paleoelevation reconstruction, Earth and Planetary Science Letters, 460, 302 – 314, https://doi.org/https://doi.org/10.1016/j.epsl.2016.11.046, 2017.

Liu, Z., Bowen, G. J., and Welker, J. M.: Atmospheric circulation is reflected in precipitation isotope gradients over the conterminous United States, Journal of Geophysical Research: Atmospheres, 115, D22 120, https://doi.org/10.1029/2010JD014175, 2010.

Masson-Delmotte, V., Hou, S., Ekaykin, A., Jouzel, J., Aristarain, A., Bernardo, R. T., Bromwich, D., Cattani, O., Delmotte, M., Falourd, S., Frezzotti, M., Gallée, H., Genoni, L., Isaksson, E., Landais, A., Helsen, M. M., Hoffmann, G., Lopez, J., Morgan, V., Motoyama, H., Noone, D., Oerter, H., Petit, J. R., Royer, A., Uemura, R., Schmidt, G. A., Schlosser, E., Simões, J. C., Steig, E. J., Stenni, B., Stievenard, M., van den Broeke, M. R., van de Wal, R. S. W., van de Berg, W. J., Vimeux, F., and White, J. W. C.: A Review of Antarctic Surface Snow Isotopic Composition: Observations, Atmospheric Circulation, and Isotopic Modeling, Journal of Climate, 21, 3359–3387, https://doi.org/10.1175/2007JCLI2139.1, 2008.

Mayer, B.: Private Communication, 2016.

Nelson, D.: Private Communication, 2018.

Oerter, E., Malone, M., Putman, A., Drits-Esser, D., Stark, L., and Bowen, G.: Every apple has a voice: using stable isotopes to teach about food sourcing and the water cycle, Hydrology and Earth System Sciences, 21, 3799–3810, https://doi.org/10.5194/hess-21-3799-2017, 2017.

Oerter, E. J., Siebert, G., Bowling, D. R., and Bowen, G.: Soil water vapour isotopes identify missing water source for streamside trees, Ecohydrology, 0, e2083, https://doi.org/10.1002/eco.2083, https://onlinelibrary.wiley.com/doi/abs/10.1002/eco.2083, 2019.

Oshun, J., Dietrich, W. E., Dawson, T. E., and Fung, I.: Dynamic, structured heterogeneity of water isotopes inside hillslopes, Water Resources Research, 52, 164–189, https://doi.org/10.1002/2015WR017485, 2016.

Rozanski, K., Araguás-Araguás, L., and Gonfiantini, R.: Relation Between Long-Term Trends of Oxygen-18 Isotope Composition of Precipitation and Climate, Science, 258, 981–985, https://doi.org/10.1126/science.258.5084.981, 1992.

Schmidt, G., Bigg, G. R., and Rohling, E. J.: Global Seawater Oxygen-18 Database - v1.22, https://data.giss.nasa.gov/o18data/, 1999.

Scholl, M. A., Torres-Sanchez, A., and Rosario-Torres, M.: Stable Isotope ($\delta^{18}O$ and $\delta^{2}H$) Data for Precipitation, Stream Water, and Groundwater in Puerto Rico, Tech. rep., U.S. Geological Survey, https://doi.org/https://dx.doi.org/10.3133/ofr20141101, open-File Report 2014–1101, 2014.

Thomas, E.: Private Communication, 2018.

Wei, Z., Lee, X., Aemisegger, F., Benetti, M., Berkelhammer, M., Casado, M., Caylor, K., Christner, E., Dyroff, C., GarcÃa, O., GonzÃ¡lez, Y., Griffis, T., Kurita, N., Liang, J., Liang, M.-C., Lin, G., Noone, D., Gribanov, K., Munksgaard, N. C., Schneider, M., Ritter, F., Steen-Larsen, H., Vallet-Coulomb, C., Wen, X., Wright, J. S., Xiao, W., and Yoshimura, K.: A global database of water vapor isotopes measured with high temporal resolution infrared laser spectroscopy, Scientific Data, 6, 180 302, https://doi.org/10.1038/sdata.2018.302, 2019.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da, S. S., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., der Lei van, van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B.: The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, 3, 160 018, https://doi.org/10.1038/sdata.2016.18, https://doi.org/10.1038/sdata.2016.18, 2016.

Williams, A. E. and Rodoni, D. P.: Regional isotope effects and application to hydrologic investigations in southwestern California, Water Resources Research, 33, 1721–1729, https://doi.org/10.1029/97WR01035, 1997.

Xie, L., Wei, G., Deng, W., and Zhao, X.: Daily $\delta^{18}O$ and δD of precipitations from 2007 to 2009 in Guangzhou, South China: Implications for changes of moisture sources, Journal of Hydrology, 400, 477–489, https://doi.org/https://doi.org/10.1016/j.jhydrol.2011.02.002, iD: 271842, 2011.

Yao, T., Masson-Delmotte, V., Gao, J., Yu, W., Yang, X., Risi, C., Sturm, C., Werner, M., Zhao, H., He, Y., Ren, W., Tian, L., Shi, C., and Hou, S.: A review of climatic controls on $\delta^{18}O$ in precipitation over the Tibetan Plateau: Observations and simulations, Reviews of Geophysics, 51, 525–548, https://doi.org/10.1002/rog.20023, 2013.

Yi, Y., Brock, B. E., Falcone, M. D., Wolfe, B. B., and Edwards, T. W. D.: A coupled isotope tracer method to characterize input water to lakes, Journal of Hydrology, 350, 1–13, https://doi.org/https://doi.org/10.1016/j.jhydrol.2007.11.008, iD: 271842, 2008.

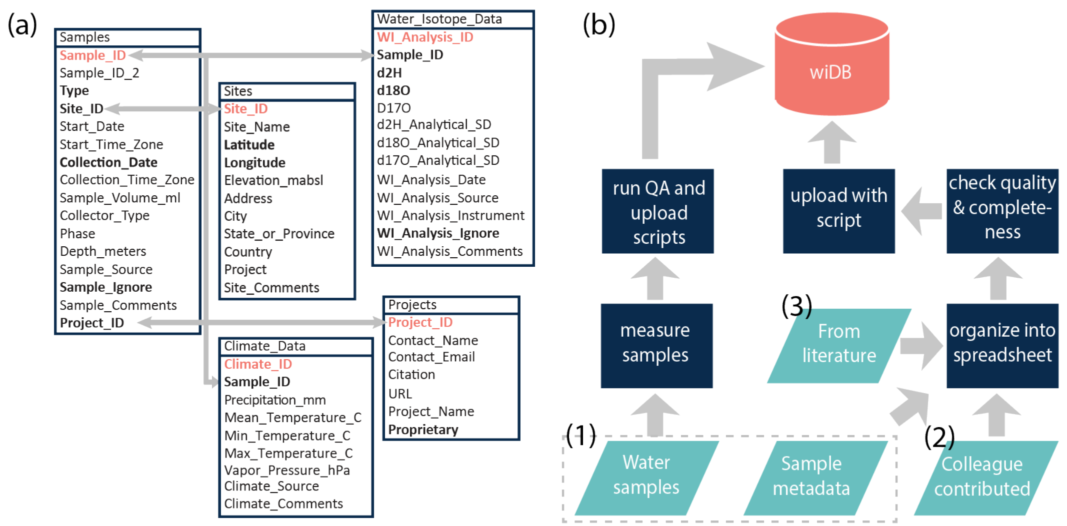

Figure 1. The (a) wiDB schema ~~.~~ and (b) flow chart detailing methods for introducing data into the database. The 5 data tables are linked to one another using primary keys, highlighted in ~~maroon~~coral. Metadata fields required for all entries are bolded. The data sources (teal) are numbered in the order they are discussed in the text.

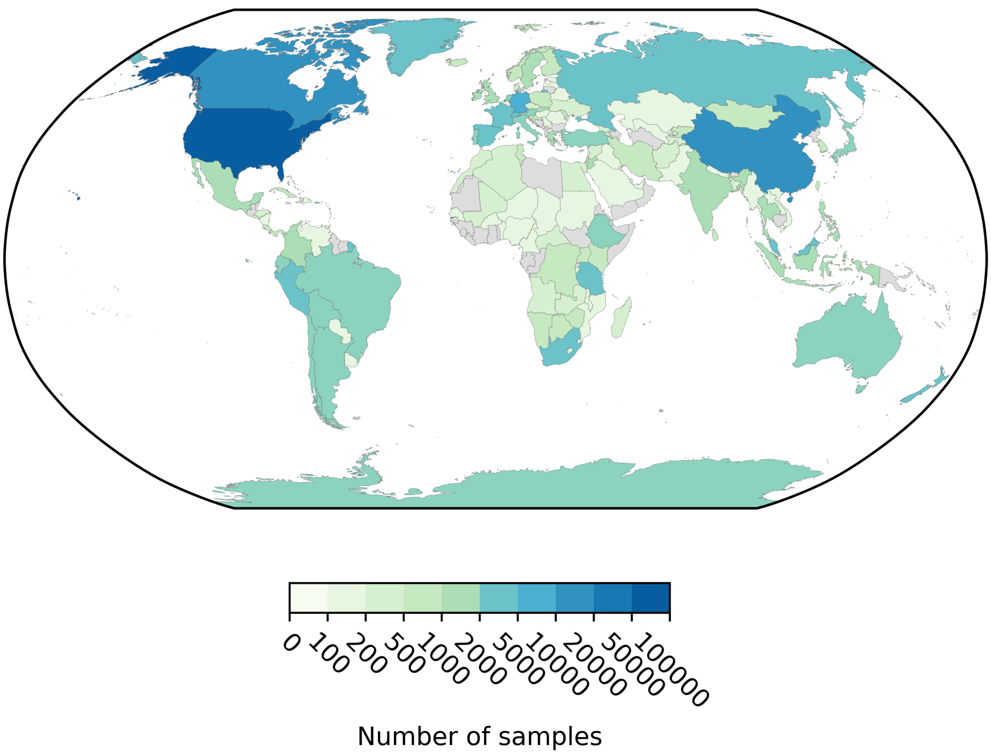

Figure 2. Number of samples (of any type) from each country. Countries in gray have no samples currently stored in the database. Ocean samples are not included in counts. Note that the segmented color scale is not linear.

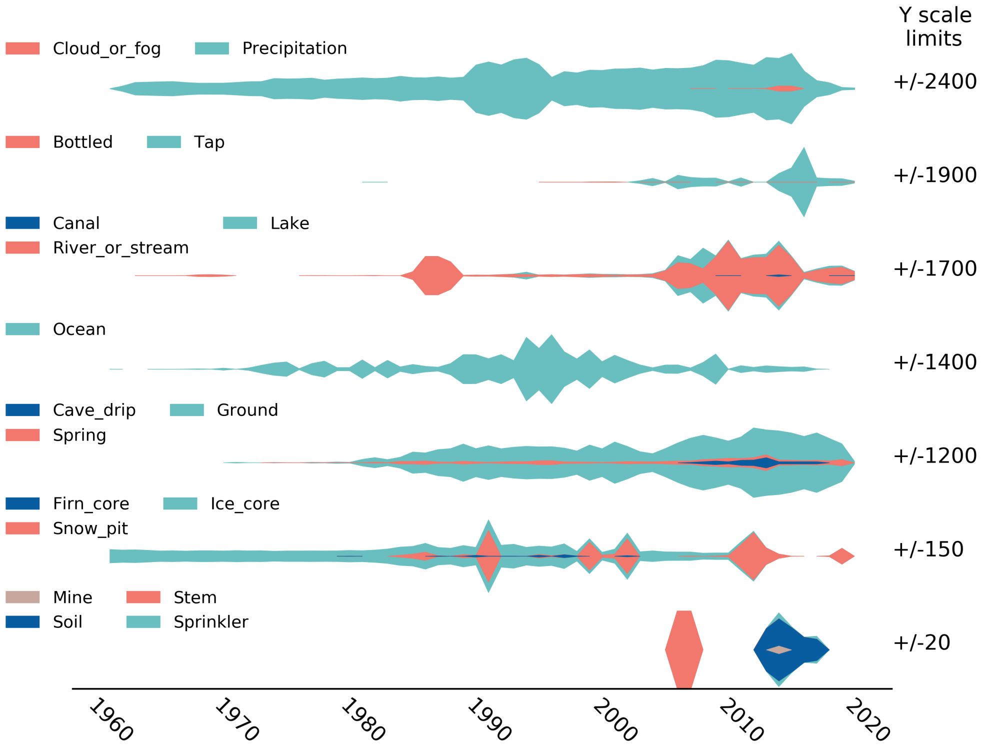

Figure 3. Temporal coverage of samples grouped by type and binned by collection year. Plots are ordered by most numerous sample types to most 'niche' sample types, and the y-scale varies by plot. The sample is attributed to the latest year within the collection period, even if it may represent integrated sampling across multiple years (e.g., *Precipitation*). We do not show the entire temporal range of *Ice_core* samples.

Table 1. Sample Types and abundance in database.

Abundance	Type
169	Bottled
97	Canal
1257	Cave_drip
708	Cloud_or_fog
41	Firn_core
~~38637~~ 39032	Ground
8377	Ice_core
~~4730~~ 5168	Lake
4	Mine
26693	Ocean
~~107894~~ 108890	Precipitation
47	Rime
~~24435~~ 26383	River_or_stream
1015	Snow_pit
97	Soil
~~3186~~ 3189	Spring
10	Sprinkler
84	Stem
~~9802~~ 9909	Tap
374	Vapor