

Wrangling data from short Antarctic ice cores

Barbara Stenni¹ and Elizabeth R. Thomas²

We share our experience of compiling ice-core data for PAGES' Antarctica 2k working group publications. Almost one third of the records were not publicly archived, despite appearing in peer-reviewed literature, highlighting the obstacles posed when performing synthesis studies.

Paleoclimate research, and particularly ice-core research, is expanding, resulting in a welcome increase in scientific data. However, we need to ensure we are following best practices for archiving our data to achieve the maximum impact and sustained use of the data now and in the future (Kaufman and PAGES 2k special-issue editorial team 2018). Paleoclimate reconstruction is moving away from studies based on single locations to a more regional- and continental-scale approach (PAGES 2k Consortium 2013, 2017). Community efforts, such as PAGES, provide a platform to bring together researchers from a wide range of disciplines and scientific backgrounds to address key scientific questions. Journals are now taking the welcome step of requesting that published data be archived, and organizations such as PAGES have taken the initiative in proposing data standards for paleoclimate archives. McKay and Emile-Geay (2016) proposed the Linked Paleo Data (LiPD) format for data archiving that has been adopted by several PAGES projects, but some issues still arise when, for example, collating historical data for climate reconstructions. LiPD is a machine-readable data container, designed for paleoclimate data, that allows multiple levels of metadata as well as descriptions of proxy relations to climate variables (McKay and Emile-Geay 2016).

As one of the PAGES 2k regional working groups, Antarctica 2k was tasked with compiling ice-core stable water isotopes (proxy for past local surface temperature) and snow accumulation (precipitation) records. Figure 1 (upper panels) shows the ice-core site locations for both compilations as well as the length of the records. The resulting reconstructions were published as part of the PAGES 2k special issue in *Climate of the Past* (Stenni et al. 2017; Thomas et al. 2017). The exercise highlighted the importance of data archiving. While collating ice-core records we also faced some difficulties that we want to share here. Moreover, we make some recommendations to the paleoclimate community to expand upon the data format proposed by McKay and Emile-Grey (2016) to facilitate future endeavors.

Experience collating ice-core data

For compiling the Antarctica 2k isotopic database, the records were identified

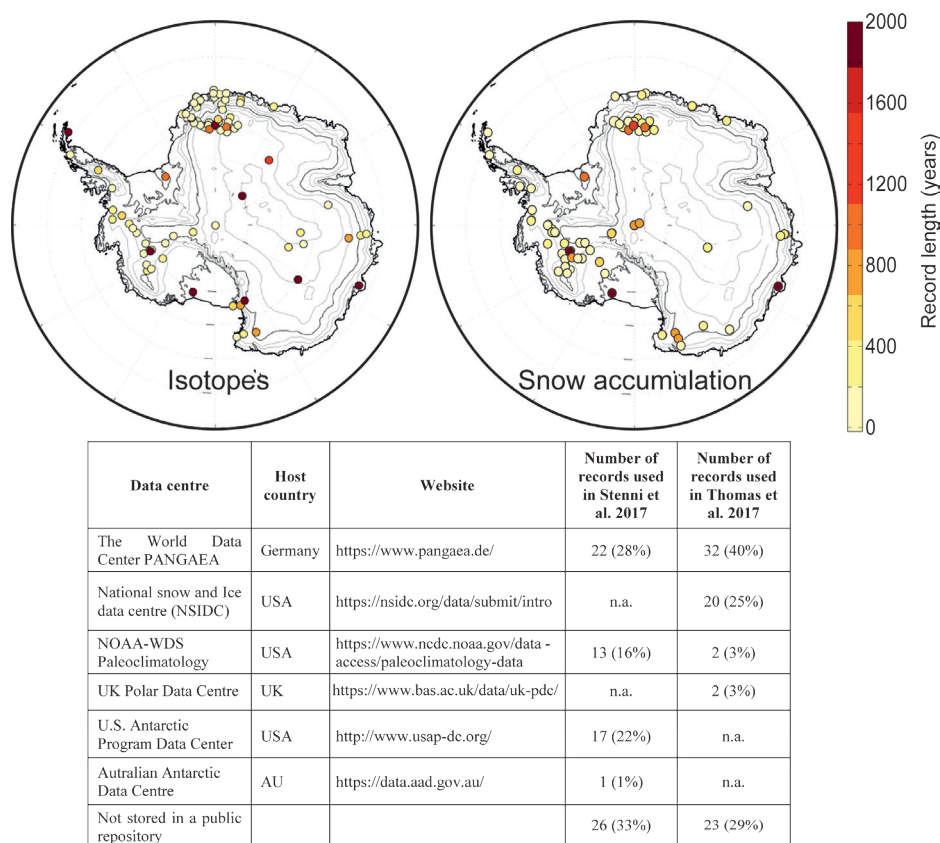


Figure 1: Upper panel: locations and length of records for the recent PAGES Antarctica 2k water isotope (left, Stenni et al. 2017) and snow accumulation (right, Thomas et al. 2017) compilations. Lower panel: list of recognized ice-core data repositories and the numbers of records used in the Antarctica 2k database.

by searching the literature and calling for data from the Antarctica 2k working group mailing list subscribers. A total of 112 records were collected but only 79 met the minimum requirement of having at least 30 years of data coverage since 1800 CE (Stenni et al. 2017). One of the selection criteria developed by the PAGES 2k Network (pastglobalchanges.org/initiative/wg/2k-network/data) was that the data used in the compilation must be published, peer-reviewed and publicly available. However, about one-third of the records used in the syntheses were not previously available publicly, despite them having been described in peer-reviewed publications. Only 53 records were publicly available, distributed among four different data repositories, while 33% of the records had not been uploaded after publication

(Fig. 1; lower panel). At this point a major effort was required to have all the data uploaded in a public repository. A request was sent to authors asking them to deliver the selected data to a public data center. These requests resulted in three different outcomes (1) the authors agreed and deposited their data, (2) they sent us the data, which we directly uploaded to NOAA-WDS Paleoclimatology, and (3) five records were made available in the article's supplementary material through the journal's website upon publication.

The task of collecting ice-core-based snow accumulation records proved more challenging than for water isotopes. Despite the existence of a large number of ice cores with annually dated stable isotope records, the number of published

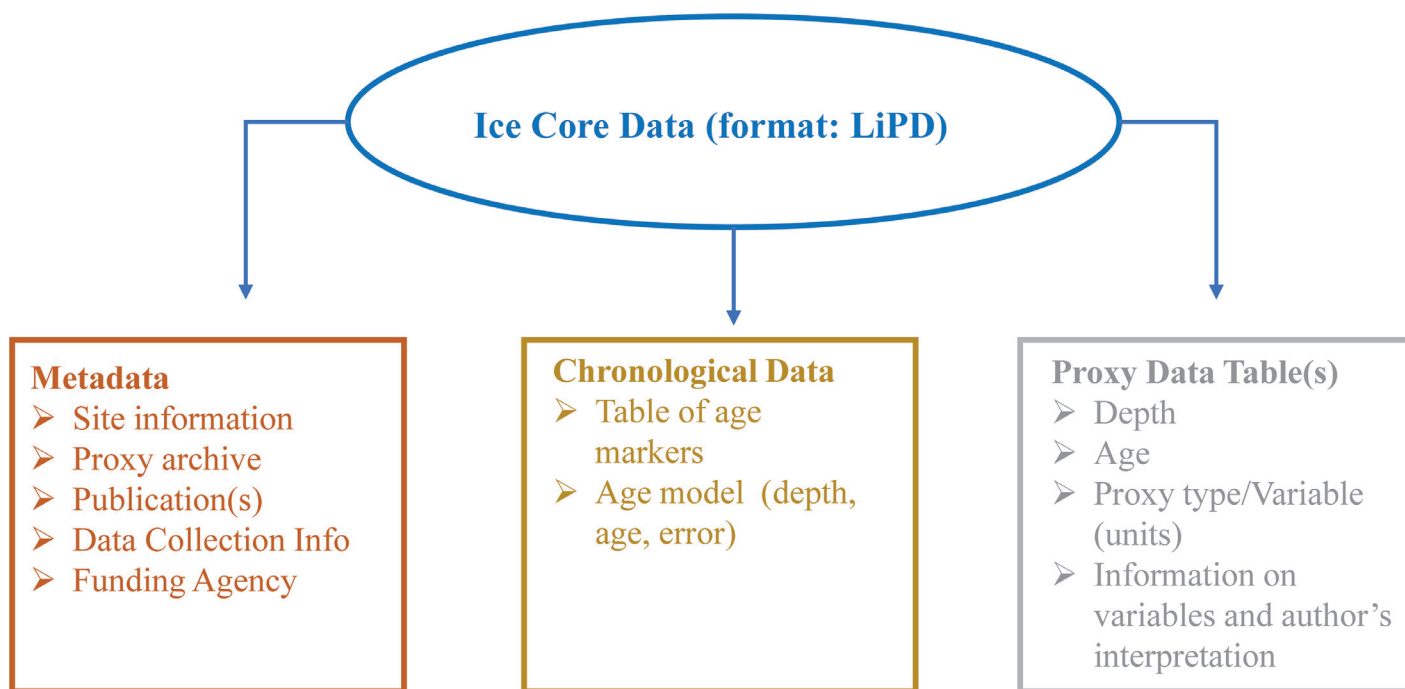


Figure 2: LiPD data container (McKay and Emile-Geay 2016) used in our regional isotopic and snow-accumulation reconstructions. This flexible framework can allocate unlimited number of climate proxy data types (sea ice, wind strengths, etc.), author interpretations and metadata fields.

snow-accumulation records is limited. Just 79 snow accumulation records were available, compared to 112 for the stable water isotopes. Twenty-two of the ice-core records submitted to the isotope database did not have a corresponding snow accumulation record, either published or publicly archived, despite the evidence that an annual depth-age scale must exist. Snow accumulation (the sum of precipitation, sublimation, melt and erosion) is the distance between dated tie-points, such as annual layers used to produce age-depth scales. This distance is corrected for compaction, based on measured density, ice thinning and flow, which can be difficult to measure in low-accumulation areas.

Another reason for the discrepancy in the number of records published may be that less scientific value is placed on snow accumulation compared to other proxies. If the additional 13 records from the East Antarctic plateau were made available, the spatial coverage in this region would have increased by 40%, while making the snow-accumulation records available for sites in the Antarctic Peninsula and Dronning Maud Land would have increased the temporal coverage in these regions from 200 to 500 years. Searching for the data was not straightforward. The 56 records that were publicly available were stored in four different archives (Fig. 1). The remaining 23 records were obtained by directly contacting original authors via email. In some cases, the ice cores were collected several decades ago and the original author was no longer working in the field. In those cases, the data were collected via third parties such as the authors of previous compilation studies or directly emailing current members of the research team. The majority of the data requested was made available, however the exercise was time-consuming, as often only the raw data was

provided and all metadata (such as dating method, thinning functions) needed to be extracted from the original publication and submitted as a new entry in the database. In accordance with the PAGES 2k and *Climate of the Past* data policies, all records had to be archived at a recognized data repository with a unique digital identity (DOI or url) prior to publication. However, given the large number of records for which this was not possible (when the original author was not able or willing to submit the data to a data center themselves), the decision was made to publish all original records in a public archive together as a single compilation, with the metadata and data citations.

Final remarks

Despite the growing number of records in public repositories and the great efforts of promoting open data, our Antarctica 2k experience pointed out that much valuable data (new and old) have not yet been transferred to public data centers. Indeed, if we look at the spatial distribution of the records included in the two compilations (Fig. 1) these are not exactly overlapping. This mismatch suggests that many datasets are still missing from public repository.

We suggest two simple actions, which are not limited to the Common Era but can be applied also to longer records. We encourage the international ice-core community to:

- archive not only new but also previously published ice-core datasets with a recognized data repository;
- adopt the flexible data container LiPD for storing multi-proxy datasets and rich metadata from ice cores (McKay and Emile-Geay, this issue) and described by the LinkedEarth ontology (Emile-Geay et al., this issue; Fig. 2).

The regional- and continental-scale temperature and snow-accumulation reconstructions carried out by the Antarctica 2k working group opened the possibility to address a longstanding question about the relationship between temperature and precipitation in Antarctica, which is one of the aims of the new CLIVASH 2k project (past-globalchanges.org/ini/wg/2k-network/projects/clivash). However, a major effort is still needed for having properly compiled and accessible records of isotopes (surface temperature), snow-accumulation rates, as well as sea-ice proxies, from all Antarctic drilling sites. The lack of available data in public repositories together with the need to increase the spatial coverage of our observations, particularly in the coastal areas, are still hampering our understanding of the recent climate variability in Antarctica.

AFFILIATIONS

¹Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Italy

²Ice Dynamics and Paleoclimate, British Antarctic Survey, Cambridge, UK

CONTACT

Barbara Stenni: barbara.stenni@unive.it

REFERENCES

- McKay NP, Emile-Geay J (2016) *Clim Past* 12: 1093-1100
- Kaufman DS, PAGES 2k special-issue editorial team (2018) *Clim Past* 14: 593-600
- PAGES 2k Consortium (2013) *Nat Geo* 6: 339-346
- PAGES 2k Consortium (2017) *Sci Data* 4: 170088
- Stenni B et al. (2017) *Clim Past* 13: 1609-1634
- Thomas ER et al. (2017) *Clim Past* 13: 1491-151