VOLUME 26 · NO 2 · NOVEMBER 2018

PAST GLOBAL CHANGES

MAGAZINE

BUILDING AND HARNESSING OPEN PALEODATA

EDITORS
John W. Williams, Alicia J. Newton, Darrell S. Kaufman and Lucien von Gunten

PAGES

futurearth

News

Host the PAGES 2021 OSM & YSM
The Open Science Meeting (OSM) and Young Scientists Meeting (YSM) are the premier events on PAGES calendar, held every four years. The 2017 events in Spain were a huge success, with approximately 900 scientists in attendance.
Be an integral part of this fantastic paleoscience community gathering. Expressions of Interest to host the next OSM and YSM, in the first half of 2021, are due 28 February 2019. All details: pages-osm.org

PAGES SSC and EXCOM news
Applications to join PAGES Scientific Steering Committee (SSC) from January 2020 are due 14 February 2019. All details: pastglobalchanges.org/about/structure/scientific-steering-committee/apply

At the end of 2018, we say goodbye and thank you to Executive Committee (EXCOM) member Pascale Braconnot (France) and SSC members Hugues Goosse (Belgium) and Yusuke Yokoyama (Japan). In January 2019, we welcome Paul Valdes (UK) and Boris Vannière (France) to the SSC and Blas Valero-Garcés (Spain) joins the EXCOM.

Congratulations to SSC member Ed Brook, who will receive the Hans Oeschger Medal at the 2019 EGU General Assembly in Vienna, Austria.

Suggest a new working group or apply for meeting support
Propose a new working group: pastglobalchanges.org/ini/wg/new-wg-proposal or apply for workshop support by 14 February 2019. This round of workshop support is an open call: pastglobalchanges.org/my-pages/meeting-support

Science Officer leaves PAGES
Soon we will be saying "Cheers and good luck" to Deputy Executive Director and Science Officer Lucien von Gunten, who has been with PAGES for eight years. Among his many contributions, he has overseen the PAGES 2k Network activities and been responsible for the *Past Global Changes Magazine*.

This issue is his final publication with us! Lucien begins a new position as Scientific Advisor at the Swiss State Secretariat for Education, Research and Innovation, with a main focus on the EU Framework Programmes for Research. Thank you, Lucien, for all you have done to advance PAGES' standing in the scientific community. Sarah Eggleston, currently at Empa in Zürich, Switzerland, joins PAGES as our new Science Officer.

Guest scientist
SSC member Darrell Kaufman joined PAGES IPO as the guest scientist from July to December 2018. Darrell worked on data stewardship activities and guest edited this magazine.

PAGES Early-Career Network (ECN)
Since launching in February, the ECN has been busy rallying early-career paleoscientists through a variety of initiatives – webinars, newsletters, regional representation and The Early Pages blog. Read on and join! pastglobalchanges.org/ecn

Data stewardship
To advance our goal of accelerating scientific discovery by facilitating open and verifiable global paleoscience, PAGES signed two new commitments concerning the preservation and reuse of the scientific data underlying the research it helps coordinate. The agreements are with ICSU-WDS and the FAIR (findable, accessible, interoperable, and reusable) guiding principles, which builds on the Coalition for Publishing Data in Earth and Space Sciences (COPDESS).

In October 2018, PAGES held its first webinar, discussing the theme of data stewardship. Presenter Nick McKay, from the Linked Paleo Data framework (LiPD), introduced and explained this new paleoscience data storage system. Watch the webinar on PAGES' YouTube Channel: youtube.com/user/PastGlobalChanges

PAGES at INQUA 2019
Have you seen the extensive list of PAGES sessions at the 20th INQUA Congress to be held from 25-31 July 2019 in Dublin, Ireland? Working groups and SSC members are well represented. Submit an abstract by 9 January 2019: pastglobalchanges.org/calendar/127-pages/1778

Help us keep PAGES People Database up to date
Have you changed institutions or are you about to move? Please check if your details are current: pastglobalchanges.org/people/people-database/edit-your-profile If you have problems updating your account, we can help. Contact pages@pages.unibe.ch

Upcoming issue of Past Global Changes Magazine
Our next magazine will be guest edited by the PALSEA working group and focuses on sea level. Although preparations are well underway, if you would like to contribute, please contact the IPO: pages@pages.unibe.ch

Calendar

LandCover6k: European Land-use at 6000BP
28-30 January 2019 – Hemmenhofen, Germany

PAGES 2k Network PALEOLINK workshop
6-8 February 2019 – Murcia, Spain

Extremes Integrative Activity workshop
18-20 February 2019 – Koblenz, Germany

VICS workshop: The Common Era and beyond
13-16 April 2019 – Cambridge, UK

DAPS 2nd workshop
29-31 May 2019 – College Park, USA

pastglobalchanges.org/calendar

Featured products

Aquatic Transitions
The group questions if we can detect ecosystem critical transitions and signals of changing resilience from paleo-ecological records (2018, *Ecosphere* 9).

C-PEAT
Angela Gallego-Sala et al. examine how peatland carbon sink is to behave under future climate warming scenarios (2018, *Nat Clim Change* 8).

Floods
The group reviews historical, botanical, and geological archives with a focus on the recording mechanisms of flood information, the historical development of the methodological approaches and the type of information that those archives can provide (2018, *WIREs Water*).

PEOPLE 3000
Members published a paper which looks at energy consumption reaching back 10,000 years and across four continents (2018, *PNAS* 115).

QUIGS
Members find that the Last Interglacial was punctuated by a series of century-scale arid events in southern Europe and cold water-mass expansions in the North Atlantic (2018, *Nat Commun* 9).

SISAL
The group published the first overview of its database's contents and structure (2018, *Earth Sys Sci Data* 10).

Warmer Worlds
Members of PAGES Warmer Worlds Integrative Activity stress the urgency of reducing CO_2 emissions to avert major environmental damages (2018, *Nat Geo* 11).

Cover

Snapshot of open paleodata available from WDS repositories PANGAEA, NOAA National Centers for Environmental Informatics, and Neotoma Paleoecology Database, as of October 2018
Data volumes continue to grow, powered by the ongoing generation of new data and the gathering and curation of these data into open, community-stewarded data resources. These open data are enabling new advances in geovisualization, data analysis, and data-model assimilation. For comparison, see the "Paleodata" PAGES news issue from 1998 (doi.org/10.22498/pages.6.2). Image designed and produced by Tanya Buckingham of the Cartography Lab at the University of Wisconsin-Madison, USA.

HOST THE 2021 PAGES OPEN SCIENCE MEETING (OSM) AND YOUNG SCIENTISTS MEETING (YSM)

- The Open Science Meeting and Young Scientists Meeting are PAGES' premier events, held every four years.
- Expressions of interest to host the next OSM and YSM meetings are due **28 February 2019**.
- All details: **pages-osm.org**

Meet our Guest Editors

Data stewardship is a topic concerning the whole range of paleoscience stakeholders. To do justice to this diversity, we invited guest editors representing three of the key actors - the data archives, the publishers and the community.

John (Jack) Williams[1] is a professor in the Department of Geography and former Director of the Center for Climatic Research at the University of Wisconsin-Madison, USA. Jack currently serves as the Chair of the Leadership Council for the Neotoma Paleoecology Database (neotomadb.org). Jack also serves on the EarthRates Steering Committee and, with Mark Uhen and others, is launching the Earth-Life Consortium (earthlifeconsortium.org) as a non-profit organization dedicated to making paleobiological and paleoecological data free and easy to access. Jack's research emphasizes the spatial and temporal responses of plant species and communities to large, novel, and abrupt environmental changes at local to continental scales. This work is indebted to the gathered data and knowledge of decades of palynologists and paleoecologists. Much of his work in recent years has focused on powering the next generation of high-quality macro-scale science by building, supporting, and interlinking open paleodata resources such as the Paleobiology Database, NOAA-Paleoclimatology, and the Neotoma Paleoecology Database.

Alicia Newton[2] graduated from the University of South Carolina, where she used foram geochemistry to reconstruct ocean conditions in the Indo-Pacific Warm Pool. She spent 11 years as an editor for *Nature Geoscience*, where she handled paleoceanography among other topics. She helped to roll out policies on data availability and FAIR data at the journal and aided authors in making their data available at the time of publication. She is currently the Director of Science and Communications at the Geological Society of London, UK.

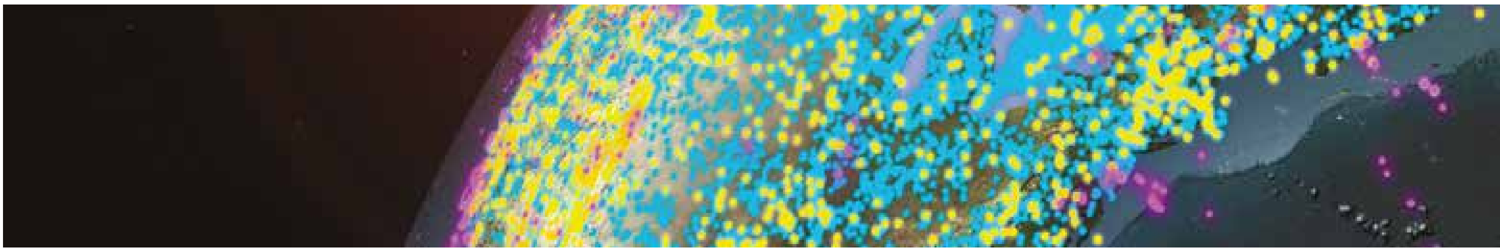
Darrell Kaufman[3] (Regents' Professor, School of Earth and Sustainability, Northern Arizona University, USA) is a Quaternary geologist and paleoclimatologist specializing in lake sediments, geochronology, and Arctic paleo environments. He is a member of the PAGES Scientific Steering Committee and is currently a guest scientist at the PAGES International Project Office and the Oeschger Centre for Climate Change Research in Bern, Switzerland. He has a special interest in PAGES' Data Stewardship Integrative Activity, an initiative that cross-cuts all PAGES working groups. He is motivated by the scientific benefit of pooling paleodata, and by the urgent need to curtail the scientific loss of valuable data resources. He recently led an open-data implementation pilot involving two special issues of the journal *Climate of the Past*, which provide examples of data stewardship for those aspiring to integrate their data into a larger global network of similar results. He is collaborating with several PAGES working groups to develop global paleo datasets, and is working with the creators of the cyber-based infrastructure that enables data sharing and analysis.

AFFILIATIONS
[1]Department of Geography, University of Wisconsin-Madison, USA. Twitter: @IceAgeEcologist
[2]Geological Society of London, UK. Twitter: @G_ruber
[3]School of Earth and Sustainability, Northern Arizona University, Flagstaff, USA. Twitter: @DKaufman1

doi.org/10.22498/pages.26.2.48

New data-availability procedures echo PAGES' long-standing commitment

PAGES Scientific Steering Committee*



The last time *Past Global Changes Magazine* highlighted the power of open paleodata was exactly 20 years ago. The cover page of the winter 1998 issue (PAGES IPO 1998; Fig. 1) featured an editorial by the PAGES Scientific Steering Committee (1998) entitled, "*Responsibilities of Data Sharing and Data Use*." Our predecessor Scientific Steering Committee members anticipated that open paleodata would fuel discoveries in past global changes. Their initiative has grown into the ongoing PAGES Data Stewardship Integrative Activity[1], and led to our recent alliance[2] with other international scientific organizations in efforts to make data publicly accessible. The 1998 editorial laid the groundwork for the first PAGES data policy by encouraging the transfer of "the highest possible proportion of existing and new, high-quality data into public domain databases…" so that access to data "…is truly easy and open to all."

We are announcing updated and expanded procedures[3] for making data available, with the goal of maximizing the long-term scientific benefit of the data generated as part of all PAGES-related activities, while fulfilling PAGES' obligation to its funders. The new PAGES guidelines build on the earlier policy and are reinforced by the FAIR (findable, accessible, interoperable, and reusable) data stewardship principles (Wilkinson et al. 2016), which have been endorsed by scientific organizations globally. They focus on publications and are adapted for paleoscience from the Author Guidelines[4] that are now being implemented by all major publishers of Earth and Space Sciences, as motivated by the Enabling FAIR Data Project[5]. They have benefited from input from managing and chief editors of paleo journals, repositories, and the community. For example, the new procedures now provide guidelines on the use of data embargoes, a topic that emerged from the paleoscience community open discussion[6] as part of the PAGES 2k open-paleodata implementation pilot (Kaufman and PAGES 2k special-issue editorial team 2018).

The 1998 editorial also called for new "realistic ways of both recognizing and rewarding the generosity of all who submit their data." The importance of crediting data generators and the value of making data reusable for future scientists is now being addressed with the advent of data citations and journals dedicated to data products. PAGES encourages the use of data citations[7], which are analogous to standard bibliographic citations, but give explicit credit to data producers, with greater exposure and citation of their work. For large-scale synthesis products, PAGES promotes the use of data-oriented publications as a means to including many data generators in the production of value-added, high-visibility data products, with inclusive authorship.

In addition to new avenues for crediting data generators, attitudes toward open data have evolved over the past two decades, and they evolve with individual's careers. Unfortunately, data that are not properly curated are liable to be lost to subsequent reuse; the time comes too quickly when the data that were made "available upon request" may never be discovered or used in future studies; a true loss for all. Now, with new means for making data available, the rewards, including higher citation rates and other benefits described by Newton (p. 52), are proportionally greater. Scientists, especially those early in their career, seek to increase the impact and recognition of their research by facilitating the reuse of their results. Indeed, according to the survey conducted by the PAGES Early-Career Network (Koch et al., p. 54), 95% of the 163 non-tenured respondents feel that data sharing is advantageous to their careers.

Figure 1: 1998 winter issue of *PAGES news*, with the PAGES SSC editorial on data stewardship.

We recognize that data stewardship requires substantial effort, but we are convinced that the benefits outweigh the (perceived) costs. It is increasingly obvious that the future of our field depends on robust and widely adopted data-sharing practices and procedures. We appreciate the community's foresight and dedication to data that are open and reusable, while curtailing the loss of valuable data.

*PAGES SCIENTIFIC STEERING COMMITTEE
A. Asrat (Addis Ababa University, Ethiopia); **P. Braconnot** (Lab. des Sciences du Climat et de l'Environnement, Gif-Sur-Yvette; **E. Brook** (Oregon State University, USA); **E. Capron** (British Antarctic Survey, UK); **C. Chiesi** (University of São Paulo, Brazil); **M.N. Evans** (University of Maryland, USA); **P. Gell** (Federation University Australia, Australia); **L. Gillson** (University of Cape Town, South Africa); **H. Goosse** (Université catholique de Louvain, Belgium); **Z. Jian** (Tongji University, China); **D.S. Kaufman** (Northern Arizona University, USA); **M. Kucera** (University of Bremen, Germany); **K. Meissner** (University of New South Wales, Australia); **W. Tinner** (University of Bern, Switzerland); **B.L. Valero-Garcés** (Spanish National Research Council, Spain); **Y. Yokoyama** (University of Tokyo, Japan).

CONTACT
Darrell S. Kaufman: Darrell.Kaufman@nau.edu

REFERENCES

PAGES IPO (1998) PAGES news 6(2)

PAGES Scientific Steering Committee (1998) PAGES news 6: 1-2

Kaufman DS, PAGES 2k special-issue editorial team (2018) Clim Past 14: 593-600

Wilkinson MD et al. (2016) Sci Data 3: 160018

LINKS

[1] pastglobalchanges.org/ini/int-act/data-stewardship

[2] pastglobalchanges.org/news/2062-pages-data-agreements-aug-18

[3] pastglobalchanges.org/my-pages/data

[4] copdess.org/enabling-fair-data-project/author-guidelines/

[5] copdess.org/enabling-fair-data-project/

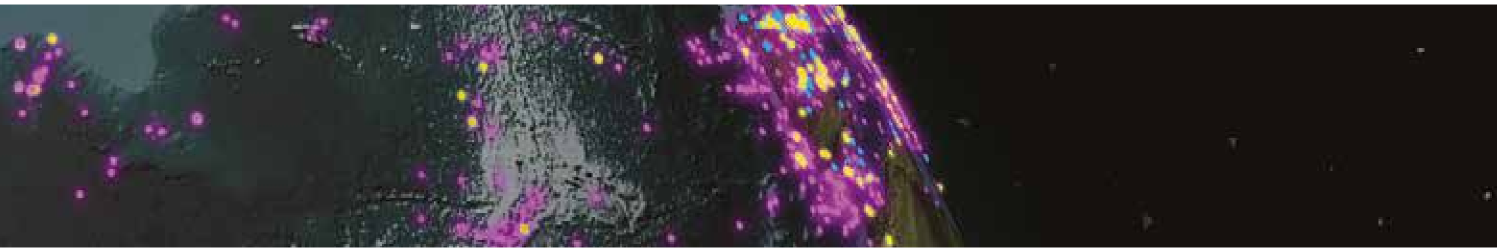
[6] clim-past.net/14/593/2018/cp-14-593-2018-discussion.html

[7] force11.org/datacitationprinciples

doi.org/10.22498/pages.26.2.49

Building and harnessing open paleodata

John W. Williams[1,2], D.S. Kaufman[3], A. Newton[4,5] and L. von Gunten[6]



Open data in the paleogeosciences have a long and fruitful history. Many of the primary open-data resources in the paleoenvironmental sciences are now at least two decades old, including the NOAA World Data Center for Paleoclimatology (Gross et al., p. 58), PANGAEA (Diepenbroek, p. 59), Paleoclimate Modelling Intercomparison Project (PMIP, Peterschmitt et al., p. 60), and the Paleobiology Database (Uhen et al., p. 78), all founded in the 1990s, with others, such as the Neotoma Paleoecology Database (Grimm, p.64), tracing their roots to constituent databases from this era and to influences spanning the last century. Indeed, this special issue can be viewed as a 20th-anniversary celebration of the 1998 "*Paleodata*" issue of PAGES news (the former name of *Past Global Changes Magazine*) that established many of the advances reviewed here (PAGES IPO 1998).

The history of open data in the paleogeosciences is long because the scientific motivation is so clear and unambiguous. In the large, complex, and ever-changing Earth system, scientific insight requires the open availability and close integration of multiple observational systems with Earth system models, to better understand the past and present, and better forecast the future (Crucifix 2012; Dietze et al. 2018). And, as the Great Acceleration continues (Steffen et al. 2015), such efforts have increased urgency; the past offers a uniquely important set of model systems for the strange new world of the coming decades.

Over these last two decades of open data, much has changed. The dividing line between "data generator" and "data user", so apparently bright in the 1990s (PAGES Scientific Steering Committee 1998), has blurred as a new generation has arisen, with cross-over expertise in data generation, synthesis, and modeling. The information revolution races on, with the data sciences emerging both as a distinct academic discipline (Blei and Smyth 2017) and as a key employment opportunity for many scientists. Access to open-data resources is now essential to career advancement for early-career scientists, while lack of access to training is a key barrier (Koch et al., p. 54).

Contributing one's data to open-data resources, once largely voluntary, is now required by most journals, funders, and professional societies (Newton, p. 52; Belmont Forum, p. 56). The bar has been raised for open-data resources, to ensure that they meet the FAIR standards of Findable, Accessible, Interoperable, and Reusable (PAGES Scientific Steering Committee, p. 48; Gross, p. 58). New funding initiatives are being launched to increase the power and interoperability of existing data resources (e.g. NSF's EarthCube; Belmont Forum, p. 56), leading to new and flexible data standards and software that leverage and link open-data resources (Uhen et al., p. 78; McKay and Emile-Geay, p. 71). New geovisualization approaches such as Flyover Country, using open data and mobile technologies, are bringing paleodata to new audiences (Myrbo et al., p. 74). And, our understanding of data is changing as well, as we recognize that open data require ongoing curation and improvement, supported by community-curated data resources and linked networks of data stewards (Williams et al., p. 50).

These advances in open-data systems are opening up new scientific frontiers. Data-model assimilation, in which paleoenvironmental inferences from data and models are closely integrated, weighted by uncertainty, are active fields in paleoclimatology (Hakim et al., p. 73) and paleoecology (McLachlan and the PalEON Project, p. 76). Computer scientists are experimenting with artificial-intelligence approaches to age-model development (Bradley et al., p. 72) and extracting geological knowledge from the peer-reviewed literature (Marsicek et al., p. 70). Open paleodata have reached new audiences, as biogeographers and macroecologists combine the fossil record with big-data genetic repositories to study the processes governing the distribution and diversity of life (Fordham and Nogues-Bravo, p. 77), and as archaeologists bring big data to bear on the interplay between humans and the environment (Kohler et al., p. 68).

More needs to be done. Many key data remain "dark", requiring inordinate effort to gather and synthesize (Stenni and Thomas, p. 66). The paleoscience communities need to commit to conventions for reporting data and essential metadata, with shared adoption by scientists, data resources, publishers, and funding agencies. Established open-data resources need commitments of sustained support from funding agencies, with opportunities to build new data resources or extend existing data models to serve new kinds of data and science. The recent advances in assigning digital object identifiers (DOIs) to datasets needs to be more fully leveraged so that data generators are appropriately credited for data use. Scientific data services are needed that better streamline the passing of data from individual labs to community data resources. And, most of all, we need better integrated training programs in paleoscience and data science, to train the next generation of cross-over scientists.

In short, these are exciting and changing times. This special issue is more progress report than final authority. Nevertheless, we hope that the articles enclosed will provide useful information about the latest updates from some of the major open-data resources in the paleogeosciences, the efforts to build new resources and interlink existing resources, the emergence of new software and science powered by open data, and the ever-evolving interplay among cultural norms, technological advances, and scientific discovery.

AFFILIATIONS

[1]Department of Geography, University of Wisconsin-Madison, USA
[2]Neotoma Paleoecology Database
[3]School of Earth and Sustainability, Northern Arizona University, Flagstaff, USA
[4]Nature Geosciences Editorial Office, London, UK
[5]Geological Society of London, UK
[6]PAGES International Project Office, Bern, Switzerland

CONTACT

John (Jack) W. Williams: jww@geography.wisc.edu

REFERENCES

Blei DM, Smyth P (2017) PNAS 114: 8689-8692

Crucifix M (2012) Quat Sci Rev 57: 1-16

Dietze MC et al. (2018) PNAS 115: 1424-1432

PAGES IPO (1998) PAGES news 6(2)

PAGES Scientific Steering Committee (1998) PAGES news 6(2): 1-2

Steffen W et al. (2015) The Anthropo Rev 2: 81-98

doi.org/10.22498/pages.26.2.50

Building open data: Data stewards and community-curated data resources

John W. Williams[1,2], D.S. Kaufman[3], A. Newton[4,5] and L. von Gunten[6]

Open data advance the pace of discovery in the paleogeosciences. Community-curated data resources and data stewards, together, offer a solution for jointly maximizing the volume and quality of open data. All can assist, at both individual and institutional levels.

Open data, long a good idea, are now mission-critical to advancing and accelerating the pace and breadth of discovery in the paleogeosciences. We seek to understand the past dynamics of the Earth system and its interacting subsystems, across a wide range of timescales, and to use this knowledge to inform society in a new era of global change. However, the scale of the system is too vast, and the volume and variety of data too large, for any single investigator or team to be able to integrate it. Open scientific data, gathered into curated data resources, are essential to integrating this information at scales beyond the capacity of any single team. Such data can then support big-data applications, where inferential power is proportional to data size and richness, such as machine learning, proxy system modeling (Dee et al. 2016), and data-model assimilation (Hakim et al. 2016). Ultimately, the goal is to form an open architecture of scientific data as complex, deep, and interlinked as the Earth system itself.

The benefits of open data extend beyond scientific objectives. For individual investigators, open-data resources provide services of data archival and increasing data visibility. In the genetics literature, papers with published data have a 9% higher citation rate than similar studies without published data (Piwowar and Vision 2013). Open data enable interdisciplinary research and knowledge exchange across disciplines. Open data also empower early-career scientists and scientists from the Global South, enable transparency and reproducibility, and return the fruits of publicly and privately funded research to the public domain (Soranno et al. 2014).

Multiple initiatives are underway to support and encourage best practices in open data. Publishers have launched the FAIR initiative: data must be findable, accessible, interoperable, and reusable (Wilkinson et al. 2016). Funding agencies are setting firmer standards for publicly funded data (National Science Foundation 2018). Multiple authors have called for open data (Soranno et al. 2014; Schimel 2017; Kaufman and PAGES 2k special-issue editorial team 2018). Open code and software are becoming the norm, facilitated by open-source languages (e.g. R, Python), platforms for sharing code (e.g. GitHub, BitBucket), and notebooks for sharing scientific workflows (e.g. RMarkdown, Jupyter).

Nonetheless, both cultural and technical barriers remain (Heidorn 2008), with only 25% of geoscientific data submitted to open-data repositories (Stuart et al. 2018). Most scientists are willing to share data once published, but many lack the time to prepare datasets and metadata for open publication, or the training and tools to do so efficiently. Some communities lack established data standards and repositories, with particular difficulties in finding an appropriate home for terabyte-scale datasets. Systems for data citation and provenancing remain underdeveloped, so it is hard for scientists to receive the credit due for data publication. Data curation adds value to open data, thereby navigating the big-data challenge of maximizing both data volume and veracity (Price et al. 2018), but effective data curation requires dedicated time by experts, which needs to be recognized and rewarded.

These challenges to open data are real but tractable and can be resolved through a combination of cultural and technological solutions.

One key emerging solution is the combined rise of community-curated data resources and linked networks of data stewards (CCDRs; Figs. 1, 2). CCDRs serve as loci where experts can contribute and refine data, establish data standards and norms, and ensure data quality. If open data are a commons, then CCDRs provide a governance framework for managing the commons. In this framework, data stewards (or data editors, see Diepenbroeck, this issue) are positions of service and leadership that are equivalent in function and prestige to journal editors, dedicating a portion of their time and expertise to ensure that published data are of high quality and meet community standards. The broader cultural goal is to establish norms of data openness – in which we commit to contributing our data to community data resources – and data stewardship, in which

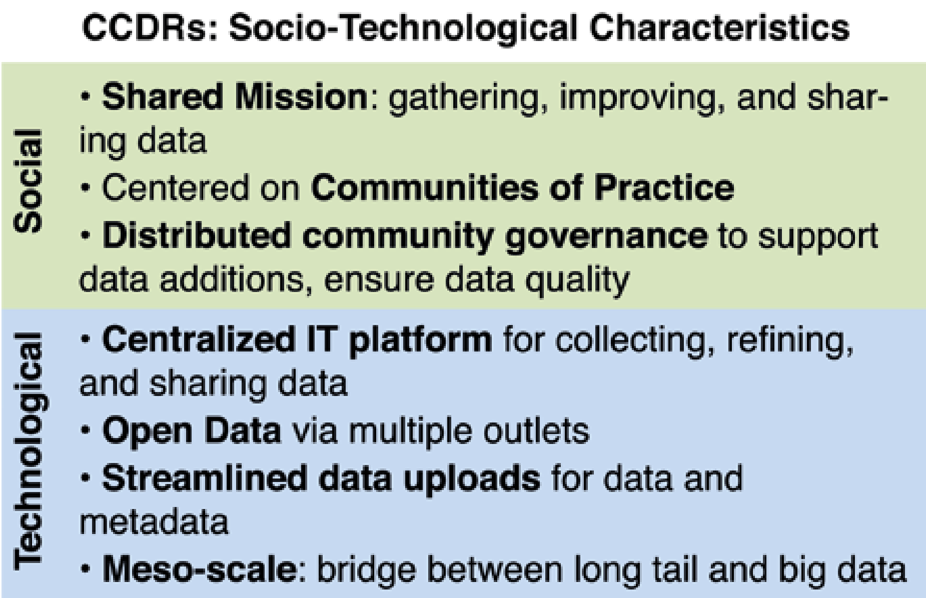


Figure 1: Community-curated data resources (CCDRs) as both social and technological solutions for supporting open data. Social characteristics include a shared scientific mission, communities of practice centered on domain experts, and governance mechanisms that facilitate participation and leadership by a broad and diverse base of experts. Technological characteristics include a central platform with support for uploading, curating, and providing data; and systems that facilitate open data access and data uploads. Because CCDRs are closely tied to their expert communities, they tend to be meso-scale intermediaries between individual data generators and big-data initiatives.

we commit to adding value to community data resources on an ongoing basis.

Multiple related initiatives are underway to build open and high-quality community data resources, stewarded by experts. Publishers have created journals specifically devoted to data publication (Newton, this issue). In paleoclimatology, PAGES 2K has established pilot examples of open data and data stewardship for global-scale data syntheses (PAGES 2k Consortium 2017). The LiPD and LinkedEarth ontologies provide flexible data standards for paleoclimatic data, with editors able to approve ontology extensions (McKay and Emile-Geay, this issue). The Neotoma Paleoecology Database has established a system of member virtual constituent databases, each with data stewards charged with prioritizing data uploads and defining variable names and taxonomies (Williams et al. 2018). The Paleobiology Database uses data authorizers to ensure quality data uploads (Uhen et al. 2013 and this issue). Some efforts focus on curating primary measurements and others on higher-level derived inferences (McKay and Emile-Geay, this issue).

Technologically, the broad need is to move open-data resources from systems of record to systems of engagement (Moore 2011), in which we move beyond models of submitting datasets to static data repositories to systems that support crowdsourcing and ongoing efforts to publish and improve data. Such infrastructure must support data discovery, archival, citation, tracking, annotation, and linking. Flexible and extensible data models are needed to support both existing and new proxies (McKay and Emile-Geay, this issue). Controlled vocabularies and common semantic frameworks are needed to tame the heterogeneity of proxy measurements. Systems for data annotation are needed to flag and correct data errors. Systems for microattribution and provenancing are needed to track data usage from initial publication to subsequent incorporation into broad-scale data syntheses. Assigning DOIs to datasets is a first step; subsequent steps are to include these DOIs in all future publications to appropriately credit data generators. Journals and citation indices will need to adopt linked data systems, tracking data usage, with ability to link to thousands of individual records, so as to avoid arbitrary limits caused by fixed limits to the number of references. New tools are needed that streamline the collection and passing of data from point of collection to data resource. Because effort is the main barrier to open data, good data management should be maximally automated.

For open data to power the next generation of scientific discovery, we must all pitch in. Scientists must commit to making their data available in open public repositories, join governance, and serve as data stewards. Publishers, as they adopt FAIR data standards, should endorse and support open community data resources that meet these standards. Funding agencies should support development of open-data standards for data types where none yet exist and provide modest but sustained support for open-data resources, under the logic that costs of supporting CCDRs are cheap relative to costs of regenerating primary data. We must launch data-mobilization campaigns that are science driven (e.g. PAGES 2k Consortium 2017), using these campaigns to prioritize rescues of dark data. Professional societies should establish mechanisms to endorse community data standards and open platforms and, where possible, provide support via a portion of membership dues. Just as professional journals were the mainstay of communicating scientific knowledge in the 19th and 20th centuries, open, high-quality community data resources will be a mainstay of communicating and advancing knowledge in the coming decades.

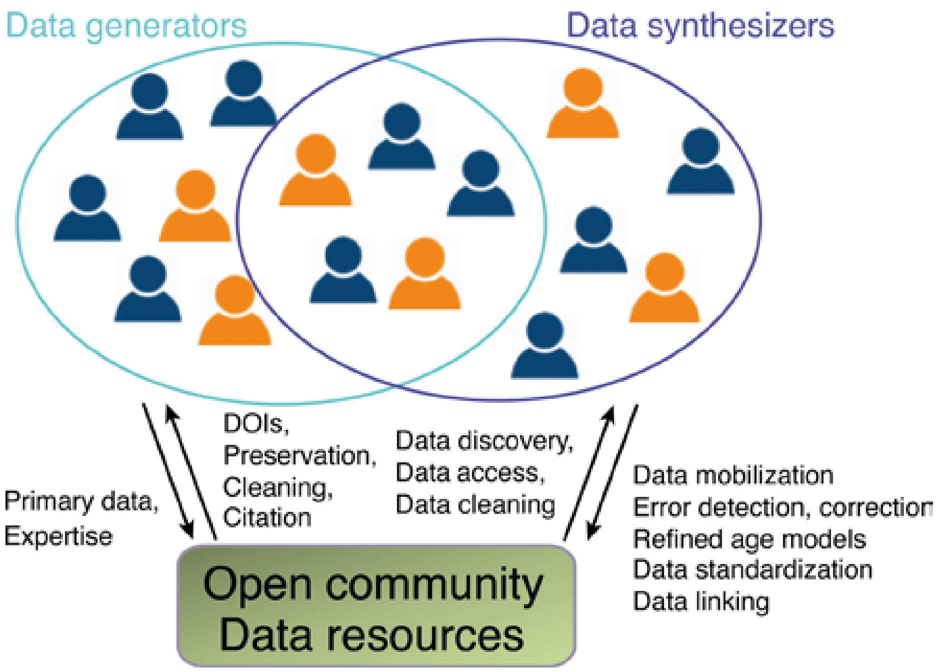


Figure 2: Paleodata CCDRs and their relationships of engagement with their overlapping research communities of data generators, stewards, and synthesizers. Data generators provide the primary data to CCDRs and receive in return DOIs for data citation and tracking and assistance in meeting community data standards. Synthesizers benefit from CCDRs through the services of improved data discovery, access, and cleaning, while returning to CCDRs the services of data mobilization for dark data, detection and correction of errors in CCDRs, updated and improved age models, and assistance in linking CCDRs with other data resources. Data stewards (orange), drawn from both communities, support data curation and ensure that community data norms are met, akin to the role of editors in peer-reviewed journals.

AFFILIATIONS
[1]Department of Geography, University of Wisconsin-Madison, USA
[2]Neotoma Paleoecology Database
[3]School of Earth and Sustainability, Northern Arizona University, Flagstaff, USA
[4]Nature Geosciences Editorial Office, London, UK
[5]Geological Society of London, UK
[6]PAGES International Project Office, Bern, Switzerland

CONTACT
John (Jack) W. Williams: jww@geography.wisc.edu

REFERENCES

Dee SG et al. (2016) J Ad Model Earth Sy 8: 1164-1179

Hakim GJ et al. (2016) J Geophys Res Atmos 121: 6745-6764

Heidorn PB (2008) Libr Trends 57: 280-299

Kaufman DS, PAGES 2k special-issue editorial team (2018) Clim Past 14: 593-600

Moore G (2011) Systems of engagement and the future of enterprise IT. A sea change in enterprise IT. AIIM, 14 pp

National Science Foundation (2018) Data and Sample Policy. nsf.gov/geo/geo-data-policies/ear/ear-data-policy-apr2018.pdf

PAGES 2k Consortium (2017) Sci Data 4: 170088

Piwowar HA, Vision TJ (2013) PeerJ 1: e175

Price GJ et al. (2018) Nature 558: 23-25

Schimel D (2017) Front Ecol Environ 15: 175

Soranno PA et al. (2014) BioScience 65: 69-73

Stuart D et al. (2018) Practical challenges for researchers in data sharing. SpringerNature, 17 pp

Uhen MD et al. (2013) J Vert Paleontol 33: 13-28

Wilkinson MD et al. (2016) Sci Data 3: 160018

Williams JW et al. (2018) Quat Res 89: 156-177

doi.org/10.22498/pages.26.2.52

Open data and the publishing landscape

Alicia J. Newton[1,2]

Every research paper is underlain by data. But, until relatively recently, the accessibility and archiving of this data has been an afterthought to the published paper. Technological advances and efforts to increase reproducibility have pushed data availability to the forefront.

Papers in the paleosciences have always been data rich: Emiliani's (1955) work illustrating glacial-interglacial cycles relied on twelve cores sampled at 10 cm intervals. And from CLIMAP (Climate: Long range Investigation, Mapping, and Prediction) to PAGES 2k Network, paleoclimatologists have also been quick adopters of big-data approaches, combining individual records to generate global maps of temperature change through time. The value of these types of efforts is immediately recognizable by the wider paleo community. However, the open data practices that support these efforts have grown more slowly.

Today, the data that underlie the CLIMAP reconstruction are available from a variety of repositories found by a simple internet search. However, at the time of the compilation in 1981, files would have been shared peer to peer, with some smaller data tables contained within publications.

Surprisingly, peer-to-peer sharing remains a prominent mode of data sharing, with 31% of Earth scientists opting not to archive data in a repository or include data in supplementary materials of publications (Stuart et al. 2018).

Peer-to-peer sharing is quick, but has a number of downsides. On a practical level, data that isn't archived may be unprotected. Many scientists still store data on personal or external hard drives, where it is vulnerable to theft, format or program obsolescence, or simply an errant cup of coffee (Baynes 2017). On a broader level, requiring personal outreach to obtain data can hinder scientists with fewer connections or who face a language barrier. And data stored in this manner may be lost when scientists retire or leave academia.

In the paleosciences, and geosciences more broadly, data archiving in open repositories takes on an additional importance: it can be exceedingly expensive to obtain samples through means such as ocean or ice-core drilling, and materials such as meteorites or certain fossils can be extremely rare. And some samples may prove irreplaceable as material is lost through erosion, land-use changes, and as glaciers melt. As signatories to the Coalition on Publishing Data in the Earth and Space Sciences (COPDESS) Statement of Commitment (copdess.org/statement-of-commitment), publishers have recognized this importance.

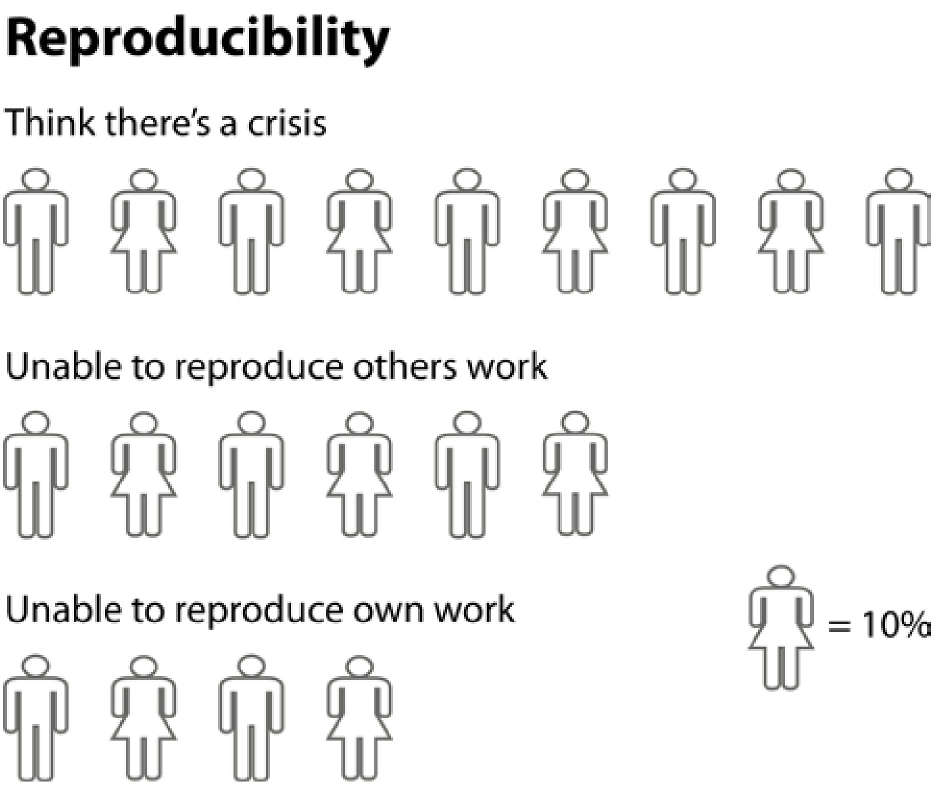


Figure 1: Respondents to a survey of 1,500 scientists raised substantial concerns about the reproducibility of the published literature, and reported their own experiences with failure to reproduce results (Baker 2016). Open data is one avenue being explored to help increase confidence in the scientific record. Image credit: Edwyn Mayhew.

Why open data?

In 2016, 90% of researchers surveyed by *Nature* raised major concerns about the reproducibility of the scientific record, with few people convinced that all of the published literature would be reproducible (Fig. 1; Baker 2016). In the Earth and environmental sciences, about 40% of respondents were unable to reproduce even their own work in at least one instance; over 60% were unable to reproduce the findings of others. Increased openness of data, methods, and code can help improve confidence in the scientific record.

Geoscientists certainly recognize the importance of data sharing, with 69% of Earth scientists making their data available in a repository or supplementary materials (Stuart et al. 2018). This movement towards data availability is driven by a growing recognition that making supporting data open offers benefits for both data producers and the broader scientific endeavor (Schmidt et al. 2016). Specifically, data sharers are motivated by the desire to help accelerate scientific research, and also to increase the visibility and dissemination of their research output (Stuart et al. 2018). Intriguingly, the survey found that funder and publisher requirements were not as strong of an incentive to release data.

But is available data always open data? In the geosciences, 28% of respondents only made data available in the electronic supplementary materials (Fig. 2). Whether or not this material sits behind a paywall varies by publisher: *Nature Geoscience* and the Nature Research journals make this material free to read, but other journals require a subscription for access. The format and content of the supplementary-data

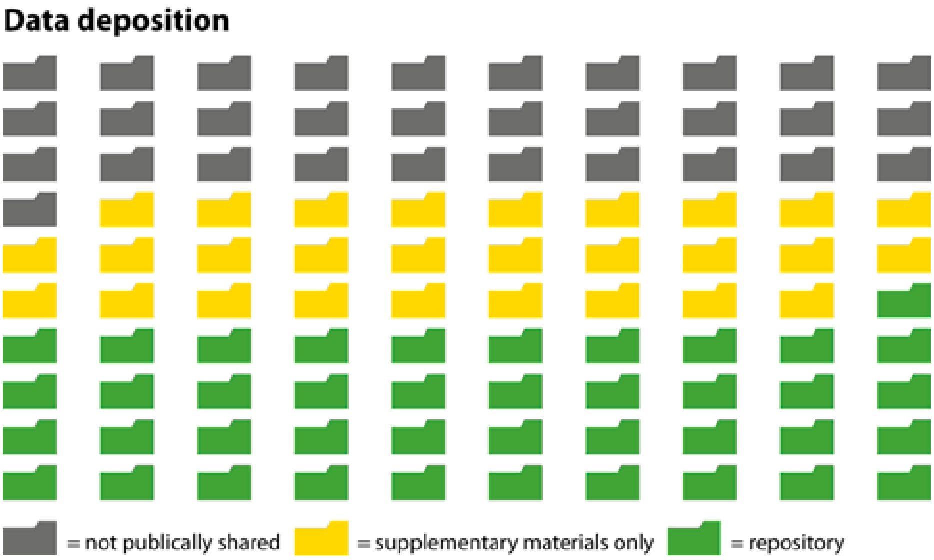


Figure 2: How discoverable is the data behind a paper? Stuart et al. (2018) surveyed 365 Earth scientists about their experiences and if and how they made the data associated with their work available. Each folder represents 1% of the survey response. Image credit: Edwyn Mayhew.

tables may also be less than ideal, with pdf tables not always easy to import into other software.

Springer Nature has started a trial in which electronic supplementary materials from articles published in BioMed Central and Springer Open journals is hosted on Figshare. These files are freely accessible and uniquely identifiable with a separate DOI, helping the data behind a paper to find its own audience (Hyndman 2016).

Recognition and reward

Beyond altruism and a desire to contribute to scientific advances, there are other benefits for researchers who make their data widely available. In *Paleoceanography*, articles that were published alongside publicly-available datasets saw a 35% greater citation rate than the journal average (Sears 2011). Across all disciplines, data availability provides a citation boost between 9 and 50% (Baynes 2017).

The rise of peer-reviewed data journals helps to provide credit for data generators, beyond a traditional scientific publication. Journals like *Scientific Data* and *Earth System Science Data* publish "data descriptors". These articles describe the collection and processing of a dataset that has been released through a public repository. The descriptors provide sufficient metadata and related information to allow for easy use of the data, but refrain from interpretation and extensive analysis. Data descriptors also can accompany a traditional scientific publication, and can allow for an expanded dataset to be released: for instance, $\delta^{13}C$ data that was collected alongside oxygen isotopes but not featured in the interpretation or additional parts of a record that were generated but not the focus on the paper. In these instances, the data descriptor can have a different lead author than the main paper, perhaps giving due credit to a student researcher who led the data collection but played a smaller role in the interpretation.

Data-descriptor papers can also serve as a way to release and promote the reuse of datasets that might otherwise live in a proverbial desk drawer: data from student summer projects, null results, or the never-written up thesis chapter can all be released for others to work from and build upon. In these cases, the data generators can receive appropriate recognition for their work – and potentially the reward of citations of the data descriptor and data set – even if the interpretation of the data might not be sufficient to warrant a traditional publication.

Into the future

In 2015, COPDESS released a statement of commitment, which was signed by most Earth and environmental science publishers and data repositories. Signatories from the publishing side agreed to promote the use of appropriate community repositories to their authors, and direct authors to relevant resources, for instance through lists maintained at the COPDESS website. The statement also encouraged publishers to develop clear statements about requirements for data availability. The Nature Research journals have long required authors to make materials, data, and code available without undue qualification. Nature Research also encourages authors to freely release data through repositories (nature.com/authors/policies/availability.html). Data-availability statements, which are now available to readers without a subscription, tell readers how to access the data reported in the manuscript, as well as any previously published data used in the analysis (Nature 2016; Hrynaszkiewicz et al. 2016). Code-availability statements require authors to report whether any code associated with the work is accessible.

Of course, much of this data still remains in supplementary information (Fig. 2), and may be only partially accessible, or lacks the essential metadata and standardization that would be provided by curators at a repository. Led by AGU, some signatories to the original COPDESS statement are addressing this concern through the Enabling FAIR Data Project. This project, which is supported by Nature Research and other publishers, will support authors to make sure that the data behind their publications are Findable, Accessible, Interoperable, and Reusable (FAIR; Wilkinson et al. 2016). Importantly, the National Computational Infrastructure of Australia is also supporting the project, providing the expertise required to start to tackle the terabyte-sized elephant in the room that is model output.

Although these and other challenges remain, the combined efforts of funders, publishers, repositories, and open-data advocates are ushering in a new era of data openness. Open data helps ensure the integrity of the scientific record, while new metrics and venues ensure that data generators are recognized and rewarded for their work. And the community stands to benefit as well, as increasingly easy data access facilitates powerful big-data approaches to understanding past environments.

AFFILIATIONS
[1]Nature Geoscience, Nature Research, London, UK
[2]Now at: The Geological Society of London, London, UK

CONTACT
Alicia J. Newton: aliciajillnewton@gmail.com

REFERENCES

Baker M (2016) Nature 533: 454-454

Baynes G (2017) In: The state of open data 2017. Holtzbrinck Publishing Group, 17-19

Emiliani C (1955) J Geol 63: 538-578

Hrynaszkiewicz I et al. (2016) Standardising and harmonising research data policy in scholarly publishing. bioRxiv, 7 pp

Hyndman A (2016) New partnership with Springer Nature to make research more accessible. Figshare blog

Nature editorial staff (2016) Nature 537: 138

Schmidt B et al. (2016) PLOS one 11: e0146695

Sears JRL (2011) Data sharing effect on article citation rate in paleoceanography, AGU Abstract IN53B-1628

Stuart D et al. (2018) Practical challenges for researchers in data sharing. Figshare, paper

Wilkinson MD et al. (2016) Sci Data 3: 160018

doi.org/10.22498/pages.26.2.54

Open-data practices and challenges among early-career paleo-researchers

Alexander Koch[1], K.C. Glover[2], B. Zambri[3], E.K. Thomas[4], X. Benito[5] and J.Z. Yang[6]

We conducted a survey on open-data-sharing experiences among early-career researchers (ECRs). While ECRs feel open-data sharing benefits their career, insufficient training in data stewardship presents a substantial challenge to data reusability.

Paleoclimate researchers readily acknowledge the benefits of open data, while identifying the need to improve best practices for data archival and sharing (Kaufman and PAGES 2k special-issue editorial team 2018). Growing data repositories are especially beneficial for ECRs, enabling the pursuit of synthetic, large-scale research questions from the start of their career. Fully implementing open-data practices throughout a project's lifecycle, however, remains time consuming and challenging.

We sought to understand how these challenges relate specifically to ECRs, and summarize here the results from a recent survey. Our survey was designed around the following questions:

- What challenges do ECRs face in following open-data practices?
- Do ECRs perceive open-data practices as advantageous?
- How can open-data practices enable ECRs' long-term scientific objectives?

While open-data practices are overwhelmingly perceived as advantageous for both one's long-term career and the advancement of science, our results highlight that the largest challenges to ECR implementation include unfamiliarity with community norms, and a lack of training and support. This perspective should inform the community's work towards greater standardization and rigor for open-data-sharing practices.

Methods

The anonymous survey consisted of 30 multiple-choice and free-response questions (see Suppl. Information). We wrote questions to target concerns raised in an ECR forum on open-data experiences (PAGES Early-Career Network 2018), and in consideration of the interactive discussion phase of the PAGES 2k Network open-data-implementation-pilot manuscript in the journal *Climate of the Past* (Kaufman and PAGES 2k special-issue editorial team 2018). Here we define ECRs as non-tenured survey respondents, since achieving tenure is unlikely within five years after PhD completion. We used Qualtrics as our survey platform, and disseminated the survey via paleoscience listservers (e.g. ECN-list; pmip-announce; paleoclimate-list; paleolim-list; Ecolog-list), Twitter, and word of mouth. The survey was open for 17 days, from 31 May to 17 June 2018.

Survey results and implications

Demographics

A total of 183 respondents completed the survey, with 163 identifying as non-tenure. The majority of respondents are students (38%) and postdocs (42%) from Europe (55%) and North America (33%; Fig. 1). Most respondents work with terrestrial (37%) or marine records (27%), or numerical models (23%). A larger proportion of respondents primarily collects or generates data (88%), rather than solely reanalyzing existing datasets (11%), for their research. Respondents commonly characterize their work as driven and dependent on quantitative data (60%). We use the survey results from the 20 tenured respondents as a point of comparison throughout the discussion below.

Data-sharing experience, opinions, and challenges

To facilitate reproducible science, Wilkinson et al. (2016) propose that published scientific data should be Findable, Accessible, Interoperable and Reusable (FAIR). Yet most non-tenured respondents (84%) are unfamiliar with the FAIR guiding principles for data management, a substantially higher proportion than in the tenured group (65%).

Tenured and non-tenured respondents equally feel that data (both 100%), metadata (both 90%) and code (e.g. data-analysis scripts; tenured: 65%; non-tenured: 70%) should be made publicly available and the proportion of respondents who regularly archive open data steadily increases from students (20%) to tenured researchers (80%; Fig. S10, supplementary information). More than two-thirds in all response groups most commonly utilize open databases or journal supplements (tenured: 72%; non-tenured: 65%) followed by personal or institutional databases (tenured: 18%; non-tenured: 12%, Fig. S11).

All respondents reported that a lack of metadata, inconsistent formatting, and data that are not centralized, not digitally available, or paywalled remain top challenges (Fig. S8). Yet, our results highlight that this problem may start at the ECR career stage: over half of the non-tenured

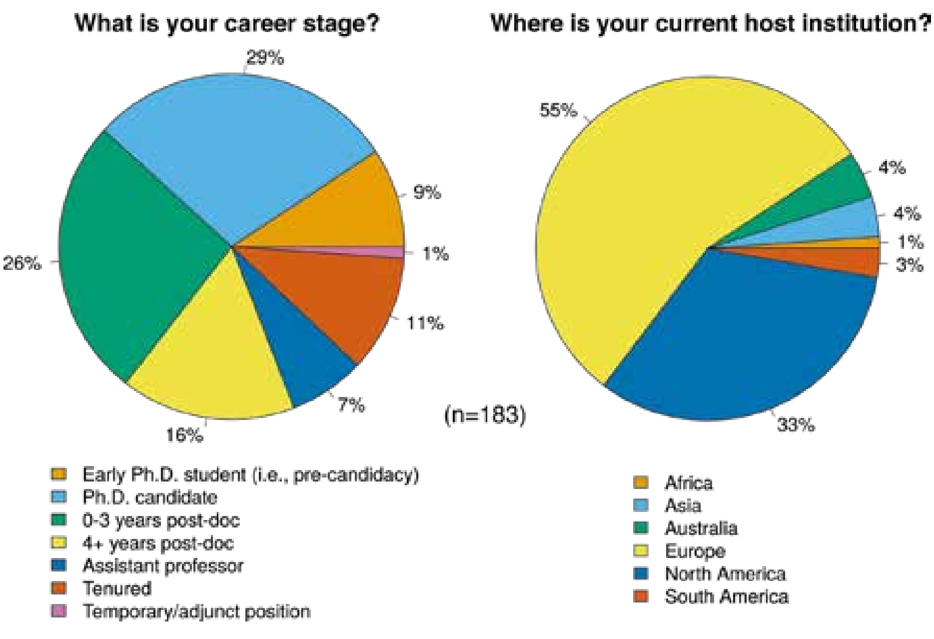


Figure 1: Selected survey demographics.

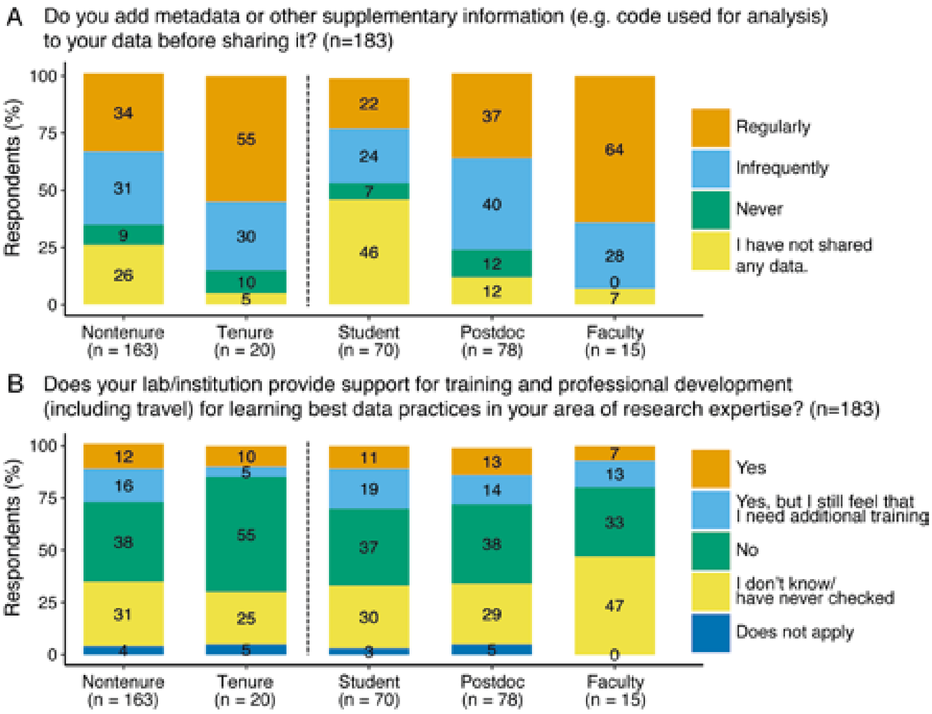


Figure 2: Selected survey responses (%; n=183) grouped by research experience. Non-tenure encompasses student, postdoc and faculty. Results for all survey questions available in the supplement.

respondents indicated "never" (12%) or "infrequently" (45%) adding metadata and code of their own to datasets, compared to 42% tenured respondents (Fig. S12). Our question on data-archival experience (Fig. S12) also reflected this split between ECR stages. If we eliminate respondents who answered "none of the above" because they had not yet published data, students were the largest group to report that the data-archiving process was difficult and the data archive they used lacked metadata templates, tutorials and upload scripts (63%). By comparison, tenured and later-stage ECRs noting this lack of guidance were less (22% each). Thus, unfamiliarity with metadata conventions and data-sharing standards may perpetuate the very problems that respondents identified in existing open datasets.

Data-sharing resources and training

The most common resources allocated to data sharing are time (tenured: 36%, non-tenured 19%) and staff help (tenured: 12%; non-tenured: 16%; Fig. S18). Over a third of the respondents that work in a lab (tenured: 36%; non-tenured: 48%) report that their lab is working towards standard operating procedures (SOPs), suggesting that labs do recognize a need for SOPs for data formatting and sharing. This is particularly important as our survey results signal that the most widespread issue may be related to labs without such SOPs (non-tenured: 89%, tenured: 78%; Fig. S17). More tenured (80%) than non-tenured (69%) respondents work in labs or institutions that offer no support for learning best practices for data sharing, or are not aware whether such support is available (Fig. 2b). Additionally, of the respondents who received training (26%), more than half feel that they need additional training.

Summary and recommendations

It is clear that the community recognizes the positive outcomes of an open-data culture: 95% of all non-tenured respondents and 90% of all tenured respondents feel that data sharing is advantageous to their career. However, equally pervasive are the difficulties surrounding open-access data preparation and publication as well as obtaining metadata-supported data (open-access or otherwise). Specifically, the lack of SOPs and institutional support paired with the unfamiliarity of best practices such as the FAIR guiding principles pose a challenge to data reusability. These benefits and challenges were widespread at all career stages.

Our survey targeting ECR practices and concerns highlighted that open-data usage tends to expand with career progression. We attribute that to researchers becoming more habituated to data-sharing procedures as they advance in their PhD programs, and career. Yet, we also found challenges unique to the ECR career stage:

- steep learning curve for new practitioners;
- widespread unfamiliarity with alternative data-sharing options such as data embargoes.

What can our community do to address these challenges for ECRs, and better promote open-data norms? ECRs working for senior (tenured) researchers may be in the position where their mentor is unfamiliar with the latest data-stewardship best practices, and thus either simply follow their mentor's practices, or must independently find other resources to support good data-sharing practices in their own work. Our survey results, however, suggest that data-management training initiatives (e.g. those offered by the Belmont Forum and Data Tree) are not widely used nor known. We therefore recommend dedicated community-led efforts to raise awareness and promote available training in data stewardship. Additionally, a continued discussion within the community regarding ways to motivate senior researchers and institutions to embrace community-wide data-sharing practices and SOPs will be key for establishing a culture of training ECRs in good data stewardship.

We therefore offer the following recommendations:

(1) Highlight existing resources, including FAIR, embargoes, and training available to ECRs (and other researchers).

(2) Encourage community efforts to the use of best practices in data stewardship and SOPs among ECRs, senior researchers and institutes.

We believe that the PAGES Early-Career Network (pastglobalchanges.org/ecn)can play an integral role in this movement by providing a platform for discourse within the community and a resource for data-stewardship training initiatives.

ACKNOWLEDGEMENTS

Our questionnaire was generated using Qualtrics software, Version May, 2018. Qualtrics and all other Qualtrics product or service names are trademarks of Qualtrics, Provo, USA.

SUPPLEMENTARY INFORMATION

Access the whole survey summary here: doi.org/10.22498/pages.26.2.54

AFFILIATIONS

[1]Department of Geography, University College London, UK
[2]Climate Change Institute, University of Maine, Orono, USA
[3]Department of Environmental Sciences, Rutgers University, New Brunswick, USA
[4]Department of Geology, University at Buffalo, USA
[5]National Socio-Environmental Synthesis Center (SESYNC), University of Maryland, Annapolis, USA
[6]Department of Communication, University at Buffalo, USA

CONTACT

Alexander Koch: alexander.koch.14@ucl.ac.uk

REFERENCES

Belmont Forum (2018) Retrieved August 6, from bfe-inf.org/action-theme-4-capacity-building-human-dimensions

Data Tree (2016) Retrieved August 6 from datatree.org.uk

Kaufman D, PAGES 2k Special Issue Editorial Team (2018) Clim Past 14: 593-600

PAGES Early-Career Network (2018) Retrieved July 6, from groups.google.com/forum/#!topic/pages-early-career-network/rOp6Hc7J6fc

Wilkinson MD et al. (2016) Sci Data 3: 160018

doi.org/10.22498/pages.26.2.56

A funder's approach to more open data and better data management

Belmont Forum e-Infrastructures & Data Management Project

The Belmont Forum partnership of funding organizations, and international and regional science councils, is committed to accelerating open-data sharing and reuse by improving researchers' data-management practices, solving e-infrastructure challenges and improving the data skills of global environmental-change scientists.

The Belmont Forum[1] is a partnership of national science funding organizations, international science councils, and regional consortia across the world committed to the advancement of global environmental science (Fig. 1). The partnership aims to accelerate delivery of data-driven environmental research to remove critical barriers to sustainability by aligning and mobilizing international resources.

The Belmont Forum activities are driven by the Belmont Challenge[2] that encourages international transdisciplinary research to provide knowledge for understanding, mitigating and adapting to global environmental change. The Belmont Forum supports multinational and transdisciplinary collaborative research through Collaborative Research Actions (CRAs)[3], bringing together natural sciences, social sciences and the humanities, as well as stakeholders, to co-create knowledge and solutions for sustainable development.

Global environmental-change research increasingly requires integrating large amounts of diverse data across scientific disciplines to deliver the policy-relevant and decision-focused knowledge that societies require to respond and adapt to global environmental change and extreme hazards, to manage natural resources responsibly, to grow our economies, and to limit or even escape the effects of poverty. To carry out this research, data need to be discoverable, accessible, usable, curated, and preserved for the long term. This needs to be done within a supporting data-intensive e-infrastructure framework that enables data exploitation, and that evolves in response to research needs and technological innovation. Without open data and the supporting e-infrastructure, policy makers and scientists will be forced to feel their way into the future without the benefit of new scientific understanding; unfocused and ill-prepared.

To accelerate the openness, accessibility and reuse of data from CRA projects, the Belmont Forum adopted an Open Data Policy and Principles[4] to stimulate new approaches to the collection, reuse, analysis, validation, and management of data, digital outputs and information, thus increasing the transparency of the research process and robustness of the results. In 2015, the Forum established the e-Infrastructures & Data Management (e-I&DM) Project[5] to help implement the Open Data Policy and reduce barriers to data sharing and interoperability. e-I&DM is promulgating procedures, standards, workflows, and other elements critical to identifying a path toward cooperative e-infrastructures and data-management policies and practices that enable and accelerate open access to, and reuse of, transdisciplinary research data.

Implementing data management for openness and reuse

The Belmont Forum is gradually implementing its Open Data Policy through its CRA funding process. All CRA calls now require a data management plan (Data and Digital Outputs Management Annex[6]) to ensure that project teams will meet both the Open Data Policy and Principles and the Force11 FAIR (Findable, Accessible, Interoperable and Reproducible) Data Principles[7], and adhere to relevant standards and community best practices. Belmont Forum researchers must consider data-management issues from the inception of a project in order to plan and budget appropriately for data curation, management and sharing. Data-management plans should also comply with public-access policies and applicable national laws of the respective funding agencies supporting CRA awards.

Research data and digital outputs are expected to be open by default and publicly accessible, possibly after a short period of exclusivity, unless there are legitimate reasons to constrain access. Data and digital outputs must be discoverable through machine-readable catalogues, information systems and search engines. A full Data and Digital Outputs Management Plan for an awarded Belmont Forum project is expected to be a living, actively updated document that describes the data-management lifecycle for the data and other digital outputs collected, processed, or reused.

A related e-I&DM initiative is a collaboration between Belmont Forum funding agencies and science publishers to articulate a coherent set of data and digital-outputs-management expectations for published research, with the ultimate result of improved sharing and data reuse. Now approved by the Belmont Forum Plenary, the Data Accessibility Statement language will be incorporated into the Data and Digital

Figure 1: Belmont Forum: An International Partnership of Funding Agencies and Science Councils.

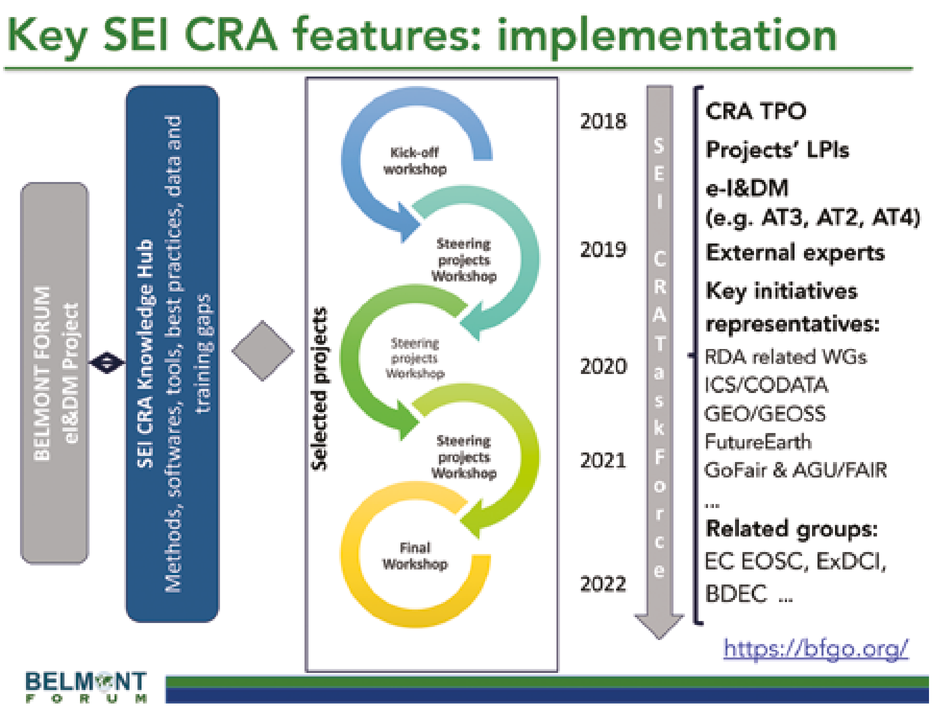


Figure 2: A "Task Force" will steer the SEI CRA implementation in synergy with the Belmont Forum e-Infrastructures & Data Management Project, continuously monitoring the progress and activities of the funded projects through regular workshops, fostering collaboration and maximizing the outcomes across projects. This will contribute to a knowledge Hub catalyzing research efficiently through sharing of best practices, methods and software implementations, to allow delivery of recommendations and priorities to the Belmont Forum for transnational federated data e-infrastructures, data policies and capacity building.

Outputs Management Annex, so researchers will understand the end-to-end expectations of both funders and publishers regarding management of their research data to maximize openness, accessibility, and reuse.

Addressing the barriers to transnational data sharing and reuse

The capability is emerging to bring computer science and technology, as well as large and complex data sets, to bear on interdisciplinary and transdisciplinary science. It is therefore critically important to establish and enable transnational frameworks so that data-driven scientific knowledge can transcend both disciplinary and geographical borders, ultimately increasing the scientific underpinnings of policy and action. International collaboration within the Belmont Forum research priorities holds the potential to establish international foundations for federated data integration and analysis systems with shared services. It can also bring together best practices from the public and private sectors, foster open-data and open-science stewardship among the science communities, including related areas such as publishing, and encourage data and cloud providers and others to adopt common standards and practices for the benefit of all.

For these reasons, the Belmont Forum recently closed a four-year competitive call on *Science-driven e-Infrastructure Innovation (SEI) for the Enhancement of Transnational, Interdisciplinary and Transdisciplinary Data Use in Environmental Change*[8]. The SEI call will fund initiatives that bring together environmental, social, and economic scientists with data scientists, computational scientists, and e-infrastructure and cyberinfrastructure developers and providers to solve methodological, technological and/or procedural challenges currently facing interdisciplinary and transdisciplinary environmental-change research.

The SEI call is being implemented under a "task force" structure (Fig. 2) that requires all funded projects to share results, participate in annual steering workshops, and contribute to a knowledge hub that will catalyze efficient research through sharing of best practices, methods and software implementations. Information in the knowledge hub may also be used to deliver research-driven recommendations to the Belmont Forum to address needs or enhance current strategies for transnational federated data e-infrastructures, data policies and capacity building.

Building researchers' data skills

The Belmont Forum e-I&DM strategy document, 'A Place to Stand'[9], recommended that a "cross-disciplinary training curriculum was required to expand human capacity in technology and data-intensive analysis methods for global change research" and that a new data literacy was required for the 21st century. Consequently, the e-I&DM Project developed the Data Skills Curricula Framework[10], based on a global survey[11] (Skills Gap Analysis), data skills workshop[12] and extensive consultation with data-management experts and trainers.

The Curricula Framework outlines core modules to enhance the skills of domain scientists specifically to make data handling more efficient, research more reproducible and data more shareable – including visualizations for end-users. The five core skills comprise programing, particulars of environmental data, visualization, data management, and interdisciplinary data exchange. Further, a number of optional modules are suggested for more-established researchers as useful introductions to widen their data skills, such as machine learning and object-oriented programing. Two additional modules aim to provide Principal Investigators with an overview of data management and skills needed for open data.

Of the core curricula, the two skill areas addressed least by existing training are 'Environmental data: expectations and limitations' and 'Interdisciplinary data exchange'. Since materials on the former are likely to exist in university courses, 'Interdisciplinary data exchange' is the current focus of the Belmont Forum, to be taught in a mixed class of environmental scientists, social scientists and engineers.

To build on existing capabilities, e-I&DM is investigating the training activities currently available from Belmont Forum member agencies. In addition, e-I&DM is working closely with the data-science community to identify existing training offerings available from around the world and augment content and provision of courses as needed.

Taken as a whole, the Belmont Forum's focus on data management, e-infrastructures and data skills is a critical step forward in advancing open-data sharing, data accessibility and data reuse.

AFFILIATION

Belmont Forum e-Infrastructures and Data Management Project, Montevideo, Uruguay

CONTACT

c3o@bfe-inf.org

LINKS

[1]belmontforum.org
[2]belmontforum.org/wp-content/uploads/2017/04/belmont-challenge-white-paper.pdf
[3]bfgo.org
[4]belmontforum.org/about/open-data-policy-and-principles/
[5]bfe-inf.org
[6]bfe-inf.org/sites/default/files/doc-repository/CRA_Data_Digital_Outputs_Management_Annex_20180501.pdf
[7]force11.org/group/fairgroup/fairprinciples
[8]bfgo.org/opportunity/index.jsp#sei2018
[9]bfe-inf.org/sites/default/files/doc-repository/Community_Edition_A_Place_to_Stand-Belmont_Forum_E-Infrastructures_Data_Management_CSIP_0.pdf
[10]bfe-inf.org/sites/default/files/doc-repository/Outline_Data_Skills_Curricula_Framework.pdf
[11]bfe-inf.org/sites/default/files/doc-repository/BF-Skills-Gap_Analysis-2017_0.pdf
[12]drive.google.com/file/d/0B-s5Of-9WNvBQd2d6ZnpxNXZlZDA/view

doi.org/10.22498/pages.26.2.58

New advances at NOAA's World Data Service for Paleoclimatology – Promoting the FAIR principles

Wendy Gross[1], C. Morrill[1,2] and E. Wahl[1]

Guided by FAIR principles as best data management practices (Wilkinson et al. 2016), the World Data Service for Paleoclimatology (WDS-Paleo) at NOAA's National Centers for Environmental Information (NCEI) has recently deployed new capabilities and data-format standards. These and planned future developments facilitate the standardization and aggregation of WDS-Paleo's small, long-tail, and heterogeneous datasets into larger standardized collections. These capacities enhance the value of the data, analogous to how large volumes of well-managed big data can be transformed into valuable information (Lehnert and Hsu 2015).

WDS-Paleo archives and provides paleoclimatology data products derived from a variety of sources, such as tree rings, ice cores, corals, and ocean and lake sediments, along with web-based services to access these products. To attain the goal of long-term professional preservation and dissemination of its data, WDS-Paleo partners with its user communities and maintains long-standing relationships with PAGES, PANGAEA and Neotoma. WDS-Paleo works with these partners to offer aggregated search capabilities. NCEI data stewardship operations meet the responsibilities of an Open Archival Information System (oais.info), and new and existing capabilities follow FAIR best practices as follows.

Findable

WDS-Paleo makes its data findable via geographic map-based searches, along with a web service featuring an application programing interface (API) for programmatic use and a graphical user interface (GUI) that also acts as an API-builder tool (ncdc.noaa.gov/paleo-search). Recently a new controlled vocabulary, *Paleoenvironmental Standard Terms* (PaST; ncdc.noaa.gov/data-access/paleoclimatology-data/past-thesaurus) has been developed for documenting variables (i.e. paleoclimate measurements, units and methods). With the input of 25 subject-matter experts, terms have been assigned to over 100,000 paleoclimatic time series, powering new search capabilities that complement other web-service features. Paleoclimate data are extremely heterogeneous, and with PaST terminology WDS-Paleo's search capabilities now capture this heterogeneity. In the future, interactive visualizations of PaST will allow users to obtain more-detailed information about terms and will enhance data discovery. The governance structure for PaST is described at the above link.

Accessible

A recently released feature of the WDS-Paleo web service provides users with capacity to bundle and download search results, thus easing the process of procuring sets of data appropriate for specific use.

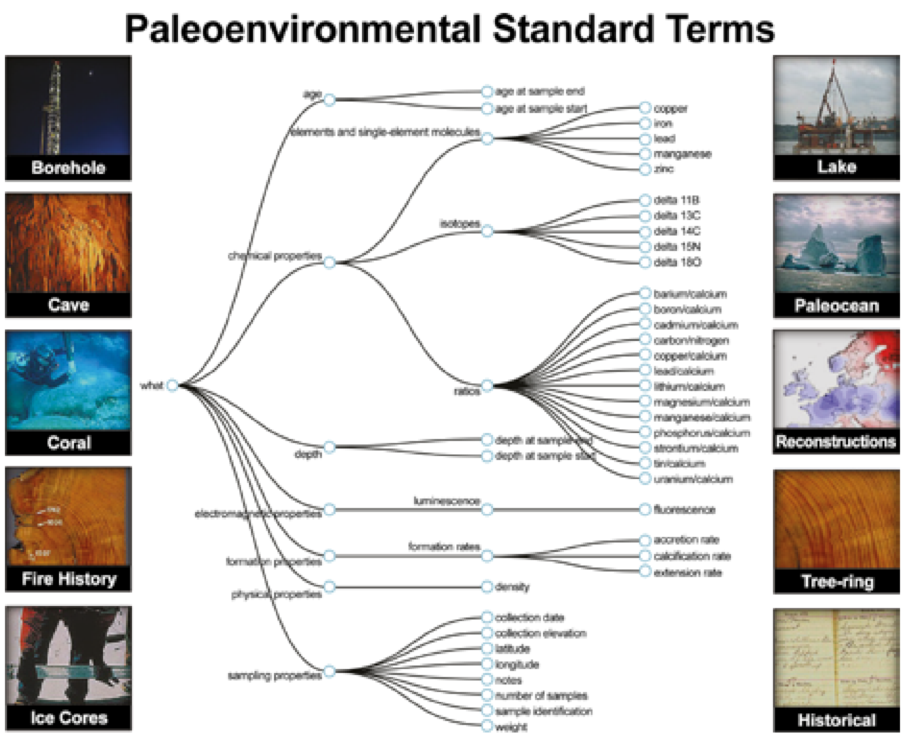


Figure 1: Example of Paleoenvironmental Standardized Terms (PaST) controlled vocabulary.

A bundle includes data files and manifest information, maintaining provenance of both data and metadata.

Interoperable

Upcoming and long-standing data and metadata formats promote interoperability via machine readability and common tools. Going ahead, a NOAA Standard for the Linked Paleo Data Format (LiPD, lipd.net) (now in development), is designed to facilitate interoperability between LiPD and the NOAA WDS-Paleo Template data format, including use of PaST terms. Standardized metadata formats, including DIF, ISO, and JSON, facilitate data discovery and federated search capabilities. Community-specific data formats and software tools, including those developed and used by the International Tree Ring Databank (ITRDB) and International Multiproxy Paleofire Database (IMPD), provide key resources for scientific discovery.

Reusable

PaST-enhanced web-service search capabilities aggregate WDS-Paleo's small, long-tailed datasets into larger, standardized collections. This can facilitate large-scale data syntheses, which is a key thrust in paleoclimatology (e.g. PAGES 2k Consortium 2017), and also promotes reuse of paleo data beyond paleoclimate specialists. WDS-Paleo is implementing persistent identifiers and locators for datasets via the provision of DOIs.

Going forward, PaST and a future project of standardizing the reporting of age determination will greatly enhance the interoperability of WDS-Paleo data formats, allowing for easier and more robust aggregation of datasets.

The WDS-Paleo website, including search, access, data contribution, and PaST information is at: ncdc.noaa.gov/paleo

AFFILIATIONS

[1]NOAA's National Centers for Environmental Information, Boulder, USA

[2]Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado Boulder, USA

CONTACT

WDS-Paleo: paleo@noaa.gov

REFERENCES

Lehnert K, Hsu L (2015) 11: 368-369

PAGES 2k Consortium (2017) Sci Data 4: 170088

Wilkinson MD et al. (2016) Sci Data 3: 160018

doi.org/10.22498/pages.26.2.59

PANGAEA - Data publisher for Earth & environmental sciences

Michael Diepenbroek

Initiated in the 1990s, PANGAEA[1] has evolved from a paleoclimate-data archive to a multidisciplinary data publisher for Earth and environmental sciences, accredited as a *World Data Center* by the International Council for Science World Data System (ICSU WDS)[2] and as *World Radiation Monitoring Center* (WRMC)[3] within the World Meteorological Organization Information System (WIS)[4].

Even in its earliest stages, data were archived consistently and carefully curated. This involves cleaning, harmonizing, and integrating data, as well as metadata, within PANGAEA's editorial workflow. Consequently, all data sets are annotated including information on how, when, and where they were produced, information about principal investigators, measurement and observation types, sampling and analysis methods, and devices as well as references to literature. In January 2005, the first data sets were registered and minted with a standard-compliant Digital Object Identifier (DOI), which enables proper citation of data and their integration within the publishing-industry workflow and bibliometric analyses. Today, PANGAEA holds around 375'000 citable data sets comprising more than 13 billion data items - numerical and textual data as well as binaries such as images, videos, or files with community specific mime types. Each data item is a georeferenced record including the parameter value, parameter type, and the spatial and temporal coverage; spatio-temporal values themselves are not data items. Over 18% of published data sets include at least one author linked to ORCID (the author identifier of the publishing industry). PANGAEA is operated as an Open Access library and is open to any project, institution, or individual scientist to use or to archive and publish data[5].

As paleoclimate research is the scientific background of PANGAEA's founders, it has a long-lasting relationship with PAGES and also looks back to a long-standing collaboration with the NOAA WDS-Paleo. The recent common focus is on interoperability and findability of paleodata. Both data centers build the archive backbone for paleodata. PANGAEA holds large inventories of all types of paleodata, for example isotope and geochemical data as well as pollen and tree-ring data. An example data collection is the data collected by the PAGES C-PEAT working group[6].

Figure 1: PANGAEA's website offers various ways to search for data.

Editorial

PANGAEA is operated by a team of data editors, project managers, and IT specialists[7]. Our editors are scientists with expertise in all fields of Earth and environmental science. They have a deep knowledge of the review and processing of scientific data. The PANGAEA data editorial ensures the integrity and authenticity as well as a high reusability of data. Archived data are machine readable and mirrored into our data warehouse which allows efficient compilations and downloads of data[8].

Data are submitted using a ticket system (Jira[9]) and assigned to an editor who is a specialist in the corresponding data domain. Preparation of the data for import is done with a highly sophisticated editorial system. Data editors check the completeness and validity of data and metadata, reformat data according to the PANGAEA ingest format, and harmonize data and metadata using standard terminologies (Diepenbroek et al. 2017). The editorial review is complemented by inviting authors and external reviewers (e.g. reviewers of articles supplemented by the data) to proofread the data sets. After being accepted, the data sets are archived, provided with a DOI, and registered at DataCite[10].

Interoperability and findability

PANGAEA is furnished with a well-developed interoperability framework based on internationally accepted standards. All interfaces to the information system are based on web services including map support (Google Earth, Google Maps)[11]. This allows most effective dissemination of metadata and data to all major internet search-engine registries, library catalogues, data portals, and other service providers, and consequently ensures the optimal findability of data hosted by PANGAEA. Scientific data portals supported include DataOne, GEOSS[12], the ICSU WDS[2], GBIF[13] and also the paleo data portal at NCEI[14]. Other infrastructures supported include DataCite[15], ORCID[16], and Scholix[17], which supplies links between scholarly literature and data. Interoperability with ORCID allows users to login with their ORCID ID and link it to their user profile in PANGAEA. In this way, data publications are automatically assigned to matching ORCID IDs.

AFFILIATION

MARUM - Center for Marine Environmental Sciences, University Bremen, Germany

CONTACT

Michael Diepenbroek: mdiepenbroek@pangaea.de

REFERENCES

Diepenbroek M et al. (2017) J Biotech 261: 177-186

LINKS

[1]pangaea.de
[2]icsu-wds.org
[3]bsrn.awi.de
[4]wmo.int/pages/prog/www/WIS
[5]pangaea.de/submit
[6]pangaea.de/?q=project%3Alabel%3APAGES_C-PEAT
[7]pangaea.de/about/team.php
[8]pangaea.de/tools
[9]en.wikipedia.org/wiki/Jira_(software)
[10]datacite.org
[11]ws.pangaea.de
[12]geoportal.org
[13]gbif.org
[14]ncdc.noaa.gov/paleo-search
[15]datacite.org
[16]orcid.org
[17]scholix.org

doi.org/10.22498/pages.26.2.60

Lessons learned from 25 years of PMIP model-data distribution

Jean-Yves Peterschmitt, P. Braconnot and M. Kageyama

Open paleo data from both observations and models underlies the success of the Paleoclimate Modelling Intercomparison Project. We present how the project has evolved from a stand-alone database to an active member of a distributed international infrastructure following community standards.

Climate models are improved iteratively, as scientific knowledge, along with computing and storage technology progress. Sharing and comparing models and their output to paleo reconstructions is an essential part of this process. This can be done by sharing data directly between individuals, but is more efficient when formally organized as a MIP (Model Intercomparison Project), where all contributors and users adopt the same standards. The Paleoclimate Modelling Intercomparison Project (PMIP), started in 1990 (Joussaume and Taylor 1995), was one of the early MIPs, following the AMIP example (Gates et al. 1998).

PMIP has been successful in terms of participation, publications, and contributions to successive IPCC Working Group 1 reports, and is now in its fourth phase, with 20 modeling groups/models from 14 countries (Kageyama et al. 2018; Kageyama et al. 2016 [PMIP4 special issue]). The first studied periods were the mid-Holocene and the Last Glacial Maximum, with the pre-industrial period used as a control run. PMIP4 now includes five additional experiments: the last millennium, the Last Interglacial, the mid-Pliocene Warm Period, the last deglaciation and DeepMIP. Thanks to improvements in model complexity, resolution, and length of the simulations, the different phases of PMIP have targeted key scientific questions on climate sensitivity, the hydrological cycle, and abrupt event and inter-annual to multi-decadal variability.

For PMIP4, experimental protocols were co-designed by the modeling and data communities (Kageyama et al. 2018). They require that the same model version be used for PMIP4-CMIP6 experiments and future climate projections so that rigorous analyses of climate processes, including both physical and biogeochemical interactions, can be performed across the range of past and future climate. This is done in collaboration with other CMIP6 MIPs (Eyring et al. 2016).

PMIP simulations address the key CMIP6 overarching questions:

- How does the Earth system respond to forcing?
- What are the origins and consequences of systematic model biases?
- How can we assess future climate changes given climate variability, predictability and uncertainties in scenarios?

	PMIP 1	PMIP 2	PMIP 3	PMIP 4
DB online	1996	2005	2011	2018
Number of groups/ models	22	18	25	20
Number of countries	11	10	12	14
Main experiments	0 k 6 k 21 k	Same as PMIP 1	PMIP 2 + Last Millennium	PMIP3 + Last Interglacial + Mid Pliocene Warm Period + Last Deglaciation + DeepMIP
DB Size	1.7 GB	482 GB	**distributed** *several* TB	**distributed** LOTS of TB...
Data distribution	ftp server LSCE (+PCMDI)	DODS server LSCE	CMIP5 ESGF	CMIP6 ESGF
Data format & Convention	NetCDF AMIP/CF	NetCDF CMIP+PMIP2/CF	NetDCF CMIP5/CF	NetDCF CMIP6/CF
Example grid IPSL atmosphere	lmcelmd5 64x50 x L11	IPSL-CM4-V1-MR 96x72 x L19	IPSL-CM5A-LR 96x95 x L39	IPSL-CM6A-LR 144x143 x L79
Example grid NCAR atmosphere	ccsm3 128x64 x L18	CCSM 128x64 x L17	CCSM4 288x192 x L26	CESM2 288x192 x L32

Box 1: PMIP database factsheet

Current work places a particular emphasis on the assessment of the different sources of uncertainties resulting from, for example, model formulation, reconstructions of forcing, and internal model noise. Model-data comparisons are key in this process (Braconnot et al. 2012; Harrison et al. 2015).

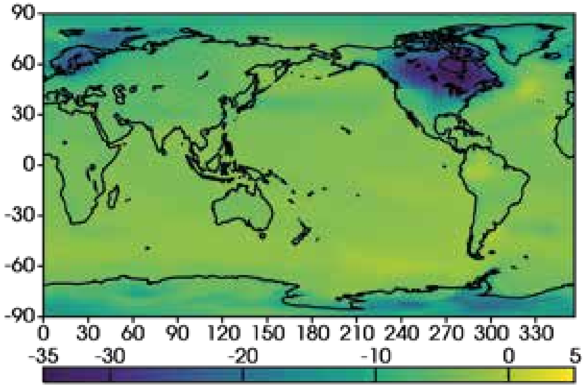
The PMIP model database has progressed from almost 2 GB for PMIP1 (~14,500 files) to a frightening (and unknown!) number of terabytes for PMIP4 (Box 1). Standards and good data-distribution tools are the key to dealing with the massive amount of data generated, along with good communication tools (mailing lists and websites), and invaluable help from the Earth System Grid Federation (ESGF; Balaji et al. 2018) community that maintains the CMIP database.

Using standards

The database of model output is too large to be accessed by ordinary database queries. Nevertheless, users need to easily access the subset of the data they need for their analyses, regardless of which research group generated it. In PMIP, this is achieved through the use of community standards. Standards are sometimes viewed as a hindrance to data production, but they are necessary to avoid chaos when working with multi-model data – the essence of a MIP. Data that is consistent across all the models and experiments eases reuse by users, and is required to automatically process numerous files, easily ingest new files, and to reprocess files when a bug is found. Such standardization also generally makes any analyses more reproducible.

Standardization is a key aspect of the long history of PMIP in international collaborations. PMIP currently follows the CMIP6 standards for file format (NetCDF format) and metadata (Climate and Forecast conventions, CMIP6 Data Reference Syntax, Controlled Vocabulary and Data Request). The NetCDF binary format has many advantages: self-describing, easily and efficiently writable/readable by programs, capacity to hold several gigabytes of data, and suitable for long-term archiving. Thanks to these choices it is still possible to access the content of PMIP1 files created more than 20 years ago. It is not easy for the modeling groups to meet the CMIP6 requirements, but the Climate Model Output Rewriter (CMOR3) library and project-specific configuration


```
In [1]: import cdms2, cdutil, vcs
In [2]: f = cdms2.open('/home/scratch01/jypeter/tas_IPSL-CM5A-LR_21-0.nc')
In [3]: v = f('tas_diff')
In [4]: v_annual_mean = cdutil.YEAR(v)
In [12]: x = vcs.init(geometry=(600,400), bg=True)
In [16]: x.plot(v_annual_mean)
Out [16]:
```




```
In [17]: x.clear()
         x.plot(v_annual_mean(latitude=(20,80), longitude=(-20,40)), ratio='autot')
Out [17]:
```

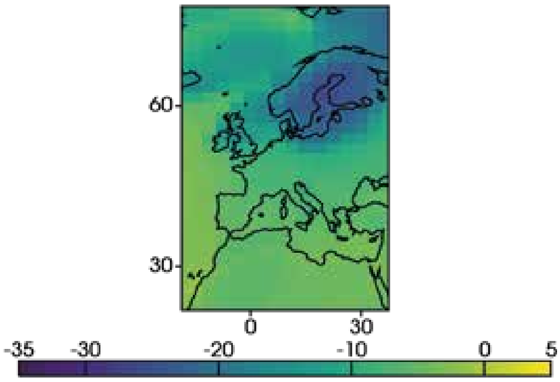


Figure 1: An example of a script using the CDAT climate-oriented Python distribution in a notebook to read PMIP3 data from a NetCDF file (surface temperature, anomaly for the Last Glacial Maximum minus the pre-industrial control, for the IPSL-CM5A-LR model), compute the annual mean and plot it on a global grid and a smaller region.

tables facilitate the creation of CMIP-compliant files.

Accessing model data

Once the data files are available in the standard format, the next goal is to ensure they move as smoothly as possible from the data provider to the data user. This is accomplished through a number of developments:

- Model-output data providers need an automatic service to answer user requests.
- Users want to determine easily if the required data is available, and then to easily access the files. Given the size of the database (Box 1) there are ongoing developments to provide computation and analysis services directly on the servers holding the data.
- Users need a good documentation of the models and how the PMIP experiments were run. For PMIP4-CMIP6, this information will be centralized on the Earth System Documentation (es-doc) site.

For CMIP5-6 (PMIP3-4), the data files are sent by the modeling groups to the closest ESGF Data Node and, after review ranging from a basic validation to an exhaustive quality control, they can be searched and downloaded from any other node of the federation. This distributed repository is scalable and is the only practical way to handle the 10-50 petabytes of data expected for CMIP6 (including PMIP4 data). ESGF also offers a fast web-search interface and bulk data-download tools. This infrastructure is powerful, but it requires substantial manpower for customized software development and local node administration, as well as sufficient storage and computing resources.

In addition to standardization, the PMIP data policy has evolved over time. For PMIP1, the full database was initially available only for the groups which had submitted data during an embargo period, prior to public release. For PMIP2, the database was also available for people proposing an analysis project. PMIP3-4 followed the CMIP5-6 data policy, which allows anyone to use the data from modeling groups, with some restrictions for commercial applications. In turn, the results of the study that uses the model output must be shared with the same open policy, without forgetting to credit the producers.

Using PMIP data

There are many ways to use PMIP model data, depending on the analyses to be done. The data complexity (number of available variables and file size) has increased substantially since the beginning of PMIP, but the programing complexity has decreased. It is now much easier to use a high-level scripting language (Fig. 1) than it was to use Fortran programs. Users can also process PMIP data with the Graphical User Interfaces provided by some programs (e.g. GIS programs such as QGIS), but they may be quickly limited by data size and available operations. There is also an ongoing effort by the PMIP community to provide some higher-level web interface; this will receive more attention in the coming years.

Conclusion

PMIP has benefited from CMIP5-6 and the ESGF infrastructure, which has eased the comparison between past and future climate simulations. One of the next challenges is to make using the data easier for non-modelers, especially experts in paleoclimate reconstructions. This will require the deployment of specific web servers similar to the ones used for impact studies, but customized for paleoclimate needs. Another challenge will be to deal with the long, transient climate simulations (thousands of years of model data) generated by the PMIP4 experiments (deglaciation, the Eemian and the Holocene) when performing model-model and model-data comparisons.

RESOURCES

PMIP: pmip.lsce.ipsl.fr
AMIP and CMIPn: pcmdi.llnl.gov/mips
PMIP3 publications: citeulike.org/user/jypeter/order/year
NetCDF: unidata.ucar.edu/software/netcdf
CF conventions: cfconventions.org
CMIP6 DRS: goo.gl/v1drZl
CMIP6 DR: earthsystemcog.org/projects/wip/CMIP6DataRequest
CMIP6 CV: github.com/WCRP-CMIP/CMIP6_CVs
CMOR3 library: cmor.llnl.gov
es-doc: search.es-doc.org
CDAT: cdat.llnl.gov
QGIS: qgis.org
DeepMIP: deepmip.org

AFFILIATIONS

Laboratoire des Sciences du Climat et de l'Environnement (LSCE/IPSL, CEA-CNRS-UVSQ), Université Paris-Saclay, Gif-sur-Yvette, France

CONTACT

Jean-Yves Peterschmitt: Jean-Yves.Peterschmitt@lsce.ipsl.fr

REFERENCES

Balaji V et al. (2018) Geosci Mod Dev 11: 3659-3680
Braconnot P et al. (2012) Nat Clim Change 2: 417-424
Eyring V et al. (2016) Geosci Mod Dev 9: 1937-1958
Gates WL et al. (1998) BAMS 73: 1962-1970
Harrisson SP et al. (2015) Nat Clim Change 5: 735-743
Joussaume J, Taylor K (1995) Proceedings of the first international AMIP scientific conference, Monterey, USA: 425-430
Kageyama M et al. (2016) CP/GMD inter-journal SI
Kageyama M et al. (2018) Geosci Mod Dev 11: 1033-1057

doi.org/10.22498/pages.26.2.62

LinkedEarth: supporting paleoclimate data standards and crowd curation

Julien Emile-Geay[1], D. Khider[1], N.P. McKay[2], Y. Gil[1,3], D. Garijo[3] and V. Ratnakar[3]

Data science in the paleo sphere has been hindered by a lack of standards that limit interoperability and interdisciplinarity. Here we describe the LinkedEarth project, which lowered some of these barriers, and offers a blueprint for further erasing them.

At present, scientists are asked to upload their data to various "silos" (loosely connected data centers like WDS-Paleo[1], PANGAEA[2], or Neotoma[3]), which use different formats and conventions, hampering interoperability. Further, there is sometimes little guidance on what information needs to be archived to provide long-lasting scientific value. While paleosciences offer a long-term perspective on environmental change, this cannot happen without a long-term perspective on environmental data stewardship. LinkedEarth[4] (Fig. 1) was funded by the EarthCube program[5] as a two-year "integrated activity", with the twin aims of putting paleoclimate data stewardship in the hands of data generators, and developing standards that promote effective reuse. Here we review LinkedEarth's successes and outstanding challenges, and take stock of its broader lessons for the PAGES community.

LinkedEarth has acted as a laboratory to advance the notion of decentralized paleodata curation, allowing data generators to curate their own and others' data, via standards and technologies. The basic premise of LinkedEarth is that no-one understands data better than the people who generated them. Therefore, data generators should be the ones describing their data, but in a consistent way to make them interoperable. Having participated in several PAGES' syntheses (e.g. PAGES 2k Consortium 2017), we also appreciate that publicly-archived datasets are nearly always incomplete, and may harbor errors - requiring collective curation and correction (that is, the ability for multiple actors to edit and annotate the same datasets). We thus set out to develop a platform that would enable paleoclimatologists to interact with data in an intuitive way, resulting in standardized datasets that are (by construction) extensible, interoperable, and discoverable.

Crowd-curation through standards

A data standard consists of three parts: (1) a standard terminology, to prevent ambiguity; (2) standard practices, which codify the information that is essential to long-term reuse and (3) a standard format for archival and exchange. The latter is emerging, in the form of Linked Paleo Data (LiPD[6]; McKay and Emile-Geay, this issue, 2016), so LinkedEarth only had to contend with the first two parts.

Standardizing terminology was accomplished by means of the LinkedEarth ontology[7]. An ontology is a formal representation of the knowledge common to a scholarly field. It allows unambiguous definitions of common terms describing a paleoclimate dataset, as well as the relationships among these terms (e.g. a proxy observation is measured on a proxy archive at a particular depth). Ontologies are necessary to organize information so machines can take advantage of digitally-archived data. Ontologies are inherently flexible, allowing to specify ecological properties such as habitat depth and seasonality to previously-archived foraminiferal-based records. Ontologies have had an enormous impact in biomedical research, ranging from genomics to drug discovery, and are beginning to permeate the geosciences[8].

Ontologies need to be sufficiently rigid so that dependent applications can rely on their structure being stable over time, yet sufficiently flexible to accommodate growth and evolution. The ontology maps closely to the LiPD structure, which serves as its stable skeleton. Extensibility was achieved via a new technology, the LinkedEarth platform[9]. At its core, it is a semantic wiki, similar to other wikis like Wikipedia, but based on the LinkedEarth ontology. The LinkedEarth wiki tracks changes and attributes them to authenticated contributors (an ORCID is all that is required to join LinkedEarth). The wiki facilitates extensions by allowing users to edit the non-core aspects of the ontology: they can define new classes or properties, create or change definitions, start discussions with other users, or request modifications to the core ontology when sufficient consensus emerges. These user roles and interactions were defined in a formal charter[10]. The flexible structure will accommodate advances in techniques and interpretations, and allow users to deprecate outdated terms.

Because LinkedEarth datasets are based on LiPD, they can be uploaded or downloaded in a few clicks, and benefit from

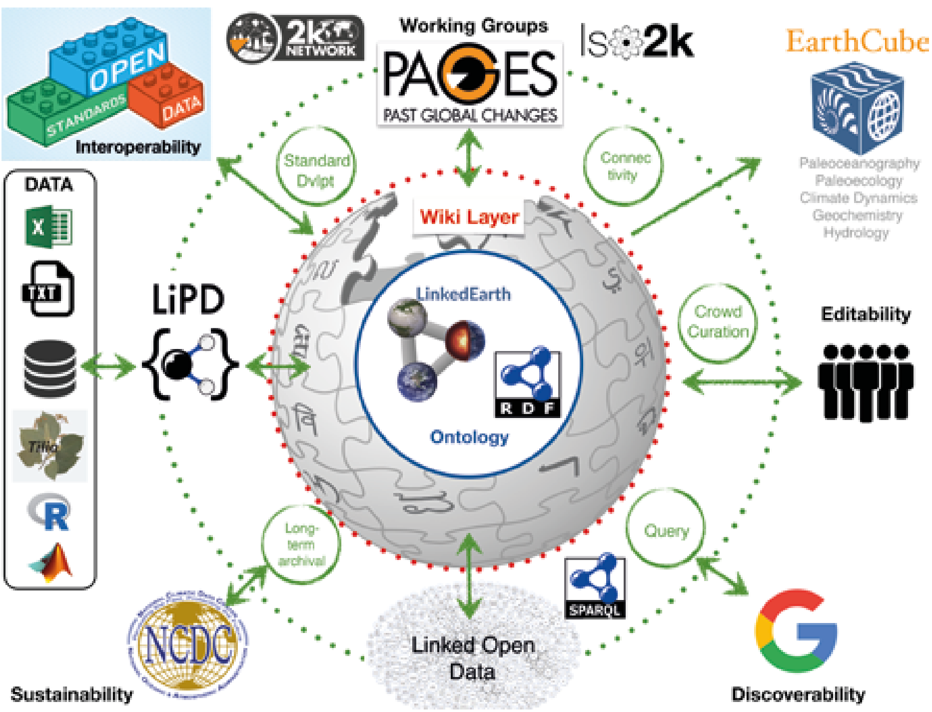


Figure 1: The LinkedEarth design (Gil et al. 2017) is structured around a semantic core (ontology), which can be easily interacted with thanks to a wiki. In addition, the framework supports import/export in LiPD, rich queries, the elaboration of community standards, the crowd-curation of datasets and a natural link to the Web of Data, ensuring discoverability.

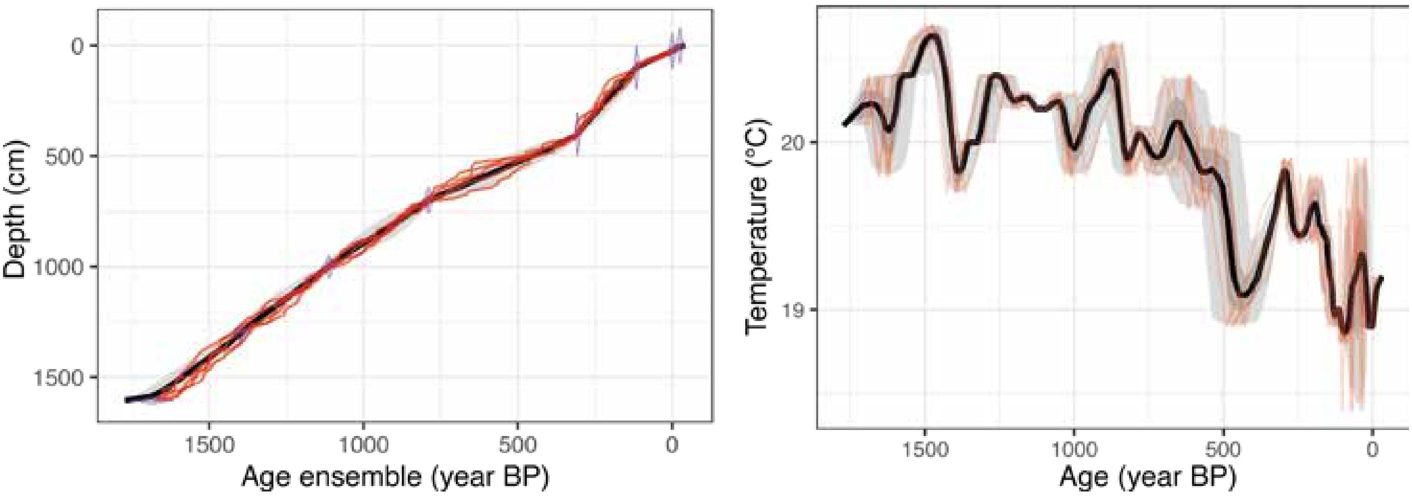


Figure 2: (A) Chronology for Basin Pond sediment core using the BACON software; **(B)** Basin Pond temperature reconstruction placed on an ensemble of age models (red line). The bold black line represents the median. The grey areas represent the 1σ (dark) and 2σ (light) uncertainty bands on calendar age. For details, see nickmckay.github.io/LinkedEarth-Neotoma-P418/LE-Neo_UseCase.html

the entire LiPD research ecosystem (McKay and Emile-Geay, this issue). This makes LinkedEarth-hosted data inherently ***interoperable***. In order to ensure the lasting utility of the data, LinkedEarth sparked the first international discussion on community-led data reporting standards[11], to build consensus on the most important information that should be reported in paleoclimate datasets. This consensus-building enterprise was facilitated by the LinkedEarth platform, including working groups, discussions, and polling (Gil et al. 2017).

Lastly, the semantic part of LinkedEarth means that datasets are broadcast to the web using standard schemas[12], which make them ***discoverable*** by various search engines, including Google. Because of this outward-facing design, LinkedEarth datasets were the first to be integrated into EarthCube's Project 418[13] (P418), an EarthCube initiative to demonstrate common publishing approaches for data holdings using such standard schemas.

Towards Interoperable Paleo Data

Discovering data pertinent to a scientific question is critical, but what to do once you find them? Imagine a user interested in the impact of time uncertainty on a pollen-based temperature reconstruction at Basin Pond, Massachusetts, USA. After a quick search through the P418 interface, our user realizes that the temperature reconstruction is hosted on LinkedEarth while the geochronological information is stored on Neotoma. Using the P418 service, they can find and download the datasets of interest. The GeoChronR software package (McKay et al. 2018) can then facilitate their analysis. GeoChronR was built around LiPD, which has been mapped to the Neotoma data model (that is, Neotoma datasets can be read by any LiPD-based code). This enables fast integration between the LinkedEarth-hosted temperature reconstruction and the Neotoma-hosted chronological data. Within the GeoChronR framework, our user has access to a variety of age-modeling tools, including Bacon (Blaauw and Christen 2011). They can then readily visualize the new age model (Fig. 2a) and assess the impact of age uncertainty on the temperature evolution (Fig 2b). Such is the promise of holistic data stewardship: more than putting data online, it's about drastically simplifying their reuse.

Beyond LinkedEarth

In a short two years, LinkedEarth has brought to life a functional platform for the crowd-curation of paleoclimate data and an emerging data standard. Along the way, it provided a nucleus for interoperability via synergistic software (GeoChronR, Pyleoclim[14]).

Despite these accomplishments, the vision still faces notable challenges. First, it has proven difficult to elicit broad participation: only 100 paleoclimatologists have answered our survey on paleoclimate data standards so far. We have found that overburdened scientists have little inclination to participate in such activities unless there are clear incentives. We argue that only publishers and funding agencies can provide these incentives, but have yet to do so. We do not envision meaningful progress until they do. Another issue concerns adoption: despite a non-trivial investment of resources (funding, personal time for participants), very few scientists are actively using LinkedEarth. PAGES is playing a leading role in incentivizing a new generation of paleoscientists to curate high-quality data compilations and take advantage of the LiPD-based research ecosystem, which was built for them. PAGES 2k[15] is a case in point, having motivated the birth of LiPD, the need for crowd-curation, and many of the ontologies' categories. One persistent obstacle to adoption is the perceived redundancy with data repositories. LinkedEarth is a framework, and works in tandem with repositories. It has strong links to WDS-Paleo, which now accepts LiPD as a submission format, and can ensure long-term archival. Because of LiPD's structured nature, LinkedEarth also integrates well with Neotoma; links to other repositories are in the works. The success of LinkedEarth will be measured over time by adoption and extension of its various tools and standards. We look forward to many more PAGES compilations being generated, discussed, and published on LinkedEarth. Every new PAGES working group brings with it new requirements; so far, LinkedEarth's intrinsic flexibility has enabled it to accommodate them all, and likely will for the foreseeable future.

AFFILIATIONS

[1]Department of Earth Sciences, University of Southern California, Los Angeles, USA

[2]School of Earth and Sustainability, Northern Arizona University, Flagstaff, USA

[3]Information Sciences Institute, University of Southern California, Marina Del Rey, USA

CONTACT

Julien Emile-Geay: julieneg@usc.edu

REFERENCES

Blaauw M, Christen JA (2011) Bayesian Anal 6: 457-474

Gil Y et al. (2017) The Semantic Web - ISWC 2017: 231-246

McKay NP, Emile-Geay J (2016) Clim Past 12: 1093-1100

McKay NP et al. (2018) GeoChronR (Version 1.0.0). Zenodo

PAGES 2k Consortium (2017) Sci Data 4: 170088

LINKS

[1]ncdc.noaa.gov/data-access/paleoclimatology-data
[2]pangaea.de
[3]neotomadb.org
[4]linked.earth
[5]earthcube.org
[6]lipd.net
[7]climdyn.usc.edu/publication/leo
[8]geoscienceontology.org/about.html
[9]wiki.linked.earth
[10]linked.earth/aboutus/governance/charter
[11]wiki.linked.earth/Paleoclimate_Data_Standards
[12]schema.org
[13]earthcube.org/group/project-418
[14]doi.org/10.5281/zenodo.1205661
[15]wiki.linked.earth/PAGES2k

doi.org/10.22498/pages.26.2.64

Constituent databases and data stewards in the Neotoma Paleoecology Database: History, growth, and new directions

Eric C. Grimm[1], J.L. Blois[2], T. Giesecke[3], R.W. Graham[4], A.J. Smith[5] and J.W. Williams[6]

The Neotoma Paleoecological Database provides critical cyberinfrastructure for paleoenvironmental research. The database can accommodate virtually any type of fossil data or paleoenvironmental proxy, and is extensible to new data types.

Scientists have long harnessed paleodata to study ecosystem dynamics across time and space. For example, to reconstruct the postglacial expansion of tree species, von Post (1924) assembled fossil-pollen data from across Sweden; Szafer (1935) assembled data from Poland and neighboring areas and invented isopolls to summarize the data; while Firbas (1949) collected data from central Europe north of the Alps, which he summarized in various ways including with isopolls, which he called "Pollenniederschlagskarten" (pollen rain maps). These early investigators assembled, organized, and processed data. In other words, they created "databases", although that term was not yet invented. Their work demonstrated the power of data collections to address emergent questions. With the advent of computers, this power was greatly amplified, for both data management and data analysis.

An early effort to harness computing power was the Cooperative Holocene Mapping Project (COHMAP Members 1988; Wright et al. 1993) in the 1970s, which developed an archive of pollen data as flat files. Many scientists contributed data to this project, which produced numerous publications and spinoff projects. Nevertheless, the data were not publicly available, accompanied by rich metadata, or stored in a relational database. That changed with the advent of the North American Pollen Database (NAPD) in the early 1990s, which was made available for public access by the National Geophysical Data Center of the U.S. National Oceanic and Atmospheric Administration. NAPD was first populated with data from COHMAP, then continued to acquire additional legacy and new data over about 15 years. The European Pollen Database (EPD) was developed simultaneously and in collaboration with NAPD, but the two databases remained separate. The FAUNMAP database, which included Quaternary data from the conterminous United States, was also launched in the early 1990s and made available on floppy disk included with its publication (FAUNMAP Working Group 1994). Following the success of these three databases, other databases were developed for other regions and data types, including the Latin America Pollen Database (LAPD), African Pollen Database, North American Plant Macrofossil Database, North American Non-Marine Ostracode Database (NANODe), Diatom Paleolimnology Data Cooperative, Northern Eurasian Palaeoecological Database, and others.

These database projects assembled large numbers of datasets, involved disciplinary experts, and supported and engendered scientific research. Nevertheless, they suffered from funding lapses and inability to cross-communicate. These issues and others led to the creation of the Neotoma Paleoecology Database (neotomadb.org) following a 2007 workshop at Pennsylvania State University (Williams et al. 2018). This database is named after the rodent genus *Neotoma*, prodigious collectors of diverse materials within their territories and which under the right conditions preserve a multiproxy record of environmental change.

Neotoma provides the underlying cyberinfrastructure for a variety of disciplinary database projects and can accommodate virtually any type of fossil data or paleoenvironmental proxy. All data in Neotoma are stored in a single centralized database but are conceptually organized into virtual constituent databases. These constituent databases, which may be organized according to data type or region, involve disciplinary specialists for data types and regions, thus providing domain scientists with quality control over their portions of the data. Neotoma is a curated resource with governance and control by disciplinary experts. "Curation" implies a high level of quality control. All data added to Neotoma are reviewed and uploaded by

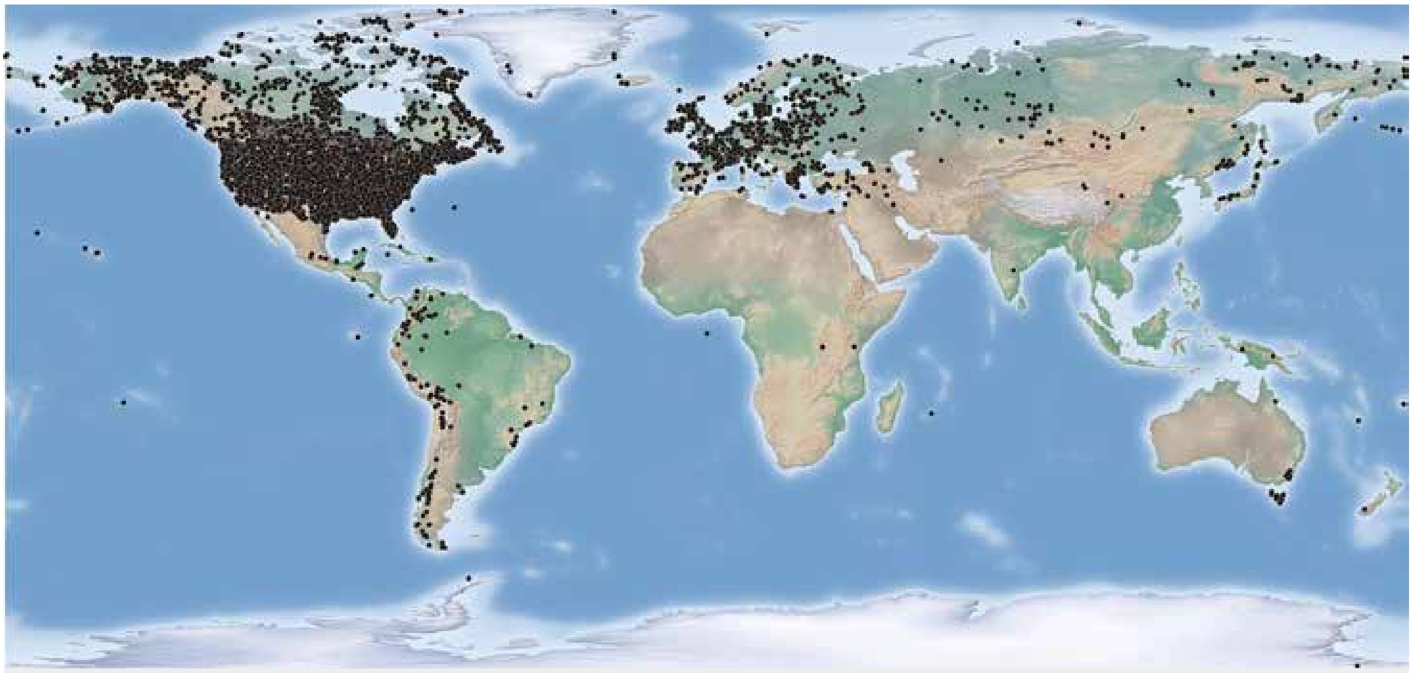


Figure 1: All data holdings in Neotoma: 22,012 datasets from 12,656 sites.

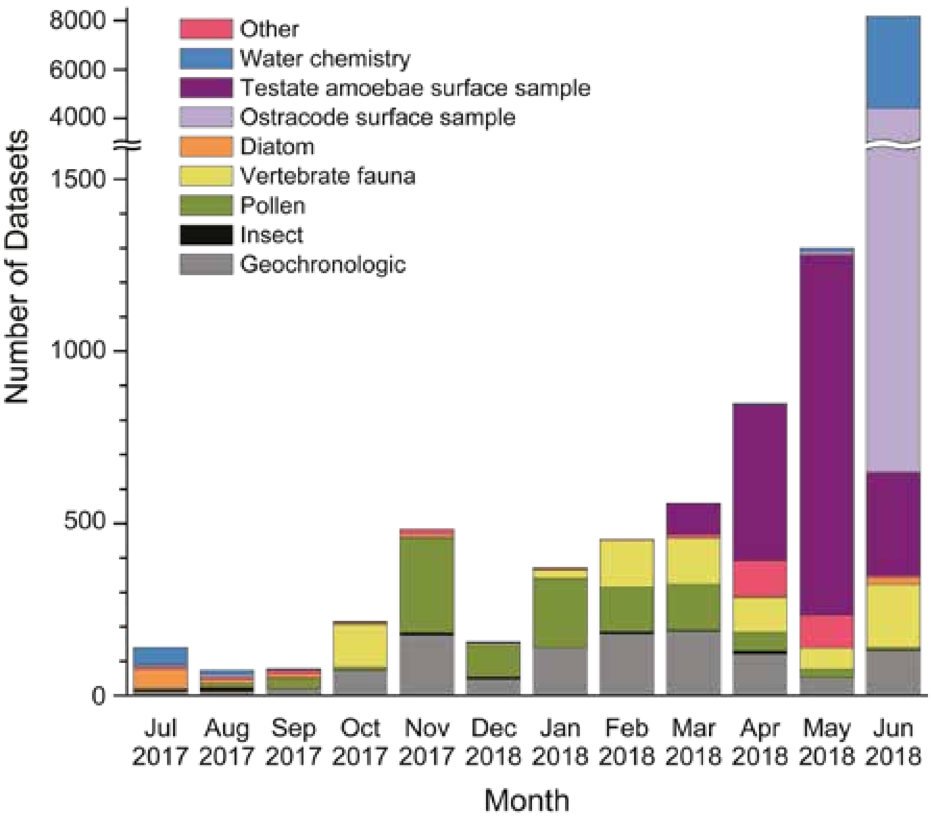


Figure 2: Recent dataset uploads to Neotoma by data type and month.

data stewards, who are appointed by the leaders of the various constituent databases.

Neotoma began by incorporating data from existing databases such as NAPD, EPD, and FAUNMAP. However, Neotoma's design model is flexible and expandable, with many open pathways for participation by new members, data contributors, stewards, and research communities. The Neotoma data model supports, or can be extended to support, any kind of paleoecological or paleoenvironmental data from sedimentary archives (Williams et al. 2018). As of 30 June 2018, Neotoma held 22,012 datasets from 12,656 sites (Fig. 1). New datasets are added almost daily at an increasing rate over the past year (Fig. 2).

Over the past year, there has been considerable push to upload surface samples for aquatic proxies, specifically testate amoebae, diatoms, and ostracodes. The latter two proxies include paired water samples, which comprise "calibration" datasets for quantitative calibration of water chemistry from diatom or ostracode assemblages. As of 30 June 2018, uploads include 1886 testate amoebae surface samples, 640 diatom surface samples, 4515 ostracode surface samples, and 5297 water chemistry samples. The ostracode surface samples have been ported from NANODe and from the Canadian Museum of Nature-Delorme Ostracoda-Surface Samples database. Most of the samples have been uploaded from the Delorme database: 3769 ostracode samples and 3776 water chemistry samples. Although these samples are from other databases, they are not ported en masse, but are subjected to the validation procedures to ensure data quality and compliance with Neotoma metadata standards.

Major efforts have been undertaken to upload data from the EPD (Giesecke et al. 2016) and FAUNMAP 2 (Uhen et al. 2013) databases and to inventory and upload pollen data from Latin America. EPD data contributed before 2007 were included in the Global Pollen Database (GPD), which was available from the World Data Center for Paleoclimatology at NOAA. Following a workshop in November 2017, EPD stewards have uploaded to Neotoma 881 new pollen datasets from 685 sites. After the new datasets are uploaded, the older EPD data ported to Neotoma from GPD will be replaced by the current EPD data, which include many updates and new age models. The original FAUNMAP database (FAUNMAP 1), was an initial compilation into Neotoma. The FAUNMAP 2 database, which includes Canada, Alaska, and the Pliocene (Blancan land mammal age), was compiled but never released nor fully vetted. Since November 2017, Allison Stegner and Mona Colburn have uploaded about a third (1009) of the FAUNMAP 2 datasets to Neotoma. For Latin America, Suzette Flantua and colleagues (Flantua et al. 2013, 2015, 2016) have inventoried pollen and associated geochronological data, and in 2017 over 50 new LAPD pollen datasets from Colombia were uploaded to Neotoma, including important, classic datasets from Thomas van der Hammen and Henry Hooghiemstra.

Another recent improvement particularly relevant for vertebrate fauna, but also other data types, is the ability to store data about individual specimens, including taxonomic and element identification, and museum catalog numbers. Other data can then be associated with these specimens, including radiocarbon dates, GenBank sequence identifiers, and isotopic measurements. In recent years, many high-quality AMS radiocarbon dates on purified collagen have been published (e.g. Widga et al. 2017), and many of these are from sites that are already in Neotoma. These new radiocarbon dates can now be added to existing or new geochronological datasets, and new age models can be built. AMS dates on identified plant macrofossils also comprise another valuable temporal record of taxon occurrences.

The flexible and expandable Neotoma data model has prompted the formation of cooperatives for data types that previously had no appropriate database. Two, in particular, are working groups for stable isotopes and organic biomarkers. The data model of Neotoma has been expanded to accommodate these proxies, and the input software has been modified to upload and validate them. Test datasets have been uploaded, and these holdings should increase during future data mobilization campaigns. We welcome inquiries from researchers interested in contributing data or launching new constituent databases. The continued growth of Neotoma in terms of data holdings and data types will increasingly enable and support paleoenvironmental reconstructions, building upon those first initiated by von Post, Szafer, Firbas, and their contemporaries in the pre-computer era.

AFFILIATIONS

[1]Department of Earth Sciences, University of Minnesota, Minneapolis, USA
[2]School of Natural Sciences, University of California, Merced, USA
[3]Department of Palynology and Climate Dynamics, University of Göttingen, Germany
[4]Department of Geosciences, Pennsylvania State University, State College, USA
[5]Department of Geology, Kent State University, USA
[6]Department of Geography, University of Wisconsin–Madison, USA

CONTACT

Eric C. Grimm: eric.c.grimm@outlook.com

REFERENCES

COHMAP Members (1988) Science 241: 1043-1052

FAUNMAP Working Group (1994) Illinois State Mus Sci Pap 25: 1-690

Firbas F (1949) Spät- und nacheiszeitliche Waldgeschichte Mitteleuropas nördlich der Alpen. G. Fischer, 480 pp

Flantua SGA et al. (2013) PAGES news 21: 88

Flantua SGA et al. (2015) Rev Palaeobot Palynol 223: 104-115

Flantua SGA et al. (2016) Clim Past 12: 387-414

Giesecke T et al. (2016) PAGES Mag 24: 48

Szafer W (1935) Bulletin de l'Académie Polonaise des Sciences et des Lettres, Série B 1: 235-239

Uhen MD et al. (2013) J Vert Paleontol 33: 13-28

von Post L (1924) Geol Fören Förh 46: 83-128

Widga C et al. (2017) Boreas 46: 772-782

Williams JW et al. (2018) Quat Res 89: 156-177

Wright HE Jr et al. (1993) Global climates since the last glacial maximum. University of Minnesota Press, 584 pp

doi.org/10.22498/pages.26.2.66

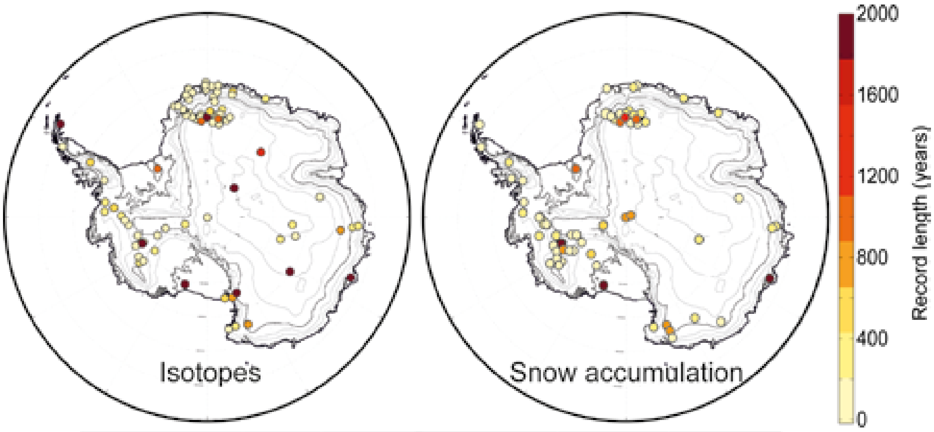
Wrangling data from short Antarctic ice cores

Barbara Stenni[1] and Elizabeth R. Thomas[2]

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.



Data centre	Host country	Website	Number of records used in Stenni et al. 2017	Number of records used in Thomas et al. 2017
The World Data Center PANGAEA	Germany	https://www.pangaea.de/	22 (28%)	32 (40%)
National snow and Ice data centre (NSIDC)	USA	https://nsidc.org/data/submit/intro	n.a.	20 (25%)
NOAA-WDS Paleoclimatology	USA	https://www.ncdc.noaa.gov/data-access/paleoclimatology-data	13 (16%)	2 (3%)
UK Polar Data Centre	UK	https://www.bas.ac.uk/data/uk-pdc/	n.a.	2 (3%)
U.S. Antarctic Program Data Center	USA	http://www.usap-dc.org/	17 (22%)	n.a.
Autralian Antarctic Data Centre	AU	https://data.aad.gov.au/	1 (1%)	n.a.
Not stored in a public repository			26 (33%)	23 (29%)

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.

Experience collating ice-core data

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

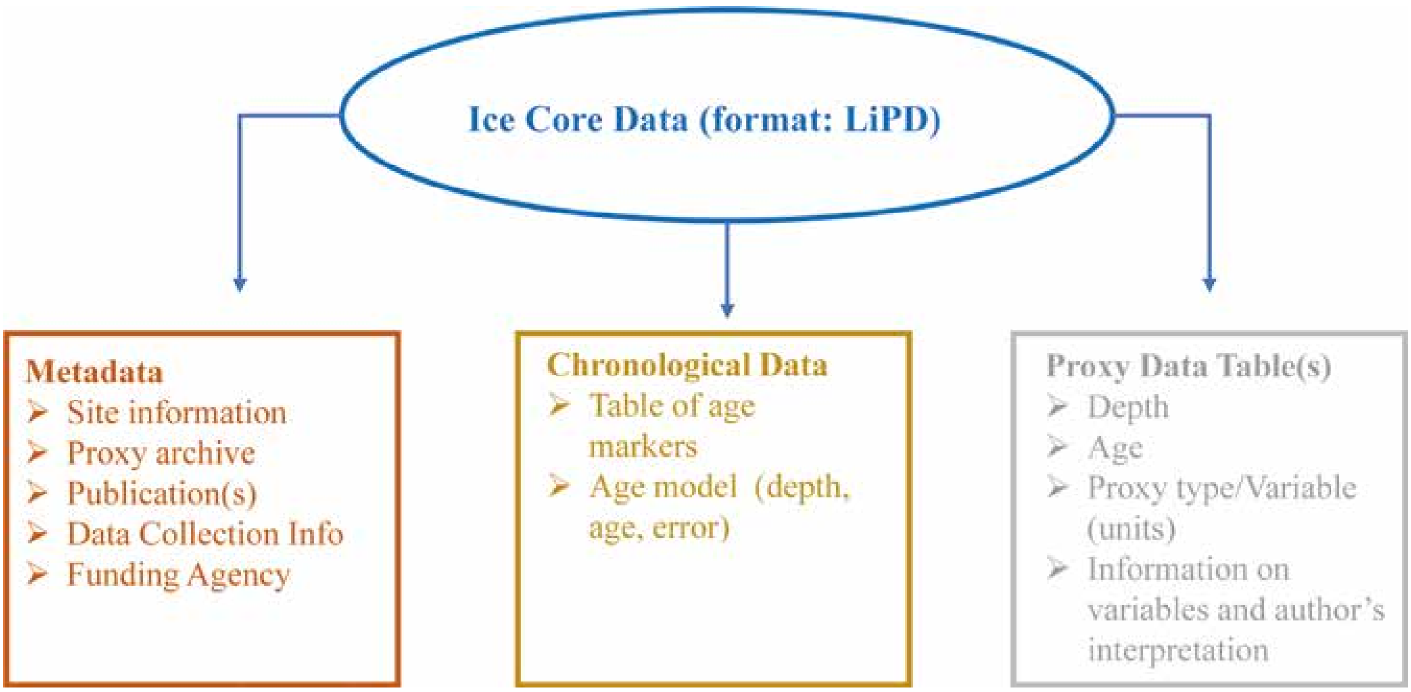


Figure 2: LiPD data container (McKay and Emile-Geay 2016) used in our regional isotopic and snow-accumulation reconstructions. This flexible framework can allocate unlimited number of climate proxy data types (sea ice, wind strengths, etc.), author interpretations and metadata fields.

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

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

[2]Ice 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

doi.org/10.22498/pages.26.2.68

Paleodata for and from archaeology

Timothy A. Kohler[1,2,3], P.I. Buckland[4], K.W. Kintigh[5], R.K. Bocinsky[3,6,7], A. Brin[8], A. Gillreath-Brown[1], B. Ludäscher[9,10], T.M. McPhillips[9], R. Opitz[11] and J. Terstriep[10]

Archaeology depends on, and generates, proxy paleoclimatic and paleoenvironmental data. We review various initiatives, most quite recent, by which archaeologists seek to make these data more readily discoverable and useful, to facilitate the cumulation of research.

Since the birth of the discipline in the mid-19th century, archaeologists have recognized that climate variability has a role in explaining the locations, densities, and practices of human populations. The variability archaeologists could then recognize was the large and dramatic sort involving the ebb and flow of glaciations that altered coastlines and changed the areas people could access, and the distributions of plants and animals they depended on (e.g. Lartet 1861; Lubbock 1890).

Along with related advances in the Earth sciences, the development of palynology in the late 19th century, tree-ring dating and dendroclimatology ca. 1930, and radiometric dating in the mid-20th century greatly increased the scope and chronological precision of paleoclimatic proxies. By the mid-20th century, studying fauna, flora, sediments, and other residues from archaeological sites became recognized subdisciplines (e.g. zooarchaeology, paleoethnobotany, geoarchaeology) and standard archaeological practice. Archaeological prospection and excavation increasingly include investigation of bogs, lakes, or packrat middens to assess local environmental change. Today, many of the "grand challenges" facing archaeology involve understanding the range of human responses to climate change and human manipulation of the landscape at various scales (d'Alpoim Guedes et al. 2016; Kintigh et al. 2014).

Barriers to addressing these challenges include lack of access to and understanding of climate-change data relevant to studies of cultural change. Environmental data from archaeological sites yield an anthropocentric view of the past, since they result from human activities including resource harvesting, hunting, and exchange. Activities in and around sites are, however, always subject to external factors; occupants' responses to changes in climate and environment will be reflected as changes in materials excavated from sediments of different ages. Indeed, the ensemble of excavated sites constitutes a "Distributed Observational Network of the Past" (DONOP; Hambrecht et al., in press) that provides the most direct evidence of our long-term interactions with our environments. As the Anthropocene debate has emphasized, human-nature interactions are not recent, simple, one-way, or local. People have been dramatically changing landscapes for over 10,000 years (Smith and Zeder 2013). Any study of paleoclimate, paleoenvironment, or paleobiodiversity, especially using broad-scale aggregated data, must evaluate the potential for human influences on proxy data used to infer natural change or variability (e.g. Li et al. 2014). As our only available line of evidence on past human and social responses to climatic variability, lessons from archaeology are critically important to forming future responses to climatic variability (Jackson et al. 2018). But just as archaeology studies climatic variability, climate change can destroy sites or their contents: we are rapidly losing archaeological data through erosion, rising sea levels, and thawing of permafrost (Hollesen et al. 2018). There is an urgent need for collecting and curating more data before key sediment archives are lost forever.

Major current efforts to curate open archaeological data

Although archaeologists have been using databases for decades, these were often project-oriented systems with short lifespans. Systematic initiatives to curate archaeological data have appeared in the last two decades, including Digital Antiquity (digitalantiquity.org) and its tDAR database (tdar.org), centered on, but not limited to, US heritage resources; Open Context (opencontext.org) (Wells et al. 2014); and the Archaeology Data Service (archaeologydataservice.ac.uk), the accredited digital repository in the UK for heritage data. Some national data services provide archaeological data, including DANS in the Netherlands (easy.dans.knaw.nl/ui) and the Swedish National Data Service (snd.gu.se/en). The Ariadne infrastructure (ariadne-infrastructure.eu) is working towards providing a single data-discovery service for all European resources. Although these archives are not specifically oriented towards paleoclimatic data, they contain much data important for understanding past environmental conditions and changes. As their interfaces are rarely designed with this in mind, considerable processing may be required to achieve paleoenvironmental insights. For example, using these data requires coping with the complexities of archaeological stratigraphy and possibly integrating archaeological dating with age-depth modeled reconstructions. Systems for standardized ontological mapping between datasets, such as tDAR's

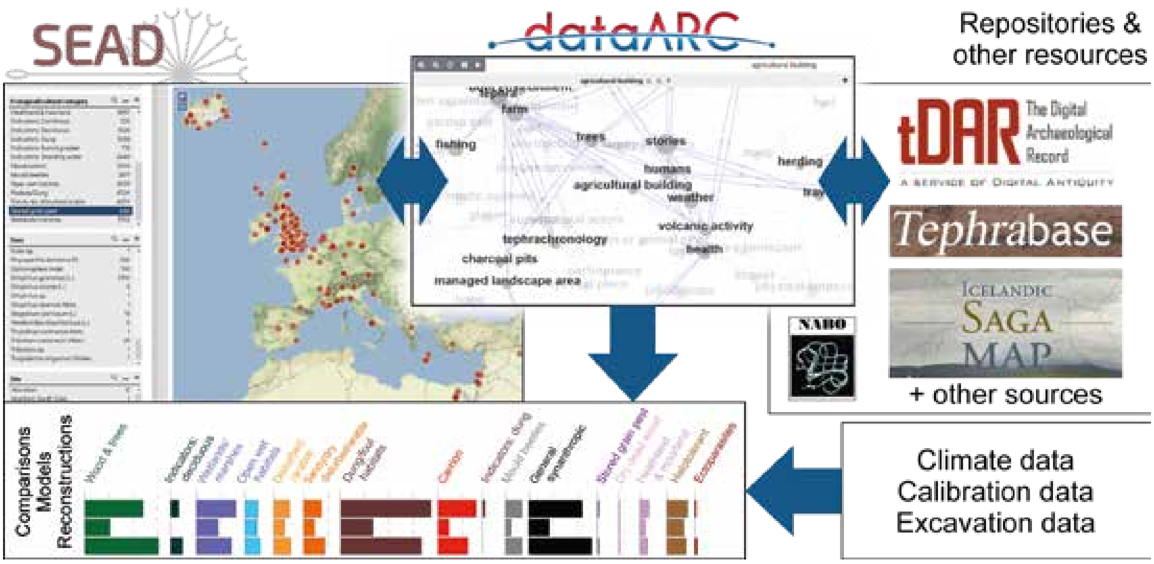


Figure 1: Data-driven research process using archaeological resources. Stored grain pest or parasite occurrences are extracted from SEAD; further information linked through 'agricultural buildings' is extracted from sources identified using DataArc's concept map; the results are visualized as environmental changes across a series of samples.

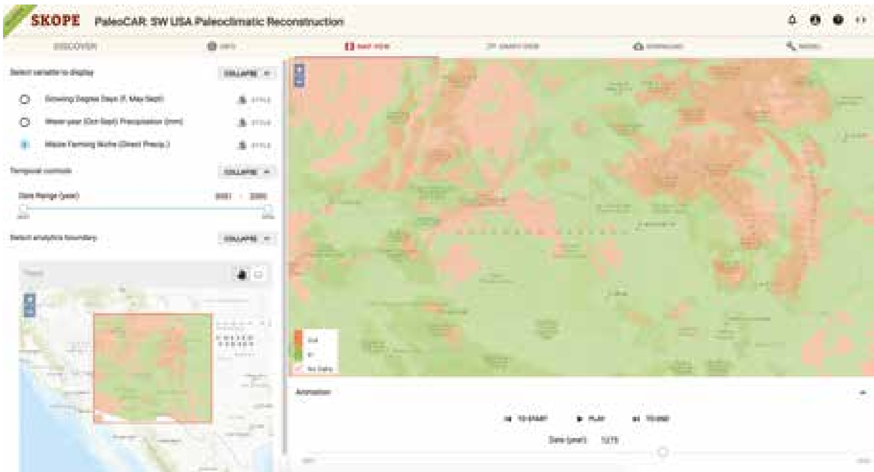


Figure 2: The SKOPE tool depicting a PaleoCAR maize-farming-niche reconstruction. Users can specify region of interest, generate time-series graphs, and download data. Future functionality will include calculating summary statistics and smoothing the time series.

data-integration framework (Kintigh et al. 2018), facilitate such processes.

Archaeological services emphasizing paleoenvironmental data

Several new projects move beyond the scope of most archaeological archives to provide data and tools for exploring the relationships between archaeological and environmental data. The Strategic Environmental Archaeology Database (SEAD: sead.se; Buckland 2014; Uhen et al., this issue) is specifically designed to provide research-level open access to proxy environmental data. These include Quaternary fossil insects, plants, bones, soil parameters, dendrochronology, and geochemistry from mainly European archaeological research (currently, some 15,000 datasets). SEAD includes species traits and cultural/environmental classifications that allow searches for and reconstructions of inferred environments or activities and past species distributions. It provides data to Neotoma and the Earth Life Consortium (see this issue) and archaeological data portals including DataARC (data-arc.org). SEAD facilitates multi-proxy approaches, such as tracking the spread of pests and parasites with people, agriculture and climate change (Fig. 1; Panagiotakopulu and Buckland 2017).

DataARC is designed to go beyond multi-proxy databases and suggest innovative links among resources. Essentially an advanced data-discovery tool, currently focused on the North Atlantic region, DataARC links diverse data types through space, time and concept – the latter using a semantic map to interlink higher-level concepts represented by different data or derived products. The suggested linkages not only expose data to users outside of core domains, but also promote novel research using less-obvious interdisciplinary relationships (Fig. 1). Thus DataARC goes further than past traditional archived data-retrieval platforms and federated systems, such as Ariadne, by providing more-advanced exploratory data-analysis tools to an expanded audience.

SKOPE (Synthesizing Knowledge of Past Environments; openskope.org) is designed to provide easy access to paleoenvironmental and paleoclimatic data that have been processed to be readily useful. Some of these datasets have been previously published; others are created through SKOPE. SKOPE focuses on delivering annual, gridded (raster) reconstructions centered on the US Southwest, including:

- High-frequency temperature, precipitation, and maize-farming niche over the last 2000 years, reconstructed for the US Southwest from networks of tree-ring chronologies using the "PaleoCAR" method (Bocinsky and Kohler 2014) (Fig. 2);
- High-frequency Palmer Modified Drought Index over the last 2000 years, reconstructed from tree rings and available as the North American Drought Atlas;
- Elevation data from the Shuttle Radar Topography Mission (SRTM) dataset, available from NASA;
- Contemporary, monthly temperature and precipitation data for the contiguous United States, available from the PRISM Climate Group at Oregon State University.

We plan to add other existing and novel datasets in coming months, including:

- Low-frequency temperature reconstructed from a network of pollen samples available in Neotoma Paleoecology Database (neotomadb.org) using the modern analog technique (MAT; Overpeck et al. 1985), extending to the early Holocene;
- A new temperature reconstruction integrating the high-frequency signal from tree rings with the low-frequency signal from pollen (MAT) through wavelet modulation for Common Era;
- Past species and vegetation community distributions based on the temperature and precipitation fields available in SKOPE;
- Contemporary potential maize productivity fields for several Native American maize landraces.

We are interested in including other legacy paleoenvironmental data that would benefit from the enhanced access and analysis provided by SKOPE.

Conclusions

Archaeological data processed to reveal socio-ecological interactions are essential to understanding past human experience and how today's world was shaped. Archaeological data that inform on paleoclimates or paleoenvironments are more available than many Quaternary scientists likely realize. The projects we describe enhance access to and facilitate use of paleoenvironmental and archaeological data. The authors welcome further collaboration with paleoclimatologists and paleoecologists to address human and climate interactions more comprehensively.

ACKNOWLEDGEMENTS

This material is based in part upon work supported by the US NSF: DataArc, 1637076 and 1439389; Digital Antiquity/tDAR, BCS 17-24713, BCS 1153115, and IIS 0624341; and SKOPE, currently, SMA 1637189 (ASU), SMA 1637155 (UIUC), and SMA 1637171 (WSU) and previously, SMA 1439591 (ASU), SMA 1439603 (UCSD) and SMA 1439516 (WSU). SEAD has been supported by the Swedish Research Council (2007-7494 and 2010-5976), the Swedish Foundation for Humanities and Social Sciences (IN15-0231:1) and Umeå University.

AFFILIATIONS

[1]Department of Anthropology, Washington State University, Pullman, USA
[2]Santa Fe Institute, USA
[3]Crow Canyon Archaeological Center, Cortez, USA
[4]Department of Historical, Philosophical and Religious Studies, Umeå University, Sweden
[5]School of Human Evolution and Social Change, Arizona State University, Tempe, USA
[6]Montana Climate Office, University of Montana, Missoula, USA
[7]Division of Earth and Ecosystem Sciences, Desert Research Institute, Reno, USA
[8]Center for Digital Antiquity, Arizona State University, Tempe, USA
[9]School of Information Sciences, University of Illinois at Urbana-Champaign, USA
[10]National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, USA
[11]School of Humanities, University of Glasgow, UK

CONTACT

Timothy A. Kohler: tako@wsu.edu

REFERENCES

Bocinsky RK, Kohler TA (2014) Nat Comm 5: 5618

Buckland PI (2014) In: Earl G et al. (eds) Archaeology in the Digital Era, Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA)

d'Alpoim Guedes JA et al. (2016) PNAS 113: 14483-14491

Hambrecht G et al. (in press) Quat Int

Hollesen J et al. (2018) Antiquity 92: 573-586

Jackson RC et al. (2018) Glo Env Ch 52: 58-65

Kintigh KW et al. (2014) Am Antiq 79: 5-24

Kintigh KW et al. (2018) Adv Arch Prac 6: 30-41

Lartet É (1861) Nouvelles Recherches sur la Coexistence de l'Homme et des Grands Mammifères Fossiles Réputés Caractéristiques de la Dernière Période Géologique. Ann Sci nat 4e série, t. XV: 177-253

Li J et al. (2014) Quat Sci Rev 99: 112-121

Lubbock J (1890) [1971] Pre-Historic Times as Illustrated by Ancient Remains and the Manners and Customs of Modern Savages. 5th edition. Books for Libraries Press, 640 pp

Overpeck JT et al. (1985) Quat Res 23: 87-108

Panagiotakopulu E, Buckland PC (2017) Quat Sci Rev 156: 23-35

Smith BD, Zeder MA (2013) Anthrop 4: 8-13

Wells JJ et al. (2014) Lit Ling Comp 29: 349-360

doi.org/10.22498/pages.26.2.70

Automated extraction of spatiotemporal geoscientific data from the literature using GeoDeepDive

Jeremiah Marsicek[1], S.J. Goring[2], S.A. Marcott[1], S.R. Meyers[1], S.E. Peters[1], I.A. Ross[3], B.S. Singer[1] and J.W. Williams[2]

Although open data resources are growing, most scientific data remain "dark" (Heidorn 2008), available only in peer-reviewed literature, where the volume and lack of structure for these data resources presents challenges to data retrieval. GeoDeepDive is an expanding digital library with toolkits that allow automated acquisition and management of published and unpublished documents, supporting large-scale text and data mining of published, peer-reviewed journal articles (Peters et al. 2014; geodeepdive.org). Initial projects have demonstrated the utility of GeoDeepDive's large-scale synthetic geoscientific research (Peters et al. 2017), with new efforts underway.

GeoDeepDive provides a corpus of documents that contain a set of user-prescribed keywords (e.g. 'IRD' and 'Pliocene' or 'Pleistocene' or 'Holocene'). Users develop a set of rules to define the kinds of data they wish to retrieve (coordinates, measurements, etc.) from a subset of the matching publications, and write a test application. The application is deployed against the full GeoDeepDive corpus once a user has developed and tested their workflow on the data subset.

Initial work with GeoDeepDive – studying the dynamics of Northern and Southern Hemisphere ice sheets during the Quaternary – has allowed us to leverage publications focusing on ice-rafted debris (IRD). Assembling information from publications documenting IRD at marine drilling sites is a non-trivial task that has traditionally involved painstaking literature compilation (Hemming 2004). GeoDeepDive allowed us to discover and extract information by searching through 7.5 million publications across a range of publishers using an R workflow based on regular expressions and natural language-processing utilities. This work also allows us to develop a general workflow for GeoDeepDive, supporting others who might use it in their future research (Fig. 1).

Future directions

Our GeoDeepDive workflow allows us to extract and plot reliable latitude-longitude pairs from publications reporting IRD events (Fig. 1). We are building a spatial database of IRD events and beginning to extract event ages from the papers. Extracting temporal information from the unstructured peer-reviewed literature is a non-trivial but tractable task using regular expressions and string matching. We are also differentiating primary, original sources from secondary studies that include previously published data, and building a GitHub repository for open code development and sharing (github.com/EarthCubeGeoChron). Next steps include building summary maps of the location, finding specific named IRD events or the timing of IRD deposits found in cores, and continuing development of an R package (github.com/EarthcubeGeochron/geodiveR). The project will result in an IRD database that can provide a better characterization of Northern and Southern Hemisphere ice sheets over the last 5.3 million years. The R package that results from this work will consist of a general set of tools for querying space and time information from GeoDeepDive, allowing other researchers to simply import their own data using their own search logic and output coordinates and subsets of the text relevant to a researcher's particular questions.

An ongoing question in this broad-scale, data-mining project is to determine the appropriate points for human intervention and interpretation, one of many questions discussed at a recent GeoDeepDive user workshop in Madison, USA (geodeepdive.org/workshop2018). These points should be minimized for reasons of scalability, but some features may not be readily automated. Future advances will likely be powered by "centaur" systems combining the relative strengths of human- and machine-learning approaches, which will then provide the basis for new applications of machine-learning methods. We view the GitHub Repository and the R package as building blocks that will serve researchers across the geosciences and allied disciplines.

AFFILIATIONS
[1]Department of Geoscience, University of Wisconsin-Madison, USA
[2]Department of Geography, University of Wisconsin-Madison, USA
[3]Department of Computer Sciences, University of Wisconsin-Madison, USA

CONTACT
Jeremiah Marsicek: jmarsicek@wisc.edu

REFERENCES

Heidorn PB (2008) Library Trends 57: 280-299
Hemming SR (2004) Rev Geophys 42: RG1005
Peters et al. (2014) PLoS ONE 9: e113523
Peters et al. (2017) Geology 45: 487-490

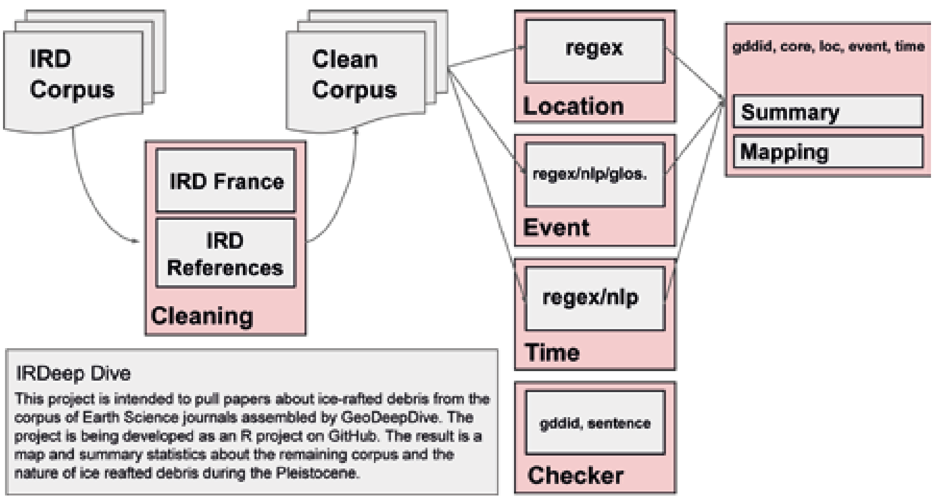


Figure 1: GeoDeepDive workflow used to build a corpus of documents that mention "ice-rafted debris" (IRD), screen a vetted set of the documents, and summarize the documents and relevant information (github.com/EarthcubeGeochron/geodiveR). 'Cleaning' = removing instances of IRD in the affiliations and references sections; regex = regular expression; nlp = natural language processing; glos. = glossary; Checker = a step to ensure sentences contain relevant IRD information; gddid = GeoDeepDive identification key.

doi.org/10.22498/pages.26.2.71

Linked Paleo Data: A resource for open, reproducible, and efficient paleoclimatology

Nicholas P. McKay[1] and Julien Emile-Geay[2]

Paleoclimatology is a remarkably diverse field of research, revolving around hard-won and complex datasets that typically represent hundreds of hours of field work, laboratory analyses and nuanced interpretation. Integrating those diverse datasets to piece together a spatio-temporal understanding of how, when and why climate has changed in the past is a grand challenge of paleoclimatology; one that requires careful handling of these data and their interpretations. Researchers often spend up to 80% of their time collecting, organizing and formatting data, before they can even begin addressing the questions they set out to tackle (Dasu and Johnson 2003). This was certainly our experience, and is why we developed the Linked Paleo Data (LiPD) framework. As the number of records relevant to paleoclimate research continues to grow, and the methodologies for investigating datasets and data networks become more complex, our community cannot afford to continue wasting time on data wrangling when there's so much science to be done!

The linked paleodata solution

The technical details of LiPD are presented in McKay and Emile Geay (2016), but the concept is simple: LiPD provides a flexible structure that contains and describes any paleoclimatic or paleoenvironmental dataset, the metadata that describe the details and complexity of the data (at any level from observations to collections), as well as models that accompany the data and their output, such as age models and their ensemble output. This powers efficient, 21st century scientific workflows, and enables open science and reproducible research.

This is why LiPD has been used by multiple data-intensive PAGES working groups, including the 2k Network Temperature Database (PAGES 2k Consortium 2017), and Iso2k[1]. Being able to rely on consistently structured data with rich metadata has greatly reduced the "time to science" for projects relying on the PAGES 2k database, such as the forthcoming global temperature reconstruction intercomparison[2], and the Last Millennium Reanalysis project.

Having structured and standardized data also enables efficient access to state-of-the-art analysis tools. One example is age-uncertain data analysis using the GeoChronR package[3]. GeoChronR relies on LiPD's capacity to contain and describe age-model ensembles to simplify quantifying the effects of age uncertainty on paleoclimate analysis. For example, quantifying and visualizing the impact of age uncertainty on a calibration-in-time with temperature, both on the regression model and the reconstruction back in time, is greatly simplified with LiPD and GeoChronR[4].

A growing LiPD "ecosystem"

Data standards and formats are only as useful as the breadth of their adoption. Thankfully, a LiPD "ecosystem" of datasets, standards, and tools is emerging (Fig. 1).

Datasets: More than 3000 datasets have now been formatted as LiPD files, largely as part of PAGES working group efforts. These data are archived at WDS-Paleo and LinkedEarth (Gil et al. 2017). LiPD is also well suited to serve as an "interchange format", facilitating the transfer of datasets from researchers to repositories and tools. As LiPD is not tied to any particular repository, initial connectivity with WDS-Paleo and Neotoma has been developed, and two-way interoperability with other repositories, including LacCore, and Open Core Data is forthcoming as part of the Throughput project[5].

Standards: From the outset, LiPD was designed to support "Linked Open Data", an international effort to connect data and concepts and make them broadly accessible through the semantic web[6]. As part of the LinkedEarth project, we created the "LiPD Ontology", the first ontology for paleoclimatology[7]. LiPD also enables community-developed data standards (Emile-Geay and McKay 2016; Emile-Geay et al., this issue), including WDS-Paleo's controlled vocabulary[8].

Tools: A wide range of tools that "speak" LiPD have been developed. This includes the LiPD Utilities, which provide basic functionality for reading, writing and querying LiPD data in R, Matlab and Python, and provides the base-level functionality for more sophisticated packages, including GeoChronR[9] and Pyleoclim[10]. A rich set of interactive, graphical, web-based tools for creating and modifying LiPD files has been created at lipd.net.

CScience, an AI-powered tool for age modeling uses LiPD as an input and output format (Bradley et al., this issue).

LiPD has always been collaborative and open-source, and we look forward to the continued expansion and evolution of these data, standards and tools by the community. To learn more about LiPD, how to use if for your research, and upcoming training opportunities, please visit lipd.net

AFFILIATIONS

[1]School of Earth & Sustainability, Northern Arizona University, Flagstaff, USA

[2]Department of Earth Sciences, University of Southern California, Los Angeles, USA

CONTACT

Nicholas P. McKay: Nicholas.McKay@nau.edu

REFERENCES

Dasu T, Johnson T (2003) Exploratory Data Mining and Data Cleaning. Wiley, 212 pp

Emile-Geay J, McKay NP (2016) PAGES Mag 24: 47

Gil Y et al. (2017) The Semantic Web – ISWC 2017: 231-246

McKay NP, Emile-Geay J (2016) Clim Past 12: 1093-1100

PAGES 2k Consortium (2017) Sci Data 4: 170088

LINKS

[1]pastglobalchanges.org/ini/wg/2k-network/projects/iso2k

[2]pastglobalchanges.org/ini/wg/2k-network/projects/gmst-recon-2k

[3]nickmckay.github.io/LinkedEarth-Neotoma-P418/LE-Neo_UseCase.html

[4]lipdverse.org/regression.htm

[5]throughput-ec.github.io

[6]w3.org/DesignIssues/LinkedData.html

[7]linked.earth/ontology

[8]ncdc.noaa.gov/data-access/paleoclimatology-data/contributing

[9]nickmckay.github.io/GeoChronR

[10]linkedearth.github.io/Pyleoclim_util

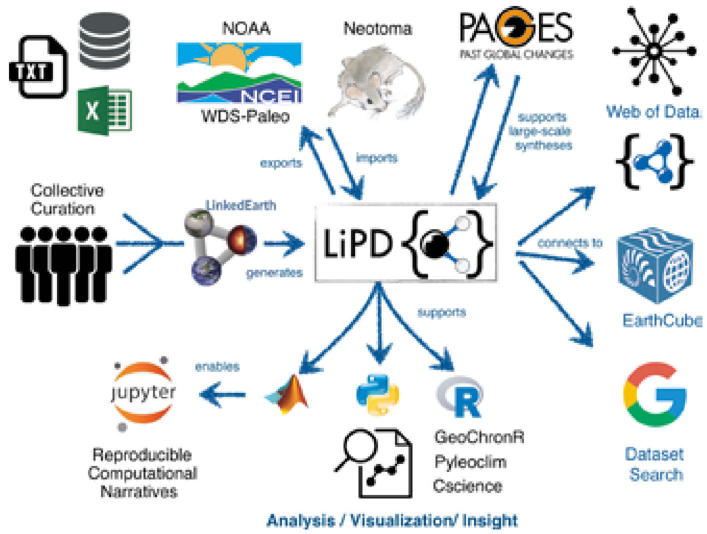


Figure 1: The LiPD ecosystem: a growing network of scientific communities, data repositories, and analysis tools connected and enabled by LiPD.

doi.org/10.22498/pages.26.2.72

CSciBox: Artificial intelligence for age-depth modeling

Elizabeth Bradley, T.H. Nelson and L. Rassbach de Vesine

Artificial intelligence (AI) provides major opportunities for scientific analysis. Automated reasoners can explore problem spaces quickly and alert practitioners to possibilities that they had not considered. As a case in point, we describe the CSciBox system. Working with data from a paleorecord, such as ^{14}C dates from a sediment core (Fig. 1a) or ^{18}O values from an ice core, CSciBox produces a set of age-depth models, plus a description of how each one was built and an assessment of its quality.

The AI field has two branches: symbolic methods capture human reasoning in closed form; statistical methods such as neural networks, aka "machine learning" (ML), fit sophisticated models to sets of labeled examples. Both have strengths and weaknesses. ML methods are powerful, but training them requires a large number of examples. This is problematic in the context of age-depth models, where there is rarely more than one published example for each core. The symbolic AI approach has its own challenges: human reasoning is remarkably difficult to capture in formalized, useful ways. However, an AI system seeded with that kind of knowledge can narrate its choices and explain its actions as it solves problems – an absolutely essential feature for a scientific assistant, and one that ML methods cannot provide.

CSciBox marries these two different types of approaches. Its toolbox includes a number of traditional data-analysis methods, along with a set of statistical methods that model the different underlying physical processes (e.g. sediment accumulation). A symbolic AI engine explores the search space of possible age-depth models: choosing among those methods, invoking them on the appropriate data fields with appropriate parameter values, analyzing the results, making appropriate modifications, and iterating until the results match the scientist's physical understanding of the world.

There can be evidence and reasoning both in favor of and against any given age model. CSciBox uses one of the few AI techniques that handle this situation, "argumentation" (Bench-Capon and Dunne 2007), which involves constructing all arguments for and against each candidate age model and then weighing them against one another (Rassbach et al. 2011). In the case of the data in Figure 1(a), CSciBox reasons from the latitude and longitude of the core to choose the IntCal marine 13 curve (Reimer et al. 2013) and the reservoir-age correction (calib.org/marine), then searches for an age-depth model to fit the calibrated, corrected age points. It first tries linear regression but discards the resulting model because the argument against it (large observed residuals) is stronger than those in favor (consistent slope, no reversals). It then tries piecewise-linear interpolation, producing the age model shown in panel C of the figure, but finds that that, too, is a bad solution (low residuals but inconsistent slope and presence of reversals). CSciBox then builds and evaluates an age-depth model using Bacon (Blaauw and Christen 2011), constructing and balancing arguments about the consistency of the slope (good) and the size of the residuals (small) against the fact that Bacon does not converge to a single distribution – as is clear from Figure 1d – and that some of the age points are outside the error bounds.

Like many powerful tools, Bacon's actions are guided by a number of free parameters. CSciBox encodes a number of rules that capture how experts tune those parameter values, which it uses to explore the parameter space and improve the Bacon model. This is a major advance; tools like Bacon are very powerful, but they can be difficult to use. At the end of the exploration process, CSciBox presents the strongest model to the user, together with a full narration of the process involved in building it. CSciBox uses LiPD (McKay and Emile-Geay 2016) to store all of this information (data and metadata), making the analyses completely documented and reproducible, as well as smoothly interoperable with any other LiPD-enabled software. Like LiPD, CSciBox is open-source; see Bradley et al. (2018) for code and documentation.

ACKNOWLEDGEMENTS

This material is based upon work sponsored by the National Science Foundation (#1245947).

AFFILIATIONS

Department of Computer Science, University of Colorado at Boulder, USA

CONTACT

Elizabeth Bradley: lizb@cs.colorado.edu

REFERENCES

Bench-Capon T, Dunne PE (2007) Artif Intell 171: 619-641

Blaauw M, Christen JA (2011) Bayesian Anal 6: 457-474

Bradley E et al. (2018) CSciBox - building age models of paleorecords. Zenodo

McKay NP, Emile-Geay J (2016) Clim Past 12: 1093-1100

Rassbach L et al. (2011) AI Mag 32: 69-78

Reimer P et al. (2013) Radiocarbon 55: 1869-1887

Xie R et al. (2012) Paleoceanography 27: PA002337

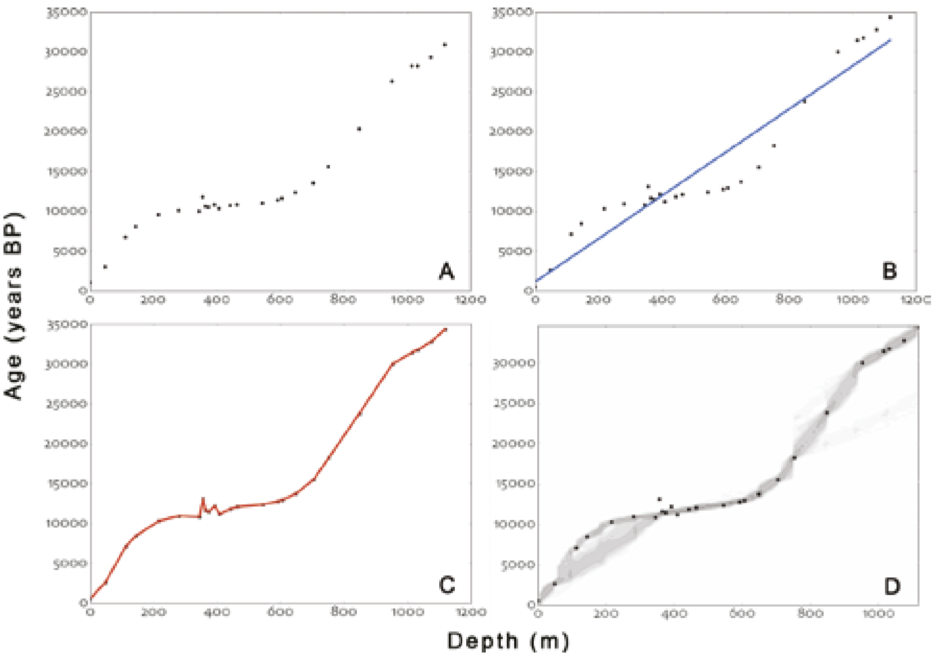


Figure 1: Screenshots of CSciBox building an age model for a marine-sediment core from the Gulf of Mexico (Xie et al. 2012). **(A)** raw ^{14}C ages (•), **(B)** linear regression, **(C)** piecewise-linear interpolation, and **(D)** Bacon model. All plots are age in years BP vs. depth in meters. (■) indicates an age point that has been corrected for reservoir age and undergone a CALIB-style calibration.

doi.org/10.22498/pages.26.2.73

Accelerating progress in proxy-model synthesis using open standards

Gregory Hakim[1], S. Dee[2], J. Emile-Geay[3], N. McKay[4] and K. Rehfeld[5]

Weather prediction has undergone a "quiet revolution" in recent decades (Bauer et al. 2015), fueled by increasing observations and the capability to assimilate them into increasingly sophisticated numerical models. Paleoclimatology today is at the cusp of such a revolution, moving away from a focus on single-site studies to the assimilation of global, multiproxy data streams into climate models (e.g. Brönnimann et al. 2013; Goosse 2016; Hakim et al. 2016; Franke et al. 2017) thanks to (1) advances in data assimilation (DA) methodology; (2) open, standardized paleoclimate datasets; and (3) proxy system models (PSMs). A critical element of DA that allows this synthesis involves mapping information from climate models to proxy measurements through PSMs (e.g. Dee et al. 2015). DA weighs the information from proxies against a climate-model simulation of the proxy value, and spreads that information in space and to other climate variables (Fig. 1). Future progress depends strongly on openness and standardization of paleoclimate proxy data, so we describe here the dependence of DA on open paleoclimate data, emerging standards, and ideas for accelerating progress.

Openness in data sharing and standardization

The currently highly heterogeneous nature of the proxy records is the main limitation to DA progress. Improvements involve three components common in data science: (1) data distribution, (2) data standardization, and (3) data-revision tracking.

Over the past two decades, distribution of paleoclimate proxy data has migrated from individual scientists sharing their data to centralized data centers, such as the World Data Service for Paleoclimatology, the International Tree Ring Databank, Neotoma, and Pangaea; however, large amounts of data have not yet been transferred to public repositories. Curated versions of paleoclimate data from these centers and from the literature, through quality control and screening, have proven critical to recent synthesis efforts (e.g. PAGES 2k Network projects). However, because these curated versions do not track uniquely from the original proxy data, future efforts either have to work with these "forks" from the source, or substantially duplicate effort by returning to the original data. Having the ability to track data from the source through the forks would allow for robust branching without returning to sources to begin anew.

Climate model output is available in standard format (NetCDF), with conventions for units and variable naming (cfconventions.org). Ongoing efforts combine PSMs in a standardized and open-source framework (e.g. PRYSM; Dee et al. 2015), but such standardization is just beginning for paleoclimate data. For example, the Linked Paleo Data (LiPD) format (McKay and Emile-Geay 2016 and this issue) provides a universal, flexible container for a wide range of paleoclimate data. Because LiPD's structure and terminology are inspired by the PSM framework, it is a natural format for DA codes, since LiPD metadata can direct PSM selection for a particular dataset. Although the emergence of LiPD offers the potential for a large increase in efficiently using proxy data in DA applications, most proxy data remain to be converted to LiPD format.

Future directions

Data standardization is the area where the greatest immediate impact can be experienced. Widespread adoption of LiPD across proxy archives would greatly facilitate the reuse of proxy data and synthesis efforts, as would standardized revision histories. As much as revision tracking has transformed productivity in software development with distributed version control software such as Git, similar practices for proxy data are compelling.

One speculative future direction involves decentralized ledgers for proxy data. Cryptographically secure ledgers, such as Bitcoin's blockchain, contain unalterable revision history that do not depend upon a central authority. For paleoclimate proxy data, this technology could be used to allow anyone to correct errors and, through a consensus algorithm, add revisions to the public ledger. One can imagine motivating public participation with micropayments of Bitcoin. A small amount of funding distributed in this way could offer rapid progress to cleaning the "bugs" from proxy data archives, with the added benefit of citizen scientist participation in paleoclimate research.

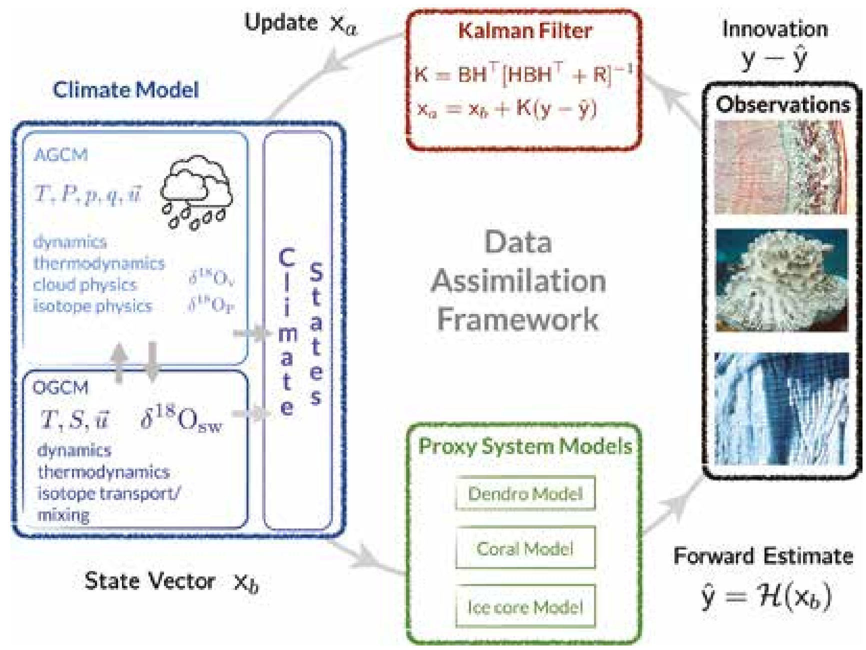


Figure 1: DA uses climate variables to estimate proxy values using PSMs, which can then be compared with the actual proxy values. The difference between these values (new information about the climate state) is weighted ("K") by the error in the proxies relative to the estimate from the climate model; most important, DA also spreads this information in space and to other climate variables. From Hakim et al. (2016).

AFFILIATIONS
[1]Department of Atmospheric Sciences, University of Washington, Seattle, USA
[2]University of Texas, Institute for Geophysics, Austin, USA
[3]University of Southern California, Los Angeles, USA
[4]Northern Arizona University, Flagstaff, USA
[5]Institute of Environmental Physics, Ruprecht-Karls-Universitaet Heidelberg, Germany

CONTACT
Gregory Hakim: ghakim@uw.edu

REFERENCES

Bauer P et al. (2015) Nature 525: 47-55
Brönniman S et al. (2013) PAGES news 21: 74-75
Dee S et al. (2015) J Adv Model Earth Sys 7: 1220-1247
Franke J et al. (2017) Sci data 4: 170076
Goosse H (2016) J Adv Model Earth Sys 8: 1501-1503
Hakim GJ et al. (2016) J Geophys Res Atmos 121: 6745-6764
McKay NP, Emile-Geay J (2016) Clim Past 12: 1093-1100

doi.org/10.22498/pages.26.2.74

Outreach and educational opportunities created by open-data resources

Amy Myrbo[1], S. Loeffler[1], A.L.C. Shinneman[2] and R. McEwan[3]

Open-data-based geovisualizations including Flyover Country can engage students and the general public with science. Mobile technology allows data and visualizations to be brought into the field, facilitating discovery while users explore the natural world in real life.

Software tools that harness open-paleodata resources are opening new frontiers for education, public outreach, and citizen science. The ability to easily and seamlessly map and visualize disparate proxy datasets from multiple databases enables not only new research insights into past Earth system dynamics, but new opportunities for formal and informal education. The free, National Science Foundation (NSF)-funded Flyover Country mobile app (flyover-country.io) takes such an approach; displaying the current location engages the user with place-based, on-demand access to real data and to ongoing scientific research. Bringing students and the general public science about the places they love – sometimes literally in their own backyard – supports place-based education (Semken 2005; Apple et al. 2014) and outreach, thereby acting as gateways for new audiences into science and scientific ways of thinking, as well as interaction with publicly funded scientific research. The ability to easily cache data to a mobile device means that users can take advantage of scientific data even while offline, outdoors, in remote settings. Here, we briefly summarize how the intersection of open scientific databases and mobile platforms is creating new opportunities for traditional classroom teaching, self-directed field trips by undergraduates or casual travelers, and citizen science.

Classroom-oriented resources increasingly use content from open, community-curated scientific data resources (CCDRs) as foundations for students to explore past global changes. For example, as of July 2018, SERC (the Science Education Resource Center at Carleton College, USA) hosted over 30 classroom activities using either the Neotoma Database (Williams et al. 2018) or the Paleobiology Database (PBDB; Uhen et al. 2013). The ability to overlay multiple datasets in a single visualization – floral and faunal records; ice-sheet and sea-level positions; tectonic plate paleolocations, etc. – is available directly in some CCDRs' online resources (e.g. Neotoma Explorer, PBDB Navigator, GeoMapApp), so they are readily adapted to support classroom lessons.

Mobile apps such as Flyover Country (Fig. 1) and Rockd (Schott 2017) extend this functionality to the outdoors. Students can use these apps to access data about a location while in the classroom or field, supporting place-based education (Semken et al. 2005; Apple et al. 2014), which uses the "hook" of a learner's hometown or homeland to communicate the relevance and excitement of the sciences. Place-based education is an especially effective approach to engaging and entraining students into the geosciences, especially members of groups who strongly identify with places, such as many Indigenous people, recent immigrants, and members of rural communities. Whether the setting is a remote landscape or a concrete-covered city, geoscience plays a fundamental role (e.g. Sanderson 2009; Broad 2018): the location, layout, economy, population structure, human history, and amenities of a place are strongly influenced by the underlying geology, climate, ecology, and geomorphology.

Flyover Country fills a need for travelers, delivering interesting information about the landscape as seen out the airplane, train, or car window. The app calls data from Macrostrat (Peters et al. 2018), Neotoma, PBDB, Wikipedia, and OpenCoreData (opencore-data.org). Information in the app has global coverage, with data density dependent upon content in the databases. Flyover Country also links and enriches sparse datasets: for instance, taxon name and age for a fossil specimen returned from PBDB are used to call to

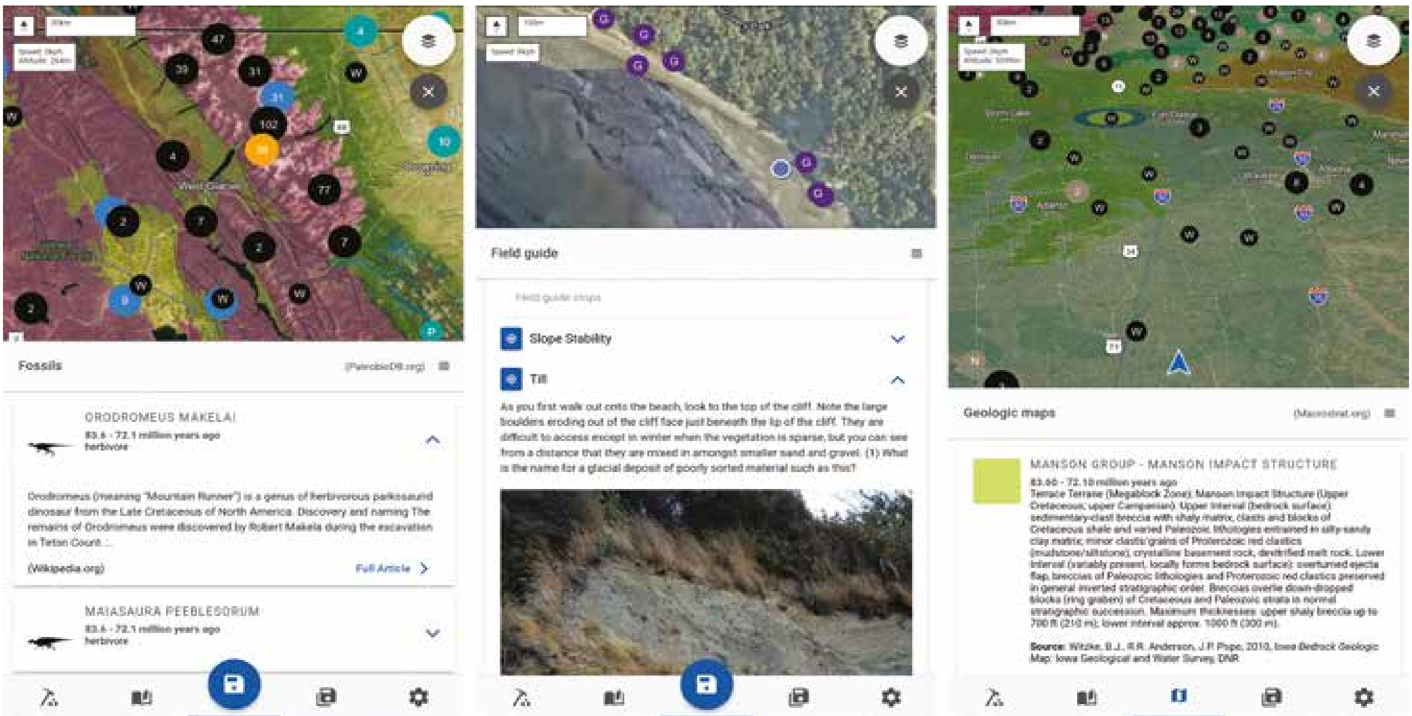


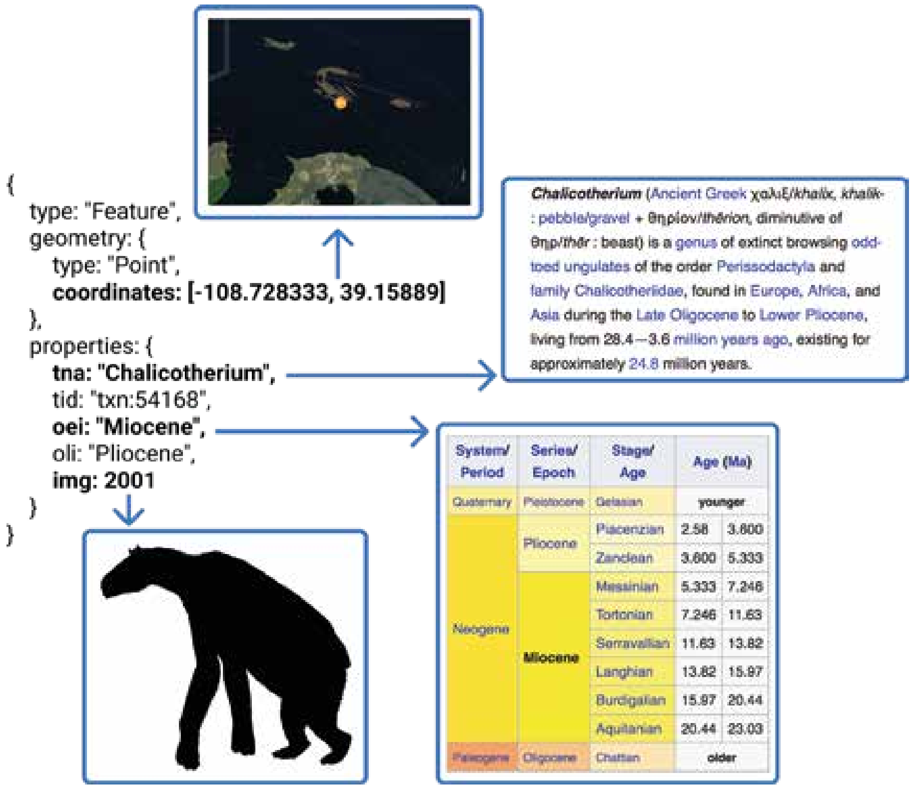
Figure 1: Examples of the Flyover Country mobile app interface. **Left**: Dots locating fossil localities from PBDB (teal and orange [currently selected]), Wikipedia articles (black), and continental sediment cores (blue), overlain on a satellite image and geologic map. Wikipedia articles on the dinosaur taxa found near the orange dot are shown below. **Center**: An instructor-submitted field-trip guide showing field trip stops (purple dots) and the user's current location (blue dot with white outline). Field-trip content and photo are shown below. **Right**: Navigation mode changes the frame of reference to match direction of travel. Gray dots are fossil localities from Neotoma, black dots from Wikipedia. Geologic map unit information from Macrostrat is shown below.

other resources, including PhyloPic (phylopic.org) for a silhouette of the taxon reconstructed and Wikipedia for the articles about that taxon and geologic time period (Fig. 2). Because maps are readily understood by people of all languages and countries, Flyover Country is appealing to users worldwide. The app's large userbase (240,000 downloads as of August 2018), reflecting a high level of interest from the general public in paleodata and related scientific information, promotes a virtuous cycle that encourages additional data resources to expose their data through Flyover Country. New data sources being added include datasets from IEDA databases (iedadata.org), state and federal geological surveys, additional paleobiological resources, tectonic plate reconstructions from GPlates (gplates.org), physiographic regions from Natural Earth Data (naturalearthdata.com), and articles from other languages' Wikipedia databases.

Mobile tools like Flyover Country also create exciting new opportunities to enrich field experiences for undergraduates and the undergraduate geoscience curriculum. Field experiences at the introductory level have been shown to improve student satisfaction and likelihood to select and persist in the geosciences (Karabinos et al. 1992; Wilson 2018; Wolfe 2018). However, logistical challenges at institutions (Bursztyn et al. 2017), as well as work, family duties, and financial hardship particularly common to students from groups underrepresented in the geosciences (Bueno Watts et al. 2014) can restrict students' ability to participate in class field trips. Faculty may provide alternative activities that are classroom- or homework-based, but these miss the opportunity to get students out into the field, making observations, developing questions, seeing real rocks and organisms, and connecting course content with the world they experience. Using mobile devices to present instructor-developed field-trip content and the user's GPS location allows students to take independent field trips that fit their own schedules. Such an approach can engage students in ways similar to instructor-led trips, which could broaden participation in the field component of undergraduate courses, and thus help broaden participation in the geosciences overall.

Mobile apps can also access and help disseminate narrative- and location-based resources, such as field-trip guides from conference guidebooks and undergraduate courses. For example, a database of georeferenced field-trip guides sourced mainly from Geological Society of America guidebooks has been developed as part of Flyover Country and is available for use in any software tool. This database is open for new contributions at z.umn.edu/fcfg. Furthermore, these field trips are spatially discoverable, so users can find them without prior knowledge or possession of the guidebooks or course material.

Mobile apps can also be used to generate data, not just disseminate it, by encouraging citizen science (e.g. Bonney et al. 2014). For example, the Rockd app crowdsources images of outcrops that can ultimately be aggregated and used in scientific studies, and Flyover



System/ Period	Series/ Epoch	Stage/ Age	Age (Ma)	
Quaternary	Pleistocene	Gelasian	younger	
Neogene	Pliocene	Piacenzian	2.58	3.600
		Zanclean	3.600	5.333
	Miocene	Messinian	5.333	7.246
		Tortonian	7.246	11.63
		Serravallian	11.63	13.82
		Langhian	13.82	15.97
		Burdigalian	15.97	20.44
		Aquitanian	20.44	23.03
Paleogene	Oligocene	Chattian	older	

Figure 2: Flyover Country enriches sparse data by using text returned from PBDB to call to other data resources: in addition to plotting the coordinates of the fossil locality on the map, the taxon (*Chalicotherium*) and age (Miocene) are used to obtain the relevant Wikipedia articles, and the PhyloPic ID is used to obtain a silhouette of the extinct organism.

Country is connecting to NASA's GLOBE Observer app to crowdsource ground-cover photographs to improve remote sensing information. As in well-established biodiversity and conservation-oriented citizen-science projects such as eBird (ebird.org) and iNaturalist (inaturalist.org), upload of new data to the project database can be facilitated through apps, and the user can see their data visualized on the map in the app in near-real time alongside data from both citizen scientists and professionals.

The future opportunities for development of education and outreach using open paleodata are equally exciting. App development is well within the capacity of small research/outreach teams with both scientific and software development skills; many materials supporting self-taught coding are available online (e.g. edx.org, which hosts courses from many institutions). Augmented reality offers new ways of overlaying data onto outcrops and landscapes viewed by the mobile device's camera (similar to star-map apps for the night sky, or the PeakFinder mobile app, peakfinder.org), further helping students and the public understand the geoscientific way of seeing. Similarly, inclusion of new high-quality narrative and visual content such as Esri Story Maps (storymaps.arcgis.com) and additional professional and crowdsourced field-trip guides bring decades of research to these new audiences. The developing field of semi-automated "text leveling" (e.g. newsela.com) when combined with geoscience semantics and ontologies, could help "translate" large volumes of content written for experts into material suitable for the general public. The availability of paleo and geoscience data in integrated travel systems (e.g. inflight entertainment, passenger trains and buses, self-driving cars) can bring the excitement of discovery, closely embedded in a sense of place, to new audiences worldwide.

ACKNOWLEDGEMENTS

Flyover Country is supported by eight NSF awards, including 1338322, 1462297, 1541800, 1550913, and 1643277.

AFFILIATIONS

[1]CSDCO/LacCore and Department of Earth Sciences, University of Minnesota, Minneapolis, USA
[2]Interdisciplinary Arts and Sciences, University of Washington - Bothell, USA
[3]Academic Health Center - Information Systems, University of Minnesota, Minneapolis, USA

CONTACT

Amy Myrbo: amyrbo@umn.edu

REFERENCES

Apple J et al. (2014) J Geosci Edu 62: 1-4

Bonney R et al. (2014) Science 343: 1436-1437

Bueno Watts N et al. (2014) In: Dalbotten D et al. (Eds) Future Earth - Advancing Civic Understanding of the Anthropocene, AGU. 67-75

Bursztyn N et al. (2017) GSA Today 27: 4-11

Broad WJ (2018) How the ice age shaped New York." New York Times, accessed 22 August 2018

Karabinos P et al. (1992) J Geol Edu 40: 302-305

Peters SE et al. (2018) Geochem Geophys Geosys 19: 1393-1409

Sanderson E (2009) Mannahatta: A Natural History of New York City. Abrams, 352 pp

Schott R (2017) Bulletin 67: 1-4

Semken S (2005) J Geosci Edu 53: 149-157

Uhen MD et al. (2013) J Vert Paleontol 33: 13-28

Williams JW et al. (2018) Quat Res 89: 156-177

Wilson C (2018) Status of recent geoscience graduates 2017. American Geosciences Institute, 49 pp

Wolfe BA (2018) J Geosci Edu 66: 36-54

doi.org/10.22498/pages.26.2.76

Forecasting long-term ecological dynamics using open paleodata

Jason S. McLachlan[1] and the PalEON Project[2]

As humans alter the environment in unprecedented ways, forecasts of the future state of ecosystems become increasingly important. Good forecasts require skilled models, and paleoecological data have played an important role validating retrospective model hindcasts. Modern analytical approaches, like data assimilation, now allow paleodata to be explicitly incorporated into forward-looking model forecasts. In particular, paleodata can provide unique empirical constraints on forecasts of slow or infrequent events that are difficult to constrain with more recent instrumental measurements.

"Forecasting" has a specific meaning here, known from meteorology (Dietze 2017). An ecological forecast is a set of quantitative predictions about the most-likely future state (or reconstructed hindcast) of an ecosystem. A forecast is comprised of both models and data, each of which is incomplete and flawed: Models are simplified and imperfect representations of reality, and paleodata are noisy, geographically sparse, usually indirect, measurements of past ecosystems. Forecasting estimates the most-likely set of predictions of ecosystem state by weighting an ensemble of model predictions by the likelihood that they match statistical estimates of empirical data (Fig. 1A).

Ecological models can be informed by paleodata via initial conditions, drivers, state variables, and parameters, each of which helps improve scientific inference (Fig 1B).

Initial conditions can have persistent impacts on ecosystem state in both models and in nature (Turney et al. 2016). Paleoecological data can thus help ensure that model runs do not entrain the consequences of flawed initialization, e.g. by initializing from well-calibrated empirical estimates of historical vegetation (Paciorek et al 2016).

Drivers of ecosystem models include reconstructions of climate and other external forces driving ecological processes. For retrospective studies, empirically estimated drivers (Tipton et al. 2016) can be assimilated into climate models using data-assimilation approaches similar to those advocated in Hakim et al. (this issue).

State variables describe the state of the ecosystem being modeled over time. Plant biomass, for instance, is a state variable whose long-term dynamics can be modeled using paleoecological observations (Fig. 1A).

Model parameters, like the growth rate (r) in a population growth model, establish links among variables. Paleodata can validate predictions of long-term ecosystem dynamics based on a particular model parameterization, or they can identify the best among a set of competing parameterizations (Fig. 1B)

In data-model assimilation, discrepancies between model predictions and paleo-observations are resolved by considering their respective uncertainties; highly certain observations will exert a correspondingly stronger constraint on state variables or parameters. Hence, accurate representation of uncertainty is paramount. In Figure 1A, data from a fossil-pollen network, calibrated against vegetation survey data, produce a statistical reconstruction of changing plant biomass, accounting for uncertainty in pollen counts, taphonomic processes, etc. (Dawson et al. 2016). The mechanistic linkages between biomass and soil carbon in an ecosystem model then allow the empirically constrained reconstruction of biomass to improve estimates of soil carbon, an unobserved state variable. By narrowing uncertainty about long-term ecosystem dynamics in the past, this approach improves the model generally and thereby reduces uncertainty in forecasts of future ecosystem dynamics.

The suite of approaches to paleodata-model fusion outlined here pose opportunities and challenges for the producers and synthesizers of open data. Win-win opportunities emerge from the iterative coupling between models and data (Dietze 2017), for example, by motivating new data campaigns to meet model demands. To capitalize on such opportunities, data stewards should work with statisticians and modelers to ensure that data are useable: For instance, when derived quantities, say temperature reconstructions, are archived, the raw data underlying them should also be archived, along with the code underlying all analyses. The rewards for this inconvenience will be new collaborations and increased predictive power!

ACKNOWLEDGEMENTS

This work is funded by the NSF Macrosystem Biology Program (EF 1241874). The ideas presented here derive from seven years of conversation among the 100+ members of the PalEON Project.

AFFILIATIONS

[1]Department of Biological Sciences, University of Notre Dame, USA

[2]sites.nd.edu/paleonproject

CONTACT

Jason S. McLachlan: jmclachl@nd.edu

REFERENCES

Dawson A et al. (2016) Quat Sci Rev 137: 156-175

Dietze MC (2017) Ecological Forecasting, Princeton U. Press, 228 pp

Turney CSM et al. (2016) PAGES Mag 24: 3

Paciorek CJ et al. (2016) PLoS ONE 11: e0150087

Tipton J et al. (2016) Environmetrics 27: 42-54

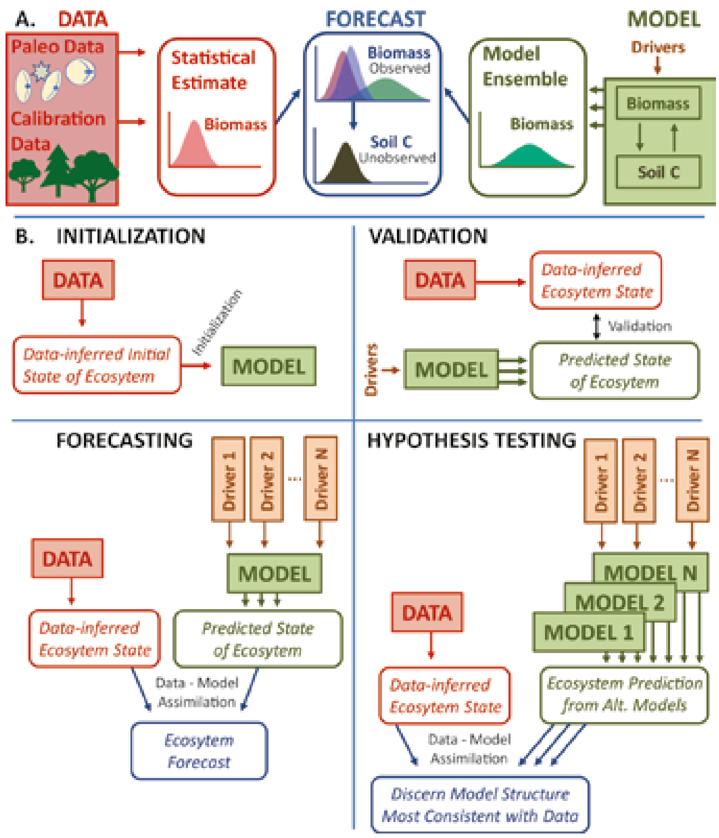


Figure 1: (A) Assimilating paleoecological estimates of aboveground plant biomass to forecast unobserved soil carbon. **(B)** Experimental designs for integrating paleoecological datasets with ecosystem models.

doi.org/10.22498/pages.26.2.77

Open-access data is uncovering past responses of biodiversity to global environmental change

Damien A. Fordham[1,2] and David Nogues-Bravo[2]

Research emerging at the frontiers between paleoecology, paleoclimatology and paleogenomics is offering exciting new prospects for unveiling the ecological and evolutionary mechanisms that have shaped past and current-day patterns of biodiversity (Nogués-Bravo et al. 2018). This frontline in paleo research is being driven by developments in high-throughput sequencing; dating and computational technologies; and open access to curated georeferenced and dated fossils, collections of genetic sequences and paleoclimate simulations. These publicly available e-resources are the result of decades of fieldwork and their combination provides innovative opportunities to use ecological and evolutionary models to connect past observed responses of biodiversity to environmental processes, particularly during the late Quaternary (from 120,000 years ago) (Barnosky et al. 2017). This integration of open-access data into biodiversity models is allowing fundamental theories in ecology and evolution to be tested and better connected to the on-ground design and implementation of effective measures to protect biodiversity (Fordham et al. 2016; Fig 1).

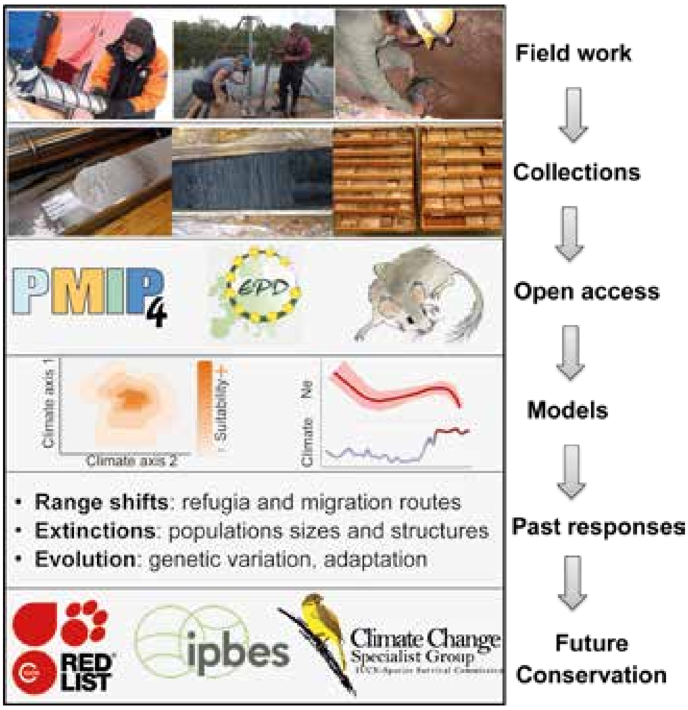


Figure 1: Information flow for using open-access data (i.e. PMIP4, European Pollen Database [EPD] and Neotoma) to uncover past ecological and evolutionary responses to environmental change, so as to improve the conservation of future biodiversity.

Until recently biodiversity modelers and other non-climate scientists have had difficulty accessing simulations of late-Quaternary climate change at the spatial and temporal scales needed to understand population-, species- and community-level responses to climatic change. Now, spatially explicit paleoclimate simulations are readily downloadable at the short-temporal scales (decades to centuries) needed to detect biotic responses to paleoclimatic change (Fordham et al. 2017).

These spatially explicit paleoclimate simulations are being used to better understand past biodiversity dynamics, and inform future conservation policies. Open-access databases, such as Neotoma (neotomadb.org) or the European Pollen Database (europeanpollendatabase.net) provide geographic localities of fossils that can be intersected with paleoclimatic simulations, then passed to statistical models to analyze changes in species climate niche properties through time, and provide spatial representations of past distributions of species, their climatic refugia and potential migration pathways (Gavin et al. 2014). These results can be used to formulate ecological hypotheses concerning changes in past population sizes and population structures, which are then testable with independent genetic sequence data from ancient DNA and modern populations – much of which is freely available via GenBank. This integrated analytical approach is providing fascinating insights into the historical biogeography of species, facilitating a better understanding of why species' population sizes and distributions change over time, and why some species survived pronounced climatic shifts during the late Quaternary and not others (Nogués-Bravo et al. 2018). Moreover, recent developments in paleogenomics are providing unparalleled opportunities to estimate not only the demographic histories of species and populations through time, but also to understand the evolutionary mechanisms that govern responses to past global environmental change (Shapiro and Hofreiter 2014).

Since genetic-sequence or genomic-level information stored in digital open-source databases often lack geographic coordinates, deep-diving algorithms and artificial intelligence are being used to georeference hundreds of thousands of genetic sequences from the peer-reviewed literature, providing new opportunities to determine the role of paleoclimatic change in structuring genetic diversity (Miraldo et al. 2016). Another major barrier is the scarcity of continuous paleoclimate simulations for the late Quaternary, which are only publicly available from the Last Glacial Maximum (LGM) to the present day (Fordham et al. 2017). Since many important biotic responses to paleoclimatic changes occurred prior to the LGM, high temporal resolution paleoclimate simulations from the last interglacial to the present day, from multiple atmosphere-ocean global circulation models, are urgently needed.

Open access to paleo resources, and their integration into macroecological models, has already played an important role in improving our understanding of how ecological and evolutionary processes regulate the severity of threats from global environmental change, providing a "real-world" foundation for better anticipating what the future may bring. The continued expansion of paleoecological information in online databases , including emerging resources, such as georeferenced ancient and modern DNA, are opening new frontiers in our understanding of past responses of biodiversity to global change.

AFFILIATIONS
[1]The Environment Institute and School of Biological Sciences, The University of Adelaide, Australia
[2]Center for Macroecology, Evolution, and Climate, National Museum of Denmark, University of Copenhagen, Denmark

CONTACT
Damien Fordham: damien.fordham@adelaide.edu.au

REFERENCES

Barnosky AD et al. (2017) Science 355: eaah4787
Fordham DA et al. (2016) Nat Clim Change 6: 909-916
Fordham DA et al. (2017) Ecography 40: 1348-1358
Gavin DG et al. (2014) New Phytol 204: 37-54
Miraldo A et al. (2016) Science 353: 1532-1535
Nogués-Bravo D et al. (2018) Trend Ecol Evol 33: 765-776
Shapiro B, Hofreiter M (2014) Science 343: 1236573

doi.org/10.22498/pages.26.2.78

EarthLife Consortium: Supporting digital paleobiology

Mark D. Uhen[1], S. Goring[2], J. Jenkins[1] and J.W. Williams[2]

The EarthLife Consortium (ELC) aims to support the accessibility, interoperability, and sustainability of paleobiological data across multiple resources. The new ELC Application Programming Interface (API) allows search and retrieval across several databases, and is readily extensible to others.

Paleobiology is a classic example of a 'long-tail' discipline, with the large majority of paleobiological data collected by individuals organized into tight guilds of specialists. Most paleobiologists have a domain of expertise centered on a particular set of organisms (or even on particular fossilized body parts within organisms), a geographic region, and a time period or timescale. For example, one paleobiologist might be an expert on leaves and seeds from the Paleogene of North America (leaving the fossil pollen and other microfossils to other specialists) (e.g. Wing et al. 2009), another might specialize in stable isotope measurements from bones and teeth (e.g. DeSantis et al. 2009), while a third might be a specialist in marine foraminifera, working with ocean-sediment cores collected from across the world (e.g. Barker et al. 2005). These scientists also pursue varied research agendas, both as individuals and research teams.

There is widespread recognition that the whole of the fossil record is greater than the sum of its parts. Many of our discipline's foundational advances – e.g. recognizing five major extinctions in Earth's history; studying speciation and extinction processes during and after extinction events (Raup and Sepkoski 1984; Sepkoski 1997; Peters and Foote 2001); demonstrating the relationship of diversity with climate and productivity variations (Marx and Uhen 2010); demonstrating that species abundances and ranges closely, but individualistically, track climate variations at timescales of 10^2 to 10^5 years during past glacial-interglacial cycles (Huntley and Birks 1983; Webb 1987) – have been made possible by the painstaking synthesis of many individual fossil occurrences into regional- to global-scale databases. Many paleobiological databases exist, some begun and maintained by individual investigators and others that have matured into open, community-curated data resources (CCDRs), with data contributed and stewarded by a broad cross section of the paleobiological community (Uhen et al. 2013; Williams et al. 2018).

The history of cyberinfrastructure development in paleobiology has been "bottom-up", with the attendant advantages and disadvantages. There has been broad and deep participation by paleobiologists in building community-supported cyberinfrastructure.

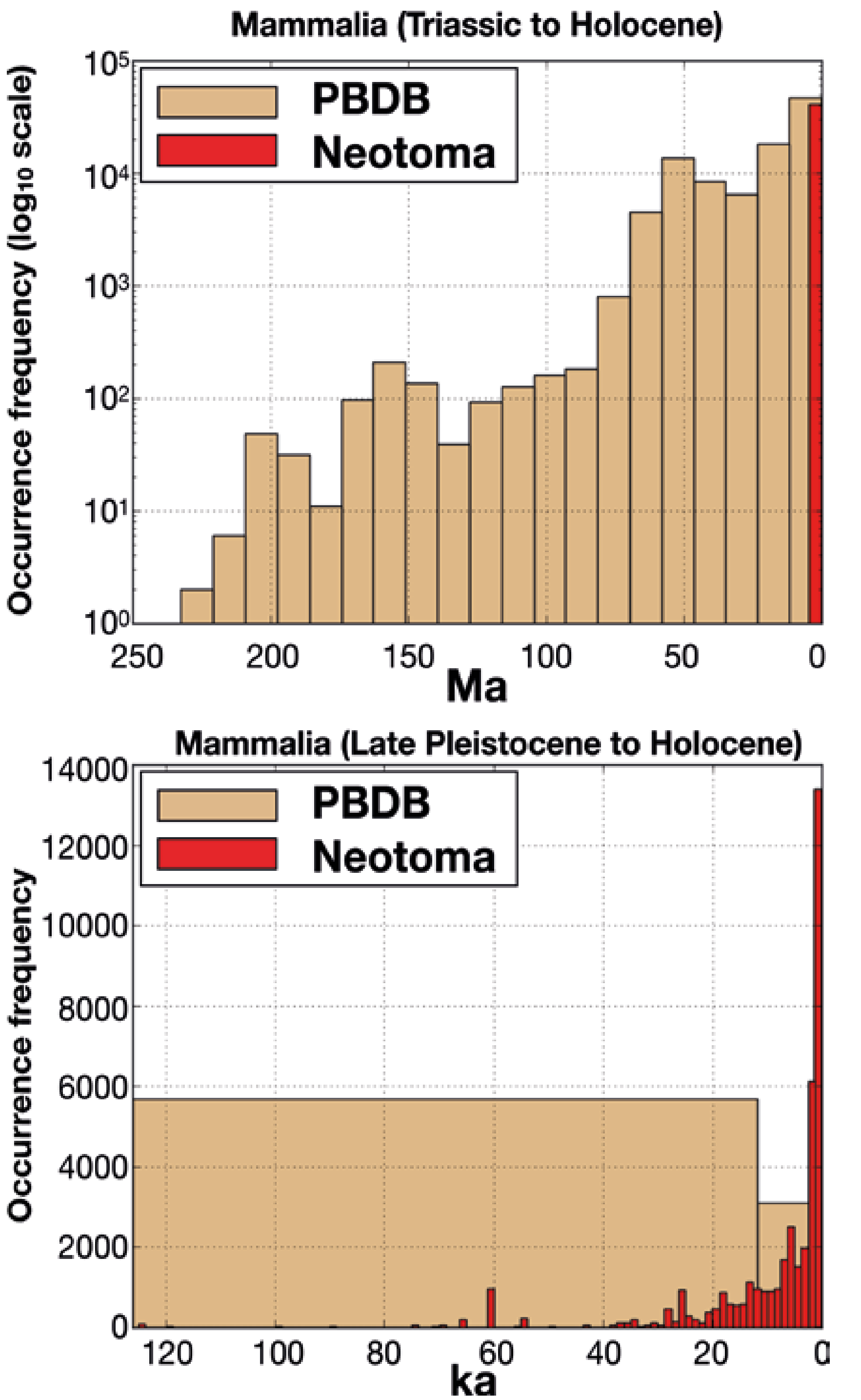


Figure 1: Comparison of the temporal distributions of occurrences of Mammalia in PBDB and Neotoma over **(A)** 250 million years (Ma); and **(B)** 120 thousand years (ka). Note that PBDB has much greater time depth, but that Neotoma has much greater time resolution over a shorter time scale.

Many hard-won lessons have been learned, and well-developed data models have been created to describe paleobiological data in geological contexts. There has also been a proliferation of many small-scale paleobiological resources, with idiosyncratic data and metadata standards and concerns about long-term sustainability of smaller resources.

ELC goals and methods

The ELC project (earthlifeconsortium.org) aims to leverage the long-tail paleobiological data to address large-scale paleobiological questions. Specifically, ELC aims to: improve and expand the interoperability of cyberinfrastructure within the paleobiosciences; promote sharing and use of paleobiological data within paleobioscience and with closely allied geoscience and bioscience disciplines; enhance the sustainability of paleobiological cyberinfrastructure by consolidating smaller resources into larger community-supported repositories; and establish a 4D framework (geography + depth + geologic time) for life and its physical environments that spans all timescales and extends back to the earliest beginnings of the fossil record.

We have advanced towards these goals with the ELC Application Programming Interface (ELC API), which returns data from Neotoma Paleoecology Database (Neotoma, neotomadb.org), which includes paleoecological and co-located paleoenvironmental data at fine temporal grains in the near past, and Paleobiology Database (PBDB, paleobiodb.org), which includes data on all fossil organisms from all of geologic time at coarser temporal grain (Fig. 1). The ELC API is fully documented on Swagger and GitHub, with the capability for extension to other related databases. ELC has already expanded to include occurrence data from the Strategic Environmental Archaeology Database (SEAD; sead.se), demonstrating the ease of database addition to the system. In doing so, we have also established a common data-interchange standard between these resources and contemporary biodiversity databases by adopting the Darwin Core format (Wieczorek et al. 2012) and further extending it for use with additional paleobiological data elements. The ELC project has also supported the incorporation of several smaller databases into Neotoma (Grimm et al. this issue).

Data from the ELC API can be returned either in comma separated value (.csv) text files, or in JSON files for further processing, display, or analysis. We have crafted eight separate endpoints for the API that return datasets based on what data the user is querying. The primary endpoints are: ***Locale***, an intersection of spatial coordinates and geologic time; ***Mobile***, which pre-packages a "light" data set on fossil occurrences for use in mobile applications such as Flyover Country (Myrbo and Loeffler, this issue); ***Occurrence***, which returns a list of occurrences of a given taxon in a specific place and time, including the subtaxa of that taxon (e.g. occurrences of fossils of all species of *Canis*, if given only the genus *Canis*); and ***Taxonomy***, which returns the metadata associated with any given taxon (e.g. ecology, time range, original author, etc.).

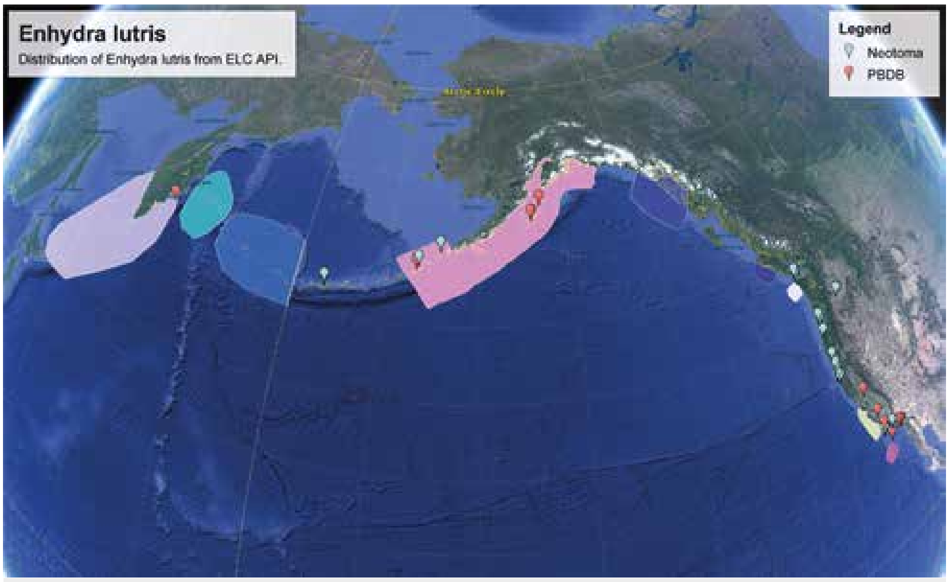


Figure 2: Distribution of the sea otter, *Enhydra lutris*, from the ELC API (All occurrences; Pleistocene-Holocene). Blue points are from Neotoma (n=82), while red points are from PBDB (n=16). Neither database has a complete picture of the distribution of fossil *E. lutris*, but the combined data from ELC more closely resembles the modern distribution of *E. lutris* (from the IUCN Red List), shown in colored polygons representing sub-populations of *E. lutris*, with fossil occurrences demonstrating presences outside the modern range. Base map from Google Earth.

Using these parameters, users can craft queries to answer many questions regarding the distribution and paleoecology of organisms through time and space, from deep geologic time scales, through glacial-interglacial time scales, into the early Anthropocene. For example, the sea otter, *Enydra lutris*, is represented in both PBDB and Neotoma, but neither has a comprehensive view of its distribution in the North Pacific fossil record. Figure 2 shows the occurrences of *Enydra lutris* derived from the ELC API which clearly shows some occurrences from both databases, yielding a much more comprehensive view of its past distribution. While the ELC API returns a limited set of data about each occurrence, end users are able to get further, richer datasets from each constituent database using provided metadata.

ELC Foundation

The Earth Life Consortium Foundation (ELC Foundation) is a non-profit organization currently in its formative stages. The ELC Foundation's missions are to provide easy, free, and global access to scientific data in paleontology, paleoenvironmental studies, and related fields and support the access, development, and sustainability of the community-curated scientific data resources that are the foundation of modern paleobiodiversity science. How best to sustain, develop, and grow these community data resources remains a persistent challenge for the paleogeosciences (Williams et al. 2017). In earlier centuries, professional societies launched peer-reviewed journals as modes of sharing data and knowledge among international networks of scientists. The time may be ripe to extend the mission of professional societies to include the support of high-quality, community-curated scientific data resources. As a starting point, the Paleontological Society and Society for Vertebrate Paleontology have contributed funds to launch the ELC Foundation.

EarthLife Consortium outlook

We welcome the participation by other paleobiological databases and societies in the ELC mission of global access to the full universe of paleobiological data. Others can also participate by joining one of the ELC participating databases, and adding data to these systems which will automatically propagate to ELC. More data in the systems will result in better-supported answers to a wider variety of questions about the history of life on Earth.

ACKNOWLEDGEMENTS

This project was supported by a grant from the US National Science Foundation, ICER 1540997.

AFFILIATIONS

[1]Department of Atmospheric, Oceanic, and Earth Sciences, George Mason University, Fairfax, USA
[2]Department of Geography, University of Wisconsin, Madison, USA

CONTACT

Mark D. Uhen: muhen@gmu.edu

REFERENCES

Barker S et al. (2005) Quat Sci Rev 24: 821-834
DeSantis LR et al. (2009) PLoS One 4: e5750
Huntley B, Birks HJB (1983) An atlas of past and present pollen maps for Europe: 0-13000 years ago. Cambridge University Press, 667 pp
Marx FG, Uhen MD (2010) Science 327: 993-996
Peters SE, Foote M (2001) Paleobiology 27: 583-601
Raup DM, Sepkoski JJ Jr. (1984) PNAS 81: 801-805
Sepkoski JJ Jr. (1997) J Paleontol 71: 533-539
Uhen MD et al. (2013) J Vert Paleontol 33: 13-28
Webb T III (1987) Vegetatio 69: 177-187
Wieczorek J et al. (2012) PLoS One 7: e29715
Williams JW et al. (2017) Authorea: 165940
Williams JW et al. (2018) Quatern Res 89: 156-177
Wing SL et al. (2009) PNAS 106: 18627-18632

doi.org/10.22498/pages.26.2.80

Advice for early-career researchers: A summary from the AMQUA/CANQUA student mixer

Vachel Carter[1,2]

Job prospects are one of the most pertinent topics for early-career researchers. At the joint AMQUA/CANQUA meeting, a panel of senior researchers shared their career advice specifically for early-career paleoscientists. Here, I'd like to share with you their advice.

Recently, I attended the joint American Quaternary Association/Canadian Quaternary Association (AMQUA/CANQUA) conference held in Ottawa, Canada, from 7-11 August 2018. The organizers arranged a 'Student Mixer and Mentoring Panel' event that was sponsored by the US National Committee (USNC) for Quaternary Research. Roughly 40-50 early-career researchers (ECRs), including Masters and PhD students, and postdocs attended. Seven panelists (Fig. 1) gave splendid advice for ECRs.

*From **Christopher Hill**, Professor at Boise State University in the Departments of Geosciences and Anthropology, on assignment as a Program Director at the US National Science Foundation.*

"As part of your graduate studies, please think about the value of non-academic internships. Non-academic internships provide an opportunity to have experiences and get skills that can prepare you for a great career in science. Examples of these types of internships may be opportunities in government agencies (at local, state, or federal levels) or national laboratories, non-profit organizations, or for-profit business (businesses of all kinds). Think about working with your graduate advisor or mentor and consider the types of internships that might be integrated into your graduate research. An ideal situation might be an internship opportunity that provides you with experiences and contacts not available at your university and is also a project that can be integrated into your graduate studies and research. Internships are an amazing opportunity to gain additional expertise and also learn about great careers where you can apply your training in science."

*From **Tony Layzell**, Quaternary geologist at the Kansas Geological Survey.*

"My success in graduate school (both Masters and PhD) was largely driven by the advisors I was fortunate enough to work with. I would suggest that graduate students find an advisor that is the best fit for them (some students prefer minimal supervision, some prefer more hands-on support, etc.). The best fit is not always the most reputable scientist in your field.

If early-career scientists are wanting to go into academia then publishing is a must. Every project I was involved with, including those unrelated to my dissertation, was published in a peer-reviewed journal of some kind. Whether we like it or not, journal articles are the coin of the realm.

Competition for academic positions is fierce and most will have to do a postdoc before a tenure-track or equivalent position. While doing a postdoc make sure that you have the time and flexibility to publish everything from your dissertation. Finding a way to stand out from the crowd is also vital. My advice is to always do what interests you the most, what you are most passionate about, but try to make it relate to your desired career in some form. This is particularly important if you have to take an alternative career path or have to take multiple postdoc jobs while seeking an academic position. Stay relevant by publishing or by teaching in some capacity if you ultimately want an academic job.

Networking is also key. Try to go to as many conferences as you are able. Seek out other scientists in your field and ask them about their research. You will need these people to write you letters of recommendation when you go up for tenure. The days of the independent scientist are waning – collaboration, communication, and collegiality are vital in science today."

*From **Rolfe Mandel**, Distinguished Professor of Anthropology at the University of Kansas, Director of the Kansas Geological Survey, and Chair of the US National Committee (USNC) for Quaternary Research-INQUA.*

"It's important to be really good at something, but not to overspecialize. Diversify your skill set to include things like GIS and remote sensing. Being adaptive allows you to be a better team member and makes you more marketable in the workforce. There are other opportunities outside of academia, namely government agencies, such as the National Park Service, US Geological Survey, and state geological surveys, and private consulting firms; they are in need of good Quaternary researchers. Leaving academia is not a negative thing; you won't disgrace your mentors if you leave academia. It's not a problem to step out for a bit, but if you do take a break, you should stay in the mix by continuing to read articles in your field, publishing your research, attending conferences, and maintaining contact with your collaborators."

*From **Kendra McLauchlan**, Full Professor of Geography at Kansas State University, on assignment as a Program Officer at the US National Science Foundation.*

"There are a couple of inherent advantages and disadvantages when you are a Quaternary scientist, or PAGES-type scientist.

Advantages: (1) By training in the paleosciences, you are quite interdisciplinary compared with most PhDs. Even a specialized PhD requires thinking about paleoecology, paleoclimatology, geochronology, Earth surface processes, and the human dimension. Just look at the diversity within and among PAGES working groups. It is a huge advantage to be trained in such an interdisciplinary science. (2) Your skills are probably quite sharp compared with most PhDs. Again, to complete a PhD you have to be proficient in several techniques, and cognizant of many others. For example, the multiproxy work that so many of us do demonstrates really advanced analytical skills. Our science still has a good component of field work, as well as lab work, and statistical and data management work. These are extremely desirable skills in many job markets.

Disadvantages: (1) You will likely never see a job ad with a title that you think exactly describes your training. You will have to be creative about connecting job ads, especially the title of job ads, with your own training. You will likely have to search several job-posting places, focused on different intellectual and geographic areas, to find a good fit. But be creative about this! (2) Landing a non-academic job can be difficult. This is true across many scientific disciplines (there is a lot of primary literature on this), due to well-documented phenomena such as the location of training in universities, the perceived prestige of academic jobs, the lack of diversity in advanced degree types, and the

Figure 1: Several of the panelists (from left to right, Stephen Wolfe, James Teller, Kendra McLauchlan, Tony Layzell, Chris Hill, and Cathy Whitlock; not pictured, Rolfe Mandel) at the Student Mixer and Mentoring Panel at the AMQUA/CANQUA conference.

lack of ability of traditional mentoring to connect with non-academic jobs. This is slowly starting to change, but you will likely have to hasten the pace if you are interested in a non-academic job. Again, the good news is that paleoscientists are well-trained for non-academic as well as academic jobs, it will just take some work to fit the match.

Now, just a few pieces of advice. Think deeply about what your goals are, as early on as possible. Once you identify those, work toward them. Everything you do in your career – every paper, every presentation, every professional conversation – should be working toward that goal. Ask your mentors and colleagues to help get you there, and seek out opportunities for training that will help you reach your goals. Finally, a career is not a race. Be open to different paces, and different paths, and even once you cross a type of finish line by landing a job, you'll need to keep growing, and creating, and pushing ahead!"

*From **James Teller**, Professor Emeritus in the Department of Geological Sciences, University of Manitoba, Winnipeg.*

"I had outstanding advisors, who understood that I needed guidance, but who respected me (after getting to know me) and realized that I worked well with a minimum of supervision. I love what I do, and never am shy to tell others that I do. I think all geoscientists should communicate with the public, media, policy makers, etc, telling them in an enthusiastic way and in understandable language about the profession you love.

My advice: Work hard and play hard. Accept all challenges and take on even more than you think you can easily manage, but with a realistic understanding of yourself and the commitments. Be vigilant about what's going on in your research area, department, and company, and never miss an opportunity. Your success in being hired and in your career will depend on your geoscience skills and on your people and communication skills – do not underestimate the latter. Secure a permanent job as soon as you can after graduating; do not take a lot of temporary jobs or postdocs, as it will negatively impact on securing permanent employment. Being able to write clearly and professionally is extremely important. Stand out (in a good way), so your resume (and your interview) puts you into the pile of "Finalists", rather than keep you as part of the herd of other applicants. Do your best, always. Cultivate opportunity. Enjoy being with your colleagues, be cheerful, be positive. As someone famous once said: 'The harder I work, the luckier I seem to be.'"

*From **Cathy Whitlock**, Professor of Earth Sciences at Montana State University and Fellow of the Montana Institute on Ecosystems.*

"If you're applying for a faculty position at a research and teaching institution, here are a few suggestions that might elevate your chances for an interview. Your cover letter is the first piece of information for the search committee, so it should be short, well written, and clearly articulate your research and teaching interests in a way that sound like a good fit for the position. It is smart to express some enthusiasm for the particular position and institution and show some appreciation for the guidance of your mentors. The search committee usually looks at your list of publications next. Even if you are in graduate school, there should be evidence that you are publishing in impactful journals and have the ability to work collaboratively. Evidence of external funding, even small research and travel grants, is important as well. The committee also looks at your teaching and public communication experience to see if you have the potential to be an inspiring educator. Beyond these items, describing your participation in broader engagement activities can help elevate your application. Working with teachers and students, communicating science to the public, and community or disciplinary service all catch notice. It will likely take more than one try to land an academic position, but just remember that each application and interview prepares you to do better on the next, so stay positive!"

Final thoughts

As an ECR, I found the information above very insightful. The most relevant bit of information for me was the advice from Kendra McLauchlan, "...seek out opportunities for training that will help you reach your goals." Personally, I have taken the opportunity to work with new colleagues post my PhD who have provided new skill set training, all in the hope that these new skill sets will help me reach my end goal. I hope you found the panelists advice as insightful as I did.

AFFILIATIONS
[1]Department of Botany, Charles University, Czech Republic
[2]PAGES ECN Steering Committee Member

CONTACT
Vachel Carter: vachel.carter@gmail.com

doi.org/10.22498/pages.26.2.82

Quaternary studies in Africa: a perspective from early-career researchers

AFQUA-PAGES ECN meeting participants*

Due to its large longitudinal, latitudinal and also altitudinal extent, the African continent crosses different climate zones and therefore hosts extremely diverse environments, such as tropical glaciers in the equatorial region, hyper-arid deserts in its subtropical latitudes, and Mediterranean ecosystems in its northern and southern fringes. Quaternary scientists have long been attracted to Africa not only because of the natural diversity of its environments, flora, and fauna, but also because Africa is the epicenter of human origins and evolution. Quaternary research in Africa therefore offers a unique opportunity to develop and test our understanding of key components of the Earth system at the intersection of past climate change, environments, and humans.

Unfortunately, despite the great interest in African paleoscience research at a global level, it remains very challenging for African early-career researchers (ECRs) to develop a scientific career in studying their own continent. However, things are positively evolving and more possibilities are now becoming available for African ECRs to conduct Quaternary science with, for instance, the new accelerator mass spectroscopy (AMS) facility at iThemba LABS[1] in Johannesburg (South Africa) that invites ECRs to receive training in radiocarbon dating and other analyses, or the Pan African Institutes of Science and Technology in Arusha, Tanzania[2], and Abuja, Nigeria[3], where research-intensive postgraduate and postdoc studies are carried out.

Opportunities to meet the international community are also more abundant, as most international congresses and workshops now offer travel grants (evaluated on a competitive basis at international levels) that are specifically dedicated to ECRs and researchers from developing countries. Participation in meetings like these is not only essential for keeping abreast with the most recent scientific developments, but is also critical for networking purposes. Recently, the African Quaternary Association[4] (AFQUA) community started to develop a series of congresses and training workshops based in Africa to bridge the gap between the regional African communities (such as SASQUA, EAQUA and WAQUA) and the more global communities. During its recent edition held in Nairobi, Kenya, in July 2018 (Chase, this issue), the presentations from the speakers and panel discussions brought to the forefront ideas and concerns, and highlighted some areas for future collaborative work. This provides an optimistic view for ECRs as future research leaders on some of the identified research niches.

However, even though opportunities exist, it is often challenging for African-based ECRs to seize them due to barriers in accessing journals, and limited opportunities to engage with scientists and non-scientists. These barriers can be partly alleviated by using social media platforms, such as ResearchGate and Twitter. Although not without their own challenges, these online media platforms represent a great way to network and share your research with local and international stakeholders. Exploring these different facets of communication will undoubtedly widen visibility within the scientific community. Joining international ECR communities can also alleviate the isolation ECRs can experience. The PAGES Early-Career Network[5] (PAGES ECN) offers an online platform to encourage discussions, exchange knowledge, and provide specific training. While still in its early phase of development, all the products that will be generated by that network, including, for instance, tips on how to improve your CV, help with applying for travel grants, and methods to communicate your research most effectively, will be accessible online.

We invite all African ECRs to get involved with this network and help build the African ECR community (contact: PAGES.ECN@gmail.com).

There is currently a pressing need to understand past environmental change in Africa, which will inform sustainable development, and offer real solutions, innovations and technology. Although infrastructure challenges exist, initiatives like AFQUA and the PAGES ECN represent unique opportunities that ECRs should seize to develop their position in their own community, to establish and extend long-term collaborations on African-based research, and eventually to become key players in international dialogues. Now is the time for African researchers to become ambassadors of their continent.

AFFILIATIONS

*Listed alphabetically: M. Chevalier (University of Lausanne, Switzerland); D. Colarossi (Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany); V.J. Hare (University of Cape Town, South Africa); F. Henselowsky (Heidelberg University, Germany); C. Martin-Jones (Cambridge University, UK); H. Moorhouse (Lancaster University, UK); C. Omuombo (University of Nairobi, Kenya); L.J. Quick (Nelson Mandela University, Port Elizabeth, South Africa); E. Razanatsoa (University of Cape Town, South Africa); R. Ssemulende (Makerere University, Kampala, Uganda)

CONTACT

Manuel Chevalier: manuel.chevalier@unil.ch

LINKS

[1]tlabs.ac.za
[2]nm-aist.ac.tz
[3]aust.edu.ng
[4]afquacongress.wixsite.com/afqua2018
[5]pastglobalchanges.org/ecn

Figure 1: Word cloud summarizing our lively discussions at AFQUA.

doi.org/10.22498/pages.26.2.83

The African Quaternary: environments, ecology and humans

Brian M. Chase

Nairobi, Kenya, 14-22 July 2018

With support from PAGES and INQUA, the 2nd AFQUA Conference (The African Quaternary: environments, ecology and humans) took place in Nairobi at the National Museum of Kenya. AFQUA was conceived to bridge the existing gap between large international meetings (>500 delegates) and regional African conferences (~50 delegates) and create a forum to share results and foster communication and collaboration at both regional, continental, and international scales. Further, AFQUA recognizes that the global distribution of research and educational funds favors developed-world researchers to become the recognized leaders in African Quaternary studies, but often their interactions with the local scientific communities are limited to their direct collaborators. AFQUA therefore seeks to (1) create a more fully integrated African research network and (2) provide opportunities for African researchers to develop, access, and share capacity that will allow for their participation in research projects at the highest level. AFQUA brings developed- and developing-world researchers together, but it also goes beyond the standard structure of most conferences, with equal time being dedicated to a series of focus groups and training workshops. These include thematic discussions on cutting-edge research topics as well as workshops that introduce and train researchers in the skills they need to develop and communicate their science in the modern research environment.

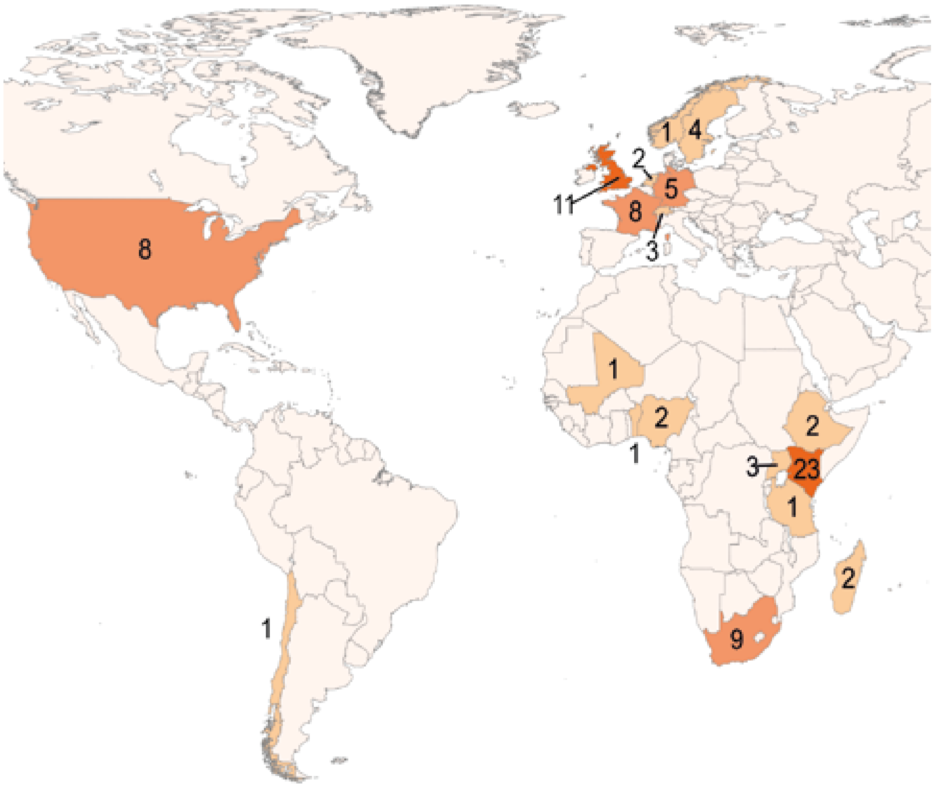


Figure 1: Distribution of delegates attending the 2018 AFQUA conference in Nairobi, Kenya.

In Nairobi, 84 researchers from 21 countries came together to share their work from the Pliocene to the projected future, spanning diverse subjects from human evolution, to climate change and vegetation dynamics, fire ecology, risk management, and the history and impacts of humans on their environments. Linking key conference themes with plenaries and sessions, Andy Cohen kicked off the conference showcasing the role that continental drilling of the large East African lake basins has had on our understanding of Quaternary environments in Africa. Subsequent plenaries included Sharon Nicholson discussing African climate change and appropriate ways to use modern systems to understand past climates; Boris Vannière and Daniele Colombaroli highlighting the past and future work of the PAGES Global Paleofire Working Group and the potential for fire ecology research in Africa; David Nash describing how historical documentary sources can be used to study climate-change history and impacts; and Daniel Olago discussing what climate change means in the African context, and how land and water resources may be impacted.

These presentations served as keystones for daily themes, which included papers on regional phenomena from across the continent, as well as papers presented in focus sessions on (1) Quantitative paleoclimatology, modeling and data-model comparisons; (2) The state-of-the-art and perspectives about fire history, fire ecology and fire-vegetation-climate interactions across tropical biomes; (3) The environmental context for hominin evolution and dispersal; (4) African archaeological landscapes; (5) Applying the Quaternary: the role of the past in supporting the future; (6) Dating and correlation of African archives of environmental change and archaeology; and (7) African paleoecology and archaeology perspectives on land-use transformation: Africa LandUse6k.

The PAGES Early-Career Network organized a splinter meeting that gathered 17 participants. Many African early-career researchers discovered the existence of this new network, recognized its potential and expressed a high level of interest to become active members in the future.

Following the five days of presentations, three days of workshops were held to provide training and foster collaboration on research projects. Participants had the opportunity to engage in international drilling programs, improve their knowledge of using lake sediments, animal remains and charcoal to understand paleosystems, learn how to integrate GIS methods in their work, and how to best apply radiocarbon techniques to create reliable chronologies (including the award of five free radiocarbon ages from the ^{14}Chrono Centre to one lucky participant!).

Building on the success of this meeting and the inaugural AFQUA conference held in Cape Town in 2015, we intend to hold the 3rd AFQUA Conference in 2021. If you would like to be added to the AFQUA mailing list, to stay informed about future developments and meetings, please write to us at afqua.congress@gmail.com.

AFFILIATIONS
[1]Institut des Sciences de l'Evolution-Montpellier, Université de Montpellier, France

CONTACT
Brian M. Chase: Brian.Chase@um2.fr

doi.org/10.22498/pages.26.2.84

Continental archives of Past Global Changes: from Quaternary to Anthropocene

Blas Valero-Garcés[1], A. Myrbo[2], A. Noren[2] and C. Jennings[2,3]

Minneapolis, USA, 24 May 2018

The Department of Earth Sciences, the Continental Scientific Drilling Coordination Office (CSDCO) and LacCore, University of Minnesota (UMN), Minneapolis, USA, hosted a Symposium after the PAGES Scientific Steering Committee (SSC) meeting. As in previous years, the SSC meeting provided an opportunity for local scientists to showcase the research related to past global changes and to interact with PAGES scientists. Minnesota has a long history of paleoclimate and paleoenvironmental research and, in particular, the Limnological Research Center has been a hub for past global changes research and the origin of LacCore and CSDCO (csdco.umn.edu). Following the steps of Herb Wright and Kerry Kelts, both facilities have been instrumental for a large number of lake-core projects in the last decades. The Symposium included 14 talks addressing local to global topics, and discussed several aspects of Quaternary and Anthropocene climate changes and ecosystem responses.

At a local scale, Amy Myrbo (UMN) showed the multiple negative consequences of sulfate pollution of freshwater shallow lakes, and how to recognize sulfate loading in the (neo)paleorecord. Randy Calcote (UMN) described several examples of how vegetation patchiness is a major control on ecosystem evolution during the Holocene and has to be considered in our future predictions. Andy Breckenridge (University of Wisconsin-Superior) illustrated how to build a 5,000-year varve chronology from glacial lakes Norwood and Agassiz. And Carrie Jennings (Freshwater Society and UMN) reviewed the Glacial Geology of the Minnesota and Mississippi Rivers (Fig. 1).

At a regional scale, Charles Umbanhowar (St. Olaf College) illustrated the differences between aquatic and terrestrial subarctic ecosystems (Manitoba, Canada) and how they tell mostly the same story at multiple sites. Kelly MacGregor (Macalester College) showed the Holocene landscape dynamics in the Paternoster lakes in Glacier National Park and the later human impacts. Kevin Theissen (St. Thomas University) presented several cases of carbonate lake sequences from Nevada and their isotopic signatures during the Holocene.

At a global scale, Larry Edwards (UMN) summarized the Chinese and Amazon cave sequences and how they pace the monsoon changes and show a clear anti-phase behavior. Larry showed us the newest U/Th and ^{14}C speleothem chronology from several Chinese caves that will help to improve the ^{14}C calibration curve. Emi Ito (UMN) presented the case of Pliocene lakes in Western North America and a possible new International Continental Scientific Drilling Program (ICDP) project. In terms of geobiochemical cycles, Jeff Havig (UMN) reviewed several examples of how to link geochemistry and geobiology in extreme lake and spring environments as found in Yellowstone, and Adam Heathcote (St. Croix Watershed Research Station) showed a review of global trends in carbon burial in lakes. Paul Glaser (UMN) summarized several decades of studies on peatlands and the changing relationship between sea level and peat growth depending on the tectonic settings. To end proceedings, Shane Loeffler (UMN) gave us an overview of Flyover Country®, a US National Science Foundation-funded offline mobile app for geoscience outreach and data discovery (see Myrbo et al., this issue).

Symposium participants toured the CSDCO and LacCore Facility, and Anders Noren (UMN) gave a summary of the support and infrastructure provided to the research community, and examples of recent high-profile projects. These facilities offer expertise in project development, planning, and management; rental equipment for field operations; core lab for processing, scanning, and subsampling; repository for cores, data, publications, and reference collections; support for outreach, diversity, and education activities; community software; and community coordination for long-range science planning. Each year, they provide support to an average of 1,400 scientists from 500 institutions worldwide, and the repository now includes more than 8,000 sites. With roots in the paleolimnology community, these facilities now offer services to continental drilling and coring projects with any scientific focus.

The following day, Carrie Jennings (Freshwater Society and UMN) led a field trip to show the post-glacial evolution of the Minnesota and Mississippi rivers' valleys. The Minnesota River valley was carved by the draining of glacial Lake Agassiz and it has a wide and active floodplain with dynamic lakes, levees and channels. The Upper Mississippi was one of the first locations where geologists demonstrated the long periods needed for geological processes to operate, as in 1876, when N.H. Winchell calculated that the retreat rate for St. Anthony Falls in Minneapolis was about a meter per year.

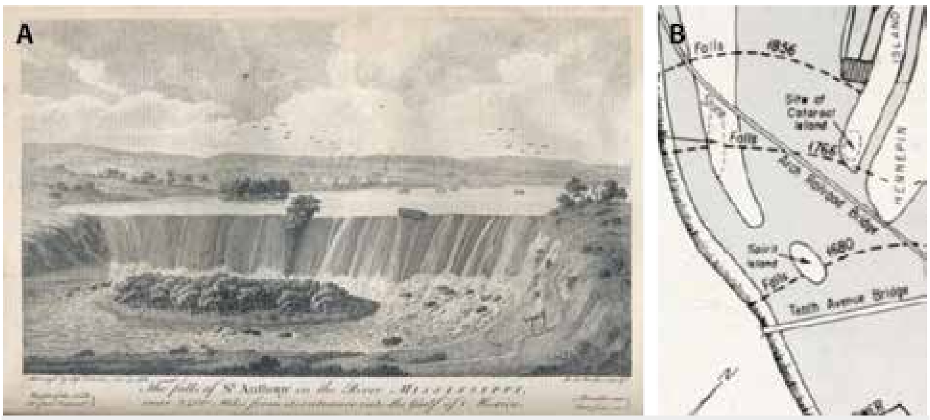


Figure 1: Left. The Falls of St. Anthony in a 1766 Sketch by Jonathon Carver. **Right**. The retreat of St. Anthony Falls from 1680 to 1871 after Sardeson, 1916. From Wright (1972).

AFFILIATIONS
[1]Pyrenean Institute of Ecology, Spanish National Research Council, Zaragoza, Spain
[2]LacCore Facility and Continental Scientific Drilling Coordination Office (CSDCO), University of Minnesota, Minneapolis, USA
[3]Freshwater Society, St. Paul, USA

CONTACT
Blas Valero-Garcés: blas@ipe.csic.es

REFERENCE

Wright HE Jr (1972) In: Sims PK, Morey GB (Eds) Geology of Minnesota: A centennial volume. Minnesota Geological Survey, 515-547

doi.org/10.22498/pages.26.2.85

An integrated proxy and simulation data initiative for the Holocene and the last deglaciation

Heather Andres[1], O. Bothe[2], K. Rehfeld[3], S. Wagner[2], N. Weitzel[4] and E. Zorita[2]

Hamburg, Germany, 16-18 April 2018

Comparing climate proxy and simulation data is fraught with challenges: age and calibration uncertainties in climate proxies, missing or incomplete processes and uncertain boundary conditions for climate models, and differences between gridded and site data are just a few examples. For the climate of the Common Era, multiple initiatives have already addressed these issues (e.g. the PAGES 2k Network regional working groups). On transient timescales beyond the late Holocene, there have been only a few integrated activities. Comparisons on these longer time scales involve large-scale changes in climate states without an equivalent during the Holocene. As such, they require methods that address both the amplitude and timing of background climate changes and account for additional processes. For example, comprehensive Earth System Models need to include changes in ice sheets and related ocean circulation changes during deglaciation. Likewise, proxy data for this period, such as lake or marine sediments, are generally less well replicated than their late Holocene counterparts (e.g. tree rings and historical documents), resulting in more uncertain climate signals (Laepple et al. 2017).

To address strategies for data-model comparisons on late Pleistocene and Holocene time scales, 30 participants, including global and regional climate modelers, statisticians and proxy experts, gathered in Hamburg for a three-day workshop. The meeting was co-sponsored by the German climate modeling initiative PalMod. The workshop started with overview talks, which provided a solid base for breakout groups. These groups covered three main categories, addressing (i) conceptual aspects of data-model comparisons, (ii) inferring Holocene, and (iii) deglacial climate changes by combining proxies and models.

The methodological breakout group began to develop a framework based on the comparison of probability distributions of both proxy and simulated data, which accounts for quantifiable uncertainties. One of their main objectives was to develop summary metrics that assess the mismatch of reconstructed and simulated climate information and are robust with respect to uncertainties. The Holocene breakout group planned an integrated analysis of different types of proxy data and model simulations for not only hemispheric and global means, but also regions, such as Europe and the North Atlantic. This group focused on apparent model-data mismatches beyond the so-called "Holocene Conundrum" (Liu et al. 2014), like the influence of large scale atmospheric and oceanic processes on regional Holocene climate variability, and discrepancies between the seasonality and regional characteristics of trends in simulations and pollen-based reconstructions. The deglaciation breakout group developed a procedure to evaluate the representation of large-amplitude events in proxies and simulations. This method compares the spatio-temporal structure of an event relative to its onset, and thus allows comparisons of manifestations of internal variability as well as events that occur at different times in the simulations and proxy records. The approach depends on metrics that quantify mismatches in time and space. This breakout group carried out initial testing of their methodology with existing datasets (Fig. 1) and will prepare, with the methodology group, a document with guidelines and a specific description of the algorithm (i.e. a cookbook) to enable others to implement the same procedure.

In addition to the methodological cookbook, workshop attendees planned future activities. These include writing an overview manuscript on the questions formulated in the Holocene breakout group. Additional follow-up plans include an interactive toolbox to compare model simulations with proxies using a hierarchy of metrics with varying complexity.

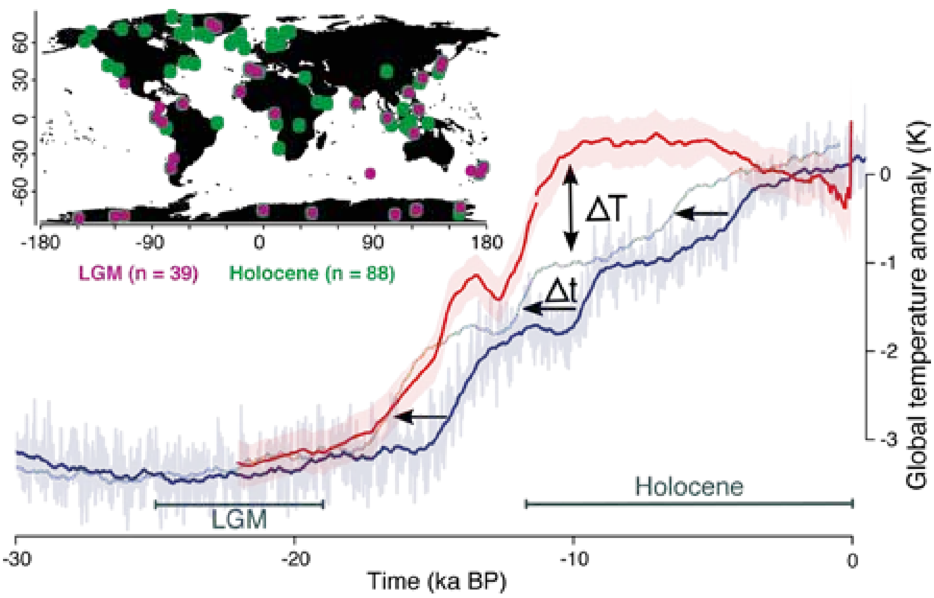


Figure 1: Comparison of an abrupt warming in global temperature during the deglaciation between climate model (blue line; Smith and Gregory 2012) and paleoclimate reconstruction data (red lines; Marcott et al. 2013; Shakun et al. 2012). The simulated time series is shifted in time to optimally match the reconstructed series for events of interest (grey line shows shifted series, horizontal arrows mark the time change Δt). The discrepancy, ΔT, is evaluated at every available spatial location (see inset; dots show data availability during the Holocene, blue, and the LGM, red; Rehfeld et al. 2018).

AFFILIATIONS
[1]Department of Physics and Physical Oceanography, Memorial University of Newfoundland, St. John's, Canada
[2]Institute of Coastal Research, Helmholtz-Zentrum Geesthacht, Germany
[3]Institut für Umweltphysik, Universität Heidelberg, Germany
[4]Institut für Geowissenschaften und Meteorologie, Universität Bonn, Germany

CONTACT
Oliver Bothe: oliver.bothe@hzg.de

REFERENCES

Laepple T et al. (2017) PAGES Mag 25: 140-141
Liu Z et al. (2014) PNAS 111: E3501-E3505
Marcott SA et al. (2013) Science 339: 1198-1201
Rehfeld K et al. (2018) Nature 554: 356-359
Shakun J et al. (2012) Nature 484: 49-54
Smith RS, Gregory J (2012) Clim Dyn 38: 1545-1559

doi.org/10.22498/pages.26.2.86

Identifying data gaps and potential synergies in forest dynamics research

Jesse Morris[1], J. Clear[2], R. Chiverell[3], R.J. DeRose[4] and I. Drobychev[5]

Liverpool, UK, 21-23 March 2017

Disturbance events, such as wildfire and bark beetle outbreaks, are natural processes that promote forest regeneration and succession. As climate change propels ecosystems and disturbance regimes along new trajectories, retrospective ecological records, like those derived from tree rings and lake sediments, become important in understanding and anticipating future environmental change. Paleoenvironmental records are essential to inform conservation and management efforts to characterize baseline ecosystem services and recovery rates from disturbances.

In March 2017, PAGES' Forest Dynamics working group assembled a team of 29 scientists, including 14 early-career researchers, from 13 nations to explore new approaches to reconstructing forest disturbances. The workshop occurred at Liverpool Hope University (UK). The first workshop day was in conference format where participants gave research talks in themed oral presentation sessions. The second day was organized around group discussions that explored four broad themes:

(1) Improving reconstructions of non-fire disturbances

While advances have occurred in recent decades in the reconstruction of paleofires from tree-ring and lake-sediment records, non-fire disturbance events, such as insect and pathogen outbreaks, are largely unaccounted in paleoenvironmental studies. Non-fire disturbances are important drivers of forest succession, yet they are challenging to reconstruct because indicators that conclusively diagnose the disturbance are infrequently recovered, as in the case with bark beetles (Morris et al. 2015). However, increasing the volume of sediment sub-samples (Whitehouse et al. 2010) may enhance macrofossil recovery, including insect remains, which would improve their utility as a proxy for outbreak events. Recent work by Stivrins et al. (2017) provides an encouraging example in reconstructing non-fire disturbances, where the authors identified and counted non-pollen palynomorphs from a fungal pathogen (*Phytophthora*) to explain the decline of alder (*Alnus* spp.) in Europe during the medieval period.

(2) Reconstructing the severity of forest disturbances

Discussions centered on how to better integrate lake-sediment proxies with tree-ring and forest-inventory data. In some instances, forest-inventory records have been combined with pollen records to produce estimates of past forest cover as aboveground biomass (Seppä et al. 2009). A key advance in assessing past disturbances centers on the quantifying tree mortality resulting from the event (i.e. severity). Biomass may aid in quantifying the severity of plant mortality as a result of a disturbance. Severity reconstructions could be enhanced by integrating mortality data from individual trees (tree rings), stands (forest hollows) to landscapes (lakes).

(3) Opportunities for database integration

One practical approach to determining potential sites to link records of forest dynamics would be to integrate existing databases. For example, integrating databases through geolocation would simplify the identification of study areas for synergistic work. Several databases were discussed, including Neotoma (neotomadb.org), the NOAA National Centers for Environmental Information (ncdc.noaa.gov/paleo), and forest inventory databases, such as the USDA Forest Inventory and Analysis (fia.fs.fed.us/tools-data). Database integration presents a key opportunity to advance reconstructions of past forest conditions, especially in leveraging quantitative approaches to reconstruct land cover (Fig. 1).

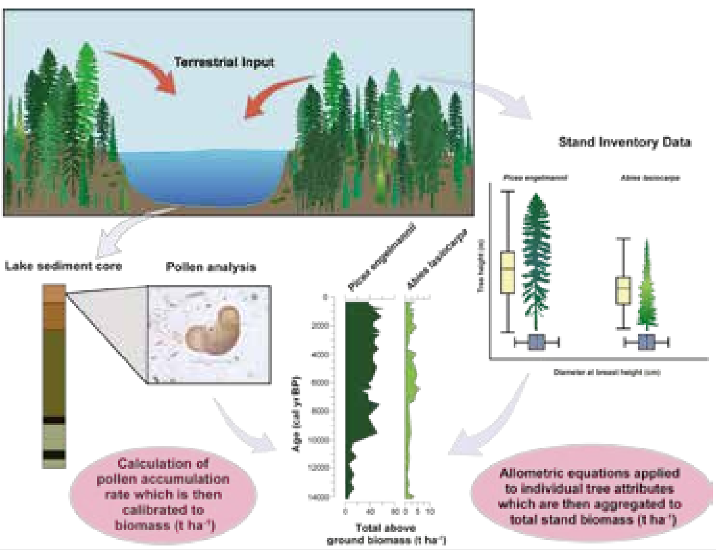


Figure 1: Conceptualization of transforming pollen data to biomass (t/ha) using stand demographic data.

(4) Key methods for data integration

The final discussion centered on new methods for integrating lake-sediment proxy data across scales. Generally, data integration across scales can be challenging even when using a single proxy (e.g. pollen). Examples from the first day's presentations were highlighted. Quantitative pollen techniques, such as the Landscape Reconstruction Algorithm (LRA) (Sugita 2007), help to compare vegetation reconstructions across spatial and temporal scales, though with noted issues (Trondman et al. 2016). Recently, Carter et al. (2018) used the LRA approach to integrate pollen profiles from 10 peat bogs with four lake records to reconstruct total land-cover abundance across local- to regional-scales. Their approach provides information that is relevant for conversation and ecosystem management strategies in Central Europe.

Future activities

The next Forst Dynamics' activity will be to write a synthesis paper with the working title "Pairing dendrochronological and paleoecological approaches to reconstructing past disturbances events."

ACKNOWLEDGEMENTS

This workshop was supported by PAGES and the Liverpool Hope University.

AFFILIATIONS

[1]Department of Geography, University of Utah, Salt Lake City, USA
[2]Department of Geography, Liverpool Hope University, UK
[3]Department of Geography, University of Liverpool, UK
[4]Forest Inventory and Analysis, USDA Forest Service, Ogden, USA
[5]Swedish University of Agricultural Sciences, Alnarp, Sweden & University of Quebec at Abitibi-Temiscamingue, Canada

CONTACT

Jesse Morris: jesse.morris@utah.edu

REFERENCES

Carter VA et al. (2018) Fron Earth Sci 8: 2268
Morris JL et al. (2015) Quat Int 387: 72-86
Seppä H et al. (2009) Holocene 19: 209-220
Stivrins N et al. (2017) J Quat Sci 32: 903-907
Sugita S (2007) Holocene 17: 229-257
Trondman A-K et al. (2016) Veg Hist Archaeobot 25: 131-151
Whitehouse NJ et al. (2010) Quat Sci Rev 29: 539-553

doi.org/10.22498/pages.26.2.87

Assessing the links between resilience, disturbance and functional traits in paleoecological datasets

Rebecca Hamilton[1], T. Brussel[2], Q. Asena[3], R. Bruél[4], K. Marcisz[5], M. Słowiński[6] and J. Morris[7]

2nd EcoRe3 workshop, Salt Lake City, USA, 8-10 May 2018

Functional traits are integral to understanding the patterns and drivers of ecological resilience. Response traits determine whether a system exhibits properties resulting in higher resistance or faster recovery rates, while effects traits are important for identifying links between environmental changes and ecosystem functioning. Although paleoecological research methods have been identified as an important tool for understanding patterns of ecological resilience across different regions, paleoecologists have only recently begun to think about how functional ecological approaches can be used in the context of sub-fossil assemblages from sediments. Thus, the second EcoRe3 workshop aimed to (i) explore developments in modeling ecosystem resilience from the paleorecord, and (ii) investigate the development of new tools to link paleoecological datasets with functional trait databases derived from neo-ecological research.

These two themes were investigated over two days at the Natural History Museum of Utah in the form of presentations and semi-structured discussion groups. Twenty-two participants from nine countries attended the meeting, with funding provided by PAGES and the Research Council of Norway. Research covered in the meeting was diverse with respect to the ecosystems studied, proxies (pollen, Cladocera, testate amoebae, diatoms) and timescales interrogated, and modeling methodologies applied. The diversity of approaches highlighted advancements the working group had made in measuring long-term patterns of ecosystem resilience since the last meeting.

Presentations themed around modeling resilience from long-term ecological data showcased progress in the development of novel techniques for measuring critical transitions, and presented particular records that serve as interesting case studies for the exploration of aspects of resilience. They highlighted the value of high-resolution data (taxonomic and temporal) that capture ecological regime shifts and/or clearly identifiable drivers of change for examination of ecological thresholds and response rates.

Further progression in using paleoecology for measuring resilience requires improvements in the capacity of associated research to capture individual and collective functional traits that can be related to the resistance or recovery rate of an ecosystem following disturbance. This is particularly apparent where no critical thresholds are breached, when there is a lead- or lag-proxy response to drivers of change, or when multiple stressors operating at different time scales are involved. Several presentations therefore sought to explore these issues via (1) case-studies focused on the role that particular traits play in resistance to, or recovery from, disturbance, and (2) demonstrating techniques for measuring changes in functional trait space through time, and linking these to ecological resilience.

While these presentations elucidated the importance of response and effect traits in controlling ecosystem function, it became apparent that there are several hurdles to overcome to progress the field of functional paleoecology and resilience. A horizon scan of key issues pertaining to what these challenges comprise, and how they can be resolved, was the key focus of the workshop discussion groups (Fig. 1). Identifying which traits are both ecologically meaningful and quantifiable in the sediment record was highlighted as a major challenge for many microfossil types. Linked to this are issues related to differential (or low) taxonomic resolution, and quantification of drivers that require the development of a unique set of analytical approaches if functional ecological approaches are applied to paleoecological data. The final discussions centered on the potential to advance understanding related to conservation outcomes if these technical challenges can be overcome. An in-depth analysis of these challenges and the solutions required to overcome them form the basis of an upcoming EcoRe3 perspective paper.

The workshop concluded with a discussion regarding EcoRe3's future steps, including (1) the publication of a special issue based on the overall goals in 2019, and (2) regrouping the members for a themed session at INQUA 2019 in Dublin, Ireland. A follow-up workshop, focused around the statistical modeling of paleo datasets using techniques developed within EcoRe3, was also proposed. Details of future EcoRe3 activities will be updated on the website (pastglobalchanges.org/ini/wg/ecore3) and on the following social media outlets.

- EcoRe3 on Facebook (visit page and request to join): facebook.com/groups/286999515057710
- EcoRe3 on Twitter: @Eco_Re3
- EcoRe3 on ResearchGate: researchgate.net/project/EcoRe3-Resistance-Recovery-and-Resilience-in-Long-term-Ecological-Systems-2

AFFILIATIONS

[1]School of Culture, History & Language, The Australian National University, Canberra, Australia
[2]Geography Department, University of Utah, Salt Lake City, USA
[3]School of Environment, University of Auckland, New Zealand
[4]UMR CARRTEL, INRA-Université Savoie Mont-Blanc, Thonon-les-bains, France
[5]Laboratory of Wetland Ecology and Monitoring and Department of Biogeography and Palaeoecology, Adam Mickiewicz University, Poznań, Poland
[6]Department of Environmental Resources and Geohazards, Polish Academy of Sciences, Warsaw, Poland
[7]Research Development Office, University of Utah, Salt Lake City, USA

CONTACT

Rebecca Hamilton: rebecca.hamilton@anu.edu.au

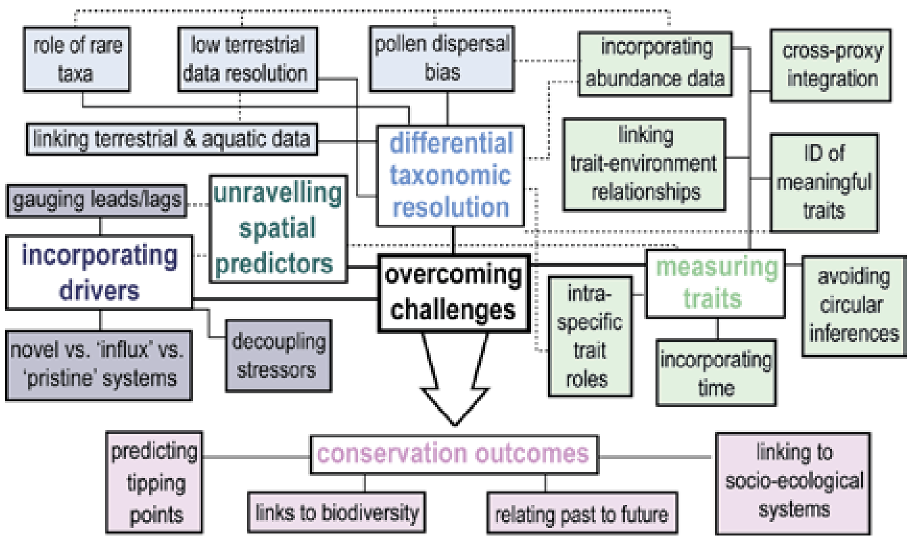


Figure 1: Key challenges to be resolved in paleoecology related to function and measuring resilience to advance conservation outcomes.

doi.org/10.22498/pages.26.2.88

African fire histories and fire ecologies

Colin J. Courtney-Mustaphi[1,2,3], D. Colombaroli[4,5], B. Vannière[6], C. Adolf[7], L. Bremond[8,9], J. Aleman[10] and the Global Paleofire Working Group (GPWG2)

Nairobi, Kenya, 19-22 July 2018

Patterns of fire are changing across African savannahs, rainforests, fynbos, woodlands, and Afroalpine and montane forests, with direct environmental and socio-ecological consequences. Fire variability has implications for biodiversity (Beale et al. 2018), vegetation patterns, grazing quality, carbon emissions, protected area management, and landscape heterogeneity.

Fire is a crucial component of savannah functioning and structure and is essential for maintaining its biodiversity. Long-term records are key to understanding drivers of fire variability and contextualize recent and ongoing land-use changes that altered fire responses to climate and vegetation changes (e.g. Ekblom and Gillson 2010, Colombaroli et al. 2014). As indigenous forest loss continues and modification through selective harvesting and land-use encroachment accelerate forest changes, the importance of historical disturbance regimes is increasingly relevant for assessing past ranges of variability and to define management targets that support more resilient socioecological systems (Whitlock et al. 2018). But how can the research community engage and integrate with land-management practitioners and policy developers? And how can we promote knowledge transfer and collaborative capacity between the international community and the next generation of African scientists?

Such themes were explored and discussed during a GPWG2 workshop following the African Quaternary Association (AFQUA) conference. The workshop gathered 18 participants from 12 countries, including 10 researchers from Africa-based institutions. It opened with introductory lectures and laboratory practical courses on study-site selection, sampling techniques, laboratory preparation and charcoal morphology analyses, and a discussion on charcoal calibration approaches for African sites (Ekblom and Gillson 2010; Adolf et al. 2018; Hawthorne et al. 2018; Fig. 1). Participants were introduced to quantitative paleoenvironmental data analysis techniques using R; those included reconstructing savannah fire responses to precipitation and biomass using Generalized Additive Models (GAMs) with data from Lake Naivasha, Kenya (Colombaroli et al. 2014). Dedicated breakout sessions involved data mining using the Global Charcoal Database (GCD) and examinations of spatiotemporal knowledge gaps, notably for western and central Africa, where spatial coverage is scant and several published records need to be imported in the GCD (Fig. 1e). The data gaps are also apparent from the driest and wettest ends of the precipitation gradient, and mostly lacking from mangrove, Afroalpine, and dry woodland study sites. Furthermore, only a limited number of study sites are located near archaeological sites, limiting analyses on human-environment interactions (Marchant et al. 2018). The discussion highlighted the need to develop high-quality charcoal series that account for potential biases in sediment accumulation and related chronological uncertainties (Colombaroli et al. 2014).

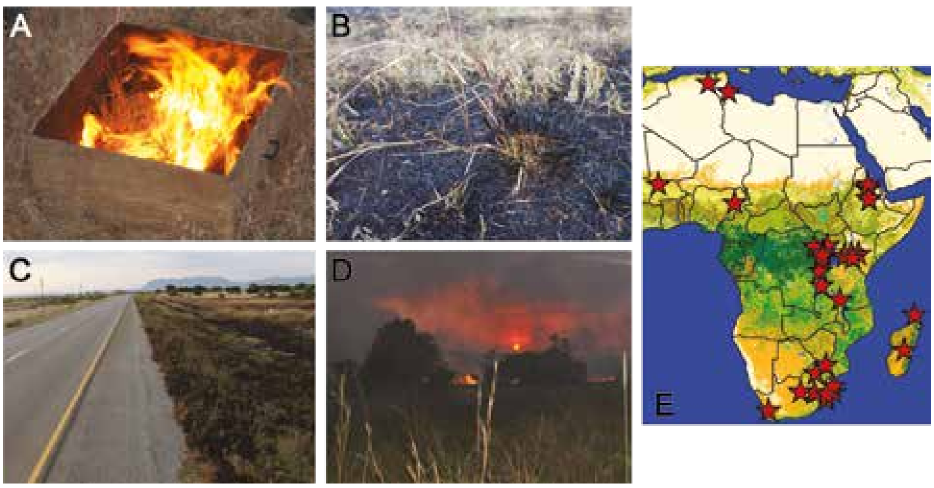


Figure 1: (A) Experiments, like this 1x1 m controlled grass burning, help relate vegetation-fire-charcoal proxy measurements to paleoecological charcoal records. **(B)** Residual culms and mainstems from grazing and burning that will regrow leaves and inflorescence. **(C)** Roads and linear infrastructure increase fragmentation altering spatial complexity for fires. **(D)** Grass and woody fuels burning near Klein's Camp, Serengeti National Park, Tanzania, 16 July 2016. **(E)** Published charcoal records that participants identified during a literature review, including records absent from the Global Charcoal Database (Map source: due.esrin.esa.int/page_globcover.php). Photographs: Colin Courtney-Mustaphi.

Finally, participants discussed the contribution of paleofire data to land-management applications (Fig. 1), including conservation and fire policy; a theme specifically addressed by the GPWG DiverseK framework-workshop (see Colombaroli et al., this issue). Participants were assigned sub-regions and discussed challenges and alternatives for managing fire as a component in socio-ecological systems (Whitlock et al. 2018). The discussion highlighted how resource management in the Menengai Forest, Kenya, and areas of southwest Madagascar are impacted by logging, unauthorized burning, resource extraction, and invasive species; while in the Bale Mountain National Park, Ethiopia, the existing contestation between pastoralists and conservation requirements necessitates new approaches to maintain ecosystem services and promote co-benefits. Approaches combining paleoecology and qualitative, local content optimize bidirectional knowledge transfers and encourage long-term engagement between the GPWG2 and new active members of the research community.

AFFILIATIONS

[1]Department of Environmental Sciences, University of Basel, Switzerland
[2]Department of Archaeology and Ancient History, Uppsala University, Sweden
[3]Environment Department, University of York, UK
[4]Department of Geography, Royal Holloway University of London, UK
[5]Institute of Plant Sciences, University of Bern, Switzerland
[6]CNRS-UMR Chrono-environnement, Université Bourgogne Franche-Comté, Besançon, France
[7]Department of Zoology, University of Oxford, UK
[8]Institut des Sciences de l'Evolution de Montpellier (CNRS, IRD, EPHE), Université de Montpellier, France
[9]École Pratique des Hautes Études, PSL Research University, Paris, France
[10]Département de Géographie, Université de Montréal, Canada

CONTACT

Colin J. Courtney-Mustaphi: colin.courtney-mustaphi@unibas.ch

REFERENCES

Adolf C et al. (2018) Glob Ecol Biogeog 27: 199-212
Beale C et al. (2018) Ecol Lett 21: 557-567
Colombaroli D et al. (2014) Glob Change Bio 20: 2903-2914
Ekblom A, Gillson L (2010) Holocene 20: 1063-1077
Hawthorne D et al. (2018) Quat Int 488: 3-17
Marchant R et al. (2018) Earth-Sci Rev 178: 322-378
Whitlock C et al. (2018) Biol Conserv 32: 84-97

doi.org/10.22498/pages.26.2.89

GLOBAL PALEOFIRE

DiverseK: integrating paleoecology, traditional knowledge and stakeholders

Daniele Colombaroli[1], J. Mistry[2], A. Milner[1], B. Vannière[3], C. Adolf[4], D. Hawthorne[5] and the Global Paleofire Working Group (GPWG2)

Egham, UK, 4-7 September 2018

Integration of ecosystem science and applied research in ecosystem management is a high priority and key challenge for the science-policy interface, as recently highlighted by the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES 2018). A more direct involvement of stakeholders and policymakers into the research agenda requires new approaches for knowledge transfer from the academic to the stakeholder community, as also emphasized during previous workshops organized by the Global Paleofire Working Group GPWG2 (Blarquez et al. 2018; Courtney-Mustaphi et al., this issue).

Thirty participants from 15 countries (Fig. 1) met at Royal Holloway University of London, UK, to discuss ongoing challenges on biodiversity conservation and fire policy, considering three approaches: (a) long-term ecology – informing on ecosystem responses to environmental change across regions and timescales (*paleoecology-informed conservation*); (b) local, traditional, and indigenous-knowledge systems on fire management that maintain biodiversity (*community-owned and -driven conservation*); and (c) conservation challenges and agendas defined by stakeholders and policymakers (*stakeholder-driven research*). The combination of long-term ecology with traditional knowledge represents a novel and alternative approach to promote a more sustainable management practice of present ecosystems under current threats, and fosters the dialogue between the different disciplines.

Before the workshop, conservation and fire-management evidence priorities were identified together with the UK Government Department for Environment, Food and Rural Affairs , and included: (1) the impact of changing climate and land use on fire regimes; (2) ecosystem recovery after fires of different severity; (3) the optimum fire regimes needed to achieve management objectives for biodiversity conservation; and (4) the effects of prescribed burning (and other land-management practices) on wildfires. The participants discussed examples of fire policies and relative impacts based on their regional expertise and on a field discussion in Chobham Common (Fig. 1). Different sub-groups then compiled evidence-based case studies on past ecosystem legacies, the role of local ecological knowledge in maintaining landscapes (Mistry and Berardi 2016), post-fire ecosystem dynamics, and burning conditions that optimize biodiversity across ecosystems (Colombaroli et al. 2013). This approach also highlighted potential conflicts in conservation targets for specific regions (natural vs cultural; e.g. Jackson and Hobbs 2009), and the need to anticipate new challenges in a future warmer world (Fischer et al. 2018). The participants produced a first draft of a policy brief summarizing best practices for sustainable ecosystem management, including how transdisciplinary knowledge can inform fire management and policy, and highlighted future priorities to foster effective science-policy knowledge exchange in this region. In addition, a group of early-career researchers focused on the compilation of a systematic review protocol to summarize the available evidence on the effects of climate change and land use on fire regimes.

Figure 1: (Left) Distribution of the participants' home countries. **(Right and bottom)** Guided discussion on fire risk and lowland heath-conservation priorities at Chobham Common, the largest National Nature Reserve in South East England. Photos: A. Milner and D. Colombaroli.

Adapting fire policy to more sustainable practices for specific regions requires approaches that focus primarily on the most urgent challenges set by policymakers (policy-driven research), in co-production with diverse knowledge and expertise (here paleoecology, cultural geography and policymakers). This dialogue can lead to the implementation of decision tools for policymakers, and further promote the integration of long-term ecology in more applied science (Willis and Birks 2006). The Global Paleofire Working Group, through the Diverse Knowledge framework (DiverseK), will continue to build a strong network of researchers, land managers, practitioners, and policymakers to tackle ongoing challenges in ecosystem management and biodiversity conservation.

ACKNOWLEDGEMENTS

Funding has been provided by PAGES, the Quaternary Research Association (QRA), and Chrono-environnement at Université Bourgogne Franche-Comté. The full meeting program is available at gpwg.paleofire.org/event/focus-group-3-workshop-london-2018. Additional material is on our Twitter account (@diverse_K). The motivation for this workshop was inspired by Cathy Whitlock, the GPWG community (gpwg.paleofire.org) and previous PAGES meetings in 2012 by K.J. Willis and E. Jeffers in Oxford, UK, (PAGES Focus 4 Biodiversity Theme Workshop, pastglobalchanges.org/calendar/127-pages/1006) and in 2017 in Bern, Switzerland, by H. Fischer, K. Meissner and A. Mix (PAGES Warmer Worlds Integrative Activity workshop, pastglobalchanges.org/calendar/127-pages/1653).

AFFILIATIONS

[1]Centre for Quaternary Research, Department of Geography, Royal Holloway University of London, UK
[2]Department of Geography, Royal Holloway University of London, UK
[3]UMR Chrono-environnement (CNRS), Université Bourgogne Franche-Comté, Besançon, France
[4]Department of Zoology, University of Oxford, UK
[5]AOC Archaeology Group, Edinburgh, UK

CONTACT

Daniele Colombaroli: daniele.colombaroli@rhul.ac.uk

REFERENCES

Blarquez et al. (2018) PAGES Mag 26(1): 40

Colombaroli D et al. (2013) Divers Distributions 19: 157-170

Fischer H et al. (2018) Nat Geo 11: 474-485

IPBES (2018) Plenary of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services Sixth session, 1151 pp

Jackson S, Hobbs RJ (2009) Science 325: 567-569

Mistry J, Berardi A (2016) Science 352: 1274-1275

Willis KJ, Birks HJB (2006) Science 314: 1261-1265

PAGES

doi.org/10.22498/pages.26.2.90

European land-use at 6000 BP: from on-site data to the large-scale view

Nicki Whitehouse[1], M. Madella[2] and F. Antolín[3]

Barcelona, Spain, 21-23 May 2018

The LandCover6k working group investigates whether prehistoric and historic human impacts on land cover (i.e. anthropogenic land-cover change due to land use - LULC) were sufficiently large to have a significant impact on regional climates. This workshop focused on reconstructing land use for the European continent by synthesizing farming and landscape management patterns derived from the archaeological record, for a 6000 BP time slice. We were also interested in the location of settlements, field systems, and industrial activities. The main objectives of the workshop were to:

(i) Review LU classifications developed for other regions (Morrison et al. 2018), ensure these are usable for Europe at 6000 BP, and develop geography-specific lower-classification categories;

(ii) Establish regions with good datasets for LU and produce preliminary LU maps following the methodology of Morrison et al. (2018);

(iii) Identify high-resolution case-study areas for detailed LU mapping;

(iv) Agree on a publication strategy and subsequent analyses.

Figure 1: Participants from the Northern Europe Group mapping their region. Photo credit: Welmoeud Out.

Workshop outcomes

Twenty-two European archaeobotanists, archaeozoologists, chronologists, GIS, climate and land-use modelers attended the workshop. In addition, some researchers joined the meeting remotely via an internet stream, or sent materials, and others have been approached since the meeting to contribute data from under-represented regions.

The first day and a half was devoted to expectations of the archaeological and modeling communities for the land-use development and best approaches towards synthesizing and reconstructing land use at a continental scale. There was considerable enthusiasm from the European community to be involved and contribute data. It was agreed that land maps should be developed from archaeological-site data points and then aggregated at a sub-regional level. Metadata should be fully accessible and traceable, and underpinned by high quality chronological data. This will allow an evaluation of uncertainty, ensure the work is archaeologically accurate and allow future revisions. A European database of ^{14}C dates developed by Marc Vander Linden (University of Cambridge, UK) was made available to facilitate mapping. Discussion also focused around "rules for interpolation" between data points and "unknown" areas, data quality and traceability. Each LU area should be assigned a percentage land-use value as far as possible, allocated at a regional or country level, based on expert knowledge; proportions of crops and animals will be assessed for particular regions.

There were extended discussions on LU categories (Morrison et al. 2018), which led to the addition of a LU category for "no evidence" and adjustments to lower-level categories for hunting-foraging-fishing, pastoralism and agriculture to fully reflect the European evidence. Further work around definitions will be needed. The final half-day of the workshop was spent undertaking initial top-level mapping on printed versions of gridded maps.

The main objectives of the workshop were met; more time was focused on item (i) than initially envisaged. However, this allowed the group to develop the necessary protocols for step (ii) and (iii). We are now progressing with step (ii) and have established areas for which there is good data coverage (iii). Datasets are well distributed across Europe, but further work will be needed to fill in data gaps and develop protocols for data quality.

Next steps

Maps and ^{14}C datasets have been transferred into a simple GIS database to produce preliminary LU maps. These will need considerable work and refining. A follow-up meeting is planned for 29-30 January 2019 at the Research Center for Wetland Archaeology, in Hemmenhofen, Germany.

We are keen for interested archaeologists to join the group, especially from regions that are currently less-well represented (e.g. France, some areas of North and central Eastern Europe, and southern parts of the Mediterranean). Visit our website for more information: pastglobalchanges.org/ini/wg/landcover6k

AFFILIATIONS

[1]School of Geography, Earth and Environmental Sciences, Plymouth University, UK

[2]ICREA and Department of Humanities, Universitat Pompeu Fabra, Barcelona, Spain

[3]Integrative Prehistory and Archaeological Science (IPAS), University of Basel, Switzerland

CONTACT

Nicki Whitehouse: Nicola.whitehouse@plymouth.ac.uk

REFERENCES

Morrison KD et al. (2018) PAGES Mag 26: 8-9

doi.org/10.22498/pages.26.2.91

New research directions for the PAGES C-PEAT working group

Julie Loisel[1] and Angela Gallego-Sala[2]

College Station, USA, 10-13 May 2018

Members of the C-PEAT working group met at Texas A&M University to define their research plan for Phase 2. The past three years have led to the development and publication of a large database containing over 500 peat records that have been primarily used to (1) reconstruct Holocene carbon and nitrogen sequestration rates across the northern peatland domain, and (2) connect centennial- and millennial-scale changes in carbon sequestration rates to key climatic forcings and environmental controls (Charman et al. 2012; Loisel et al. 2014; Treat et al. 2015). C-PEAT is in the process of making its entire peatland database available on WDS-PANGAEA; 164 sites are readily available under the project name PAGES_C-PEAT. Those same peat profiles have also been ingested in the International Soil Carbon Network's database (ISCN) and are accessible on its website. These two data sharing activities took place during a "data hackathon" shortly before this workshop. An article in *Earth System Science Data*, describing the entire database, is in development.

Building on the success of Phase 1, there is a clear need to grow C-PEAT's research scope. Workshop participants identified future tasks that ultimately aim to further integrate peatlands into land-surface models:

Expand the peatland database to encompass sites from the tropics and extra-tropics. C-PEAT now includes a large group of scientists with expertise in tropical peatlands. We know that humans are a major agent of change in tropical peatlands; our team intends to focus its effort on (a) quantifying carbon stocks and fluxes, (b) evaluating peatland ecosystem services, and (c) measuring the impact of land management on these peatlands. A paleo perspective is essential to develop a better functional understanding of these ecosystems and link their dynamics with global carbon cycling and land-use change.

Predict peatland responses to natural and anthropogenic disturbance. Workshop participants agreed that "sensitive processes" such as (a) peatland dynamics that govern decadal-scale vertical peat accumulation and net carbon balance, and (b) margin dynamics that control horizontal peat development (expansion vs. contraction) need to be further connected to disturbance regime including fire, permafrost thaw, invasion by new species, drainage, prescribed burning, and other land-management scenarios. Here, the paleo perspective acquired during Phase I will be used to compare and contrast the importance of climatic forcings, environmental controls, and land management on peat formation and carbon sequestration. Workshop participants are preparing a manuscript that addresses these new research directions and highlights the relevance of peatland dynamics in land-surface models.

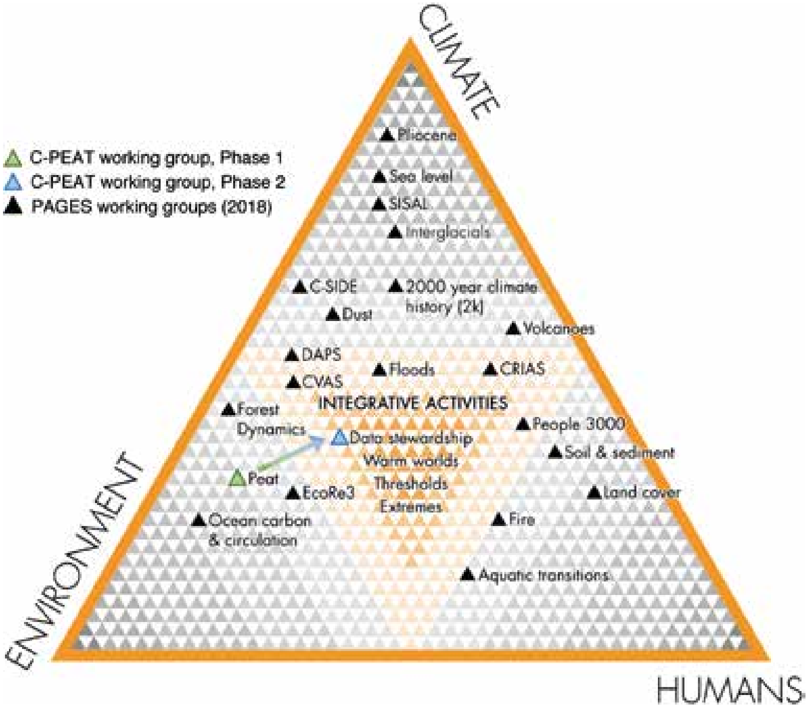


Figure 1: A new, more-integrative place for the C-PEAT working group within PAGES' science structure. This change for Phase 2 reflects the inclusion of tropical and extra-tropical peatlands, which are largely impacted by land management, and the development of new peat-based paleoclimatic proxies.

Further develop and promote the use of multi-proxy peatland records as paleoclimatic archives. The peat cores included in the C-PEAT database could be used in conjunction with those from other archives such as lake sediments, tree rings and ice cores. For example, there is an array of traditional and novel peat-based proxies for temperature and hydrology. Of particular interest to the C-PEAT group is (a) a combination of compound-specific stable isotope measurements to back-calculate changes in rainfall regimes (Amesbury et al. 2015), (b) a suite of novel organic biomarkers that are sensitive to temperature and pH or that provide insight into the carbon cycle (Naafs et al. 2017), and (c) the integration of more-traditional proxies such as plant macrofossils and testate amoebae into process-based peat models such as DigiBog (Baird et al. 2012) and the Holocene Peat Model (Frolking et al. 2010) to further understand the encoding of these proxies into the peat matrix over time. Many of these datasets are already integrated to WDS-Neotoma.

Thus far, the C-PEAT working group has been associated with "environmental research" in the PAGES science structure (Fig. 1), as peatlands are key biosphere ecosystems that interact with the climate and introduce feedbacks into the Earth system. While this remains true, Phase 2 brings peat towards the center of the triangle (Fig. 1), as this working group is integrating land-use change and natural disturbance.

ACKNOWLEDGEMENTS

We thank PAGES, INQUA, and Texas A&M's Department of Geography for financial and logistical support.

AFFILIATIONS

[1]Department of Geography, Texas A&M University, College Station, USA

[2]Department of Geography, College of Life and Environmental Science, University of Exeter, UK

CONTACT

Julie Loisel: julieloisel@tamu.edu

REFERENCES

Amesbury MJ et al. (2015) Geochim Cosmochim Acta 164: 161-174

Baird AJ et al. (2012) Ecohydrology 5: 242-255

Charman DJ et al. (2012) Biogeosciences 10: 929-944

Frolking S et al. (2010) Earth Sys Dyn 1: 1-21

Naafs BDA et al. (2017) Geochim Cosmochim Acta 208: 285-301

Loisel J et al. (2014) Holocene 24: 1028-1042

Treat CC et al. (2015) J Geophys Res Biogeosci 121: 78-94

doi.org/10.22498/pages.26.2.92

The role of dust in climate change: A biogeochemistry perspective

Gisela Winckler[1], F. Lambert[2] and E. Shoenfelt[1]

Las Cruces, Chile, 8-10 January 2018

Mineral-dust aerosols are critically important components of climate and Earth system dynamics as they affect radiative forcing, precipitation, atmospheric chemistry, surface albedo of ice sheets, and marine and terrestrial biogeochemistry, over significant portions of the planet. Dust-borne iron is recognized to be an important micronutrient in regulating the magnitude and dynamics of ocean primary productivity and affecting the carbon cycle under past and modern climate conditions. Paleodata suggest large fluctuations in atmospheric dust over the geological past. However, dust-transport models struggle to reproduce observed spatial and temporal dust-flux variability. In addition, observational and modeling studies based in the current climate suggest that not all iron in dust is equally available to continental and ocean biota. Iron solubility varies dramatically, depending on mineralogy and state of soils, as well as atmospheric processing by acids. Modeling studies, however, still mostly assume constant solubility.

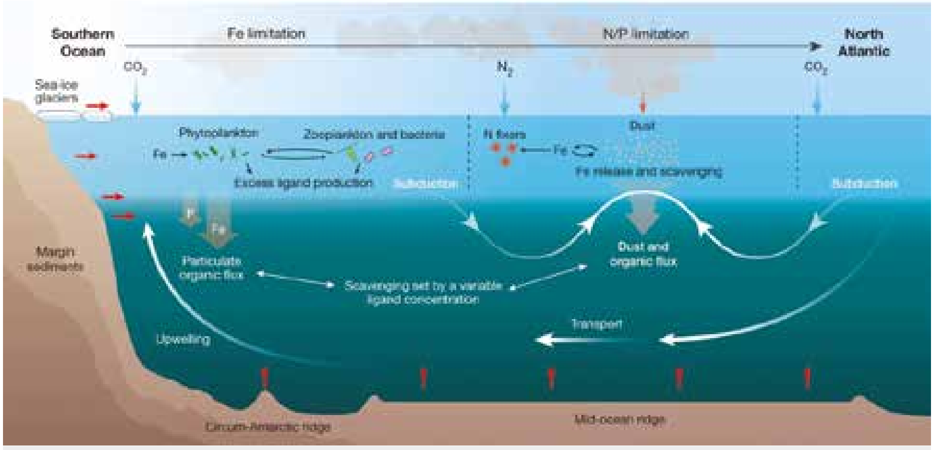


Figure 1: Representation of the major processes in the ocean iron cycle from Tagliabue et al. 2017. Reprinted by permission from Springer Nature.

The PAGES Dust Impact on Climate and Environment (DICE) working group held its first workshop on *The Role of Dust in Climate Change: A Biogeochemistry Perspective*. Twenty-seven experts from nine different countries came together at the Coastal Marine Research Station of the Catholic University of Chile in Las Cruces for the three-day workshop that was jointly supported by PAGES and the Chilean Comisión Nacional de Investigación Científica y Tecnológica. About half of the participants were early-career researchers or scientists from developing countries. The workshop format combined keynote talks with shorter thematically matching pop-up talks, followed by discussion and brainstorming of future research avenues among all participants.

This workshop was a highly interdisciplinary effort to better quantify and simulate biogeochemical impacts of dust deposition. It gave observationalists and modelers the chance to combine perspectives on the role of dust in ocean biogeochemical cycles and the greater carbon cycle. The specific goal of the workshop was to more precisely quantify the effects of mineral dust, and specifically iron, during various climate states, and to strategize how to further foster this relationship in future work.

The major themes of the keynote presentations followed the pathways of dust emissions, from geochemically identifying and tracing source regions to deposition in the surface ocean. Participants discussed the effects on the solubility and bioavailability of iron at each of these stations in the dust cycle, including the influence of dust source and mineralogy on dust solubility and bioavailability, the role of atmospheric processing in dust solubility, and the importance of organic ligands and aggregation to iron lifetime in the mixed layer of the ocean for bioavailability in modern and past oceans.

Halfway through the workshop, participants split into three breakout groups for extended discussion on the following three avenues of future research (1) Ocean iron sources (aeolian versus others), (2) Iron solubility past, present, future, and (3) Dust particles: shape, size and composition.

As a direct interdisciplinary outcome of the workshop, participants are developing a model that evaluates the relative importance of atmospheric processing, mineralogy, size fraction, settling rate, and ligand-mediated dissolution on the solubility of dust-borne iron that reaches the ocean (Fig. 1). This is in an effort to combine experimental and modeling results to improve the estimates of iron solubility in biogeochemical models – moving beyond the assumptions of constant iron solubility. The group plans to combine observational data on dust-source mineralogy, dust deposition, atmospheric processing, and ligand concentrations and strengths with kinetic models of dust dissolution in the water column. Such a model can be applied to both modern as well as past dust sources that may have been more highly impacted by glacier physical weathering, and have been shown to have a different mineral composition as a result. The goal of this effort is to combine the highly varied range of expertise to better quantify the bioavailable iron in different dust sources across space and time.

AFFILIATIONS
[1]Lamont-Doherty Earth Observatory, Columbia University, New York, USA
[2]Department of Physical Geography, Pontifical Catholic University of Chile, Santiago, Chile

CONTACT
Gisela Winckler: winckler@ldeo.columbia.edu

REFERENCES

Tagliabue A et al. (2017) Nature 543: 51-59

doi.org/10.22498/pages.26.2.93

Ocean circulation and carbon cycling during the last deglaciation: global synthesis

Janne Repschläger[1], A. Schmittner[2], J. Lippold[3], L. Skinner[4] and J. Muglia[2]

IPODS-OC3 workshop, Cambridge, UK, 6-9 September 2018

The transition of Earth's climate from the cold Last Glacial Maximum (LGM) to the warm Holocene is still not fully understood. It is associated with a rise in atmospheric CO_2 acting as an important driver of the global warming. Although changing ocean circulation during the deglaciation has been linked to the rates of CO_2 increase, its exact mechanism remains unclear, and quantifying changes in the ocean circulation and carbon storage remain challenging tasks.

The overarching goal of PAGES' working group Ocean Circulation and Carbon Cycling (OC3) is to compile global foraminiferal carbon ($\delta^{13}C$) and oxygen ($\delta^{18}O$) isotope data and compare them with model simulations to better understand changes in ocean circulation and carbon cycling during the last deglaciation. INQUA's Investigating Past Ocean Dynamics (IPODS) focus group works towards the same scientific objective by synthesizing complementary global ocean circulation proxy data such as radiocarbon ($\Delta^{14}C$), εNd and Pa/Th. This joint IPODS-OC3 workshop aimed to make progress through new regional and global data syntheses and their combination with isotope-enabled model simulations.

The workshop brought together 40 international scientists – sea-going paleoceanographers and modelers – to report the progress within the working groups, to discuss potential overlaps between the communities and to identify potential joint aims.

On the first day of activities, OC3 regional syntheses group leaders gave updates of current activities and work progress from different ocean basins (North and South Atlantic, and Indian Ocean). Additionally, posters were presented on regional datasets and modeling approaches. $\delta^{13}C$ and $\delta^{18}O$ data were discussed with respect to the importance of deep and intermediate water distribution and circulation, and importance of the different ocean regions for CO_2 storage and deglacial CO_2 release.

The focus of the discussion was on new age-model and data-visualization software tools, including age-model approaches PARIS (Claire Waelbroeck), HMM-Match (Lorraine Lisiecki) and the age-model software and visualization tool PaleoDataView (Stefan Mulitza). Discussions included the differences of the approaches, age-model error assessments and uncertainties in reservoir age corrections.

Age models are an important backbone of OC3 and IPODs databases. As nearly all cores in the IPODS database are also included in the OC3 database (Fig. 1), the same age models will be used for the global synthesis to enable the direct comparison of different proxy records from the same cores.

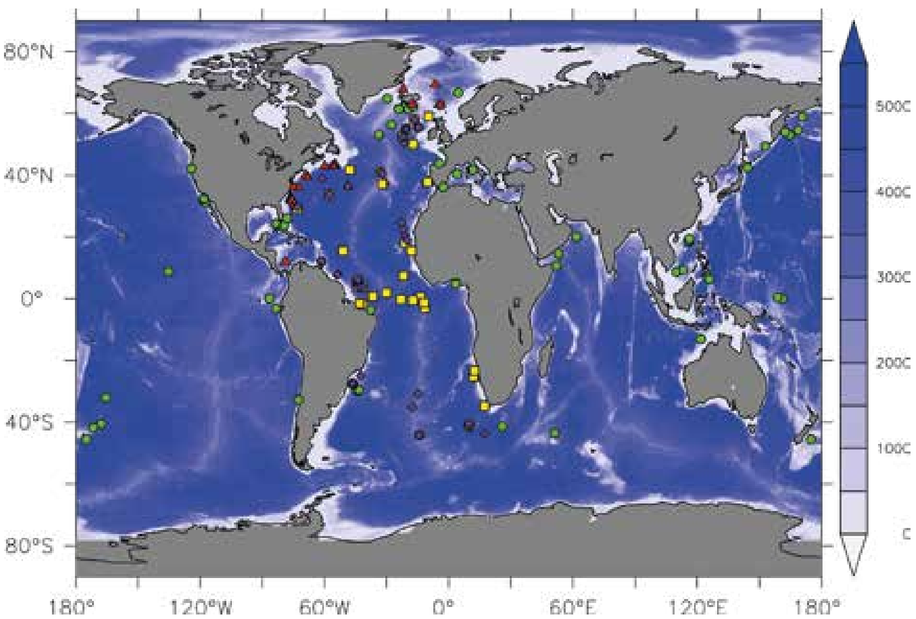


Figure 1: Map showing cores included in the OC3 and IPODS databases covering the last deglaciation. Green circles show published high-resolution $\delta^{13}C$ data prioritized for quality control. Red triangles are radiocarbon data from Zhao et al. (2018). Yellow squares are Pa/Th data (Ng et al. 2018) and purple diamonds are εNd data.

The second day of the conference was designated to the IPODS working group. General summaries of the current knowledge in εNd, Pa/Th and $\Delta^{14}C$ were given by work-group leaders, followed by presentations of recent compilation studies (Pietrowski, Blaser; Ng et al. 2018).

The final day of the workshop was dedicated to future plans of OC3 and IPODS. The participants encouraged OC3 to proceed with the database effort and developed a strategy for the refinement of the data synthesis, including the development of quality-check guidelines and consistent age-model generation. Extensive labor is needed to generate consistent age models and to ensure data quality. A detailed work plan for the next year includes the release of the final database structure with already published regional synthesis datasets, publishing of the Atlantic Ocean synthesis in the beginning of 2019, and work on syntheses papers in the summer of 2019. The chair of OC3 was passed from Andreas Schmittner to Janne Repschläger. Juan Muglia will join the steering committee in the function of database manager.

The workshop added momentum to IPODS and OC3 activities such as the writing of an εNd review paper, the creation of a joint database including $\delta^{13}C$, $\delta^{18}O$, εNd, Pa/Th and $\Delta^{14}C$, and the exploration of common plans for the future.

AFFILIATIONS

[1]Max-Planck Institute for Chemistry, Climate Geochemistry, Mainz, Germany
[2]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, USA
[3]Institute of Earth Sciences, Heidelberg University, Germany
[4]Department of Earth Sciences, University of Cambridge, UK

CONTACT

Janne Repschläger: J.Repschlaeger@mpic.de

REFERENCES

Ng HC et al. (2018) Nat Commun 9: 2947

Zhao N et al. (2018) Paleoceanogr Paleoclimatol 33: 128-151

doi.org/10.22498/pages.26.2.94

Increasing social complexity, climate change, and why societies might fail to cope

David A. Byers[1], M. Lima[1], A. Gil[2], E. Gayo[3], C. Latorre[3], E. Robinson[4] and R. Villalba[2]

San Rafael, Argentina, 14-18 May 2018

The PEOPLE 3000 working group focuses on integrating archaeological and paleoecological case studies with mathematical modeling. We seek to understand how co-evolving human societies and ecosystems can successfully cope with the interrelated forces of population growth, increasing social complexity and climate change, and the diversity of trajectories of reorganization that social-ecological systems follow. Our work focuses on the observation that human societies experienced periods of social and economic development followed by major reorganizations throughout the Holocene. Thus, we are investigating explanations for what appears to be widespread and, potentially, climate-driven patterns.

Much of our research builds on the Variance Reduction Safe Operating Space (VRSOS) hypothesis (Anderies 2006). The VRSOS proposes that population growth, increasing complexity, and increasing energy consumption reduce variation in human subsistence economies. This, in turn, results in systems where individuals are well adapted to a specific range of climate variation, but where those same strategies are easily disrupted by climate change outside the range to which a society has adapted. Whether social-ecological systems follow a collapse trajectory or a new growth trajectory depends on the adaptive capacity of individuals within ecological and institutional constraints. To date, we have begun to explore such relationships with a series of related publications and grant proposals (Freeman et al. 2018a,b).

The May 2018 workshop moved our research forward by synthesizing radiocarbon, paleoecological, and subsistence data in ways that allow the working group to integrate them with mathematical models. The workshop resulted in the following outcomes:

1. The collation of radiocarbon and paleoenvironmental records from North and South America, Europe and Australia.

2. Outlines of manuscripts exploring ways the VRSOS hypothesis can predict broad shifts in human subsistence and settlement decisions across the Holocene. This hypothesis allows for the identification of periods when we would predict societal reorganizations and how those adaptive shifts may occur. Our data suggest that Late Holocene prehistoric societies across the globe experienced trajectories of adaptive reorganization, subsistence specialization, and sometimes societal collapse in response to environmental changes as predicted by the VRSOS.

3. Outlines of manuscripts deploying well-established logistical growth models derived from population ecology to understand how shifts in carrying capacity can condition demographic regimes across the Holocene. In this case we define a demographic regime as a one defined by a fixed carrying capacity. Shifting carrying capacity moves a population into a new demographic regime. We explore this idea by identifying different patterns of growth and climate-mediated collapses in the more distant past (Lima Arce 2014). For example, Figure 1a illustrates the long-term growth (trends) and medium-term fluctuations (peaks and valleys) of human populations and economies from the California (USA) case study discussed at the workshop. This case displays long-term growth, but with various medium-term spurts of growth and decline. However, although the broad trend displays long-term growth, the type of growth is variable. Figure 1b applies discrete time logistic growth models to the data and suggests that in California, the long-term growth of human societies was driven by a shift in their carrying capacity, either through social-technological change or climate change. This methodology allows us to compare demographic regimes and better understand how and why they change.

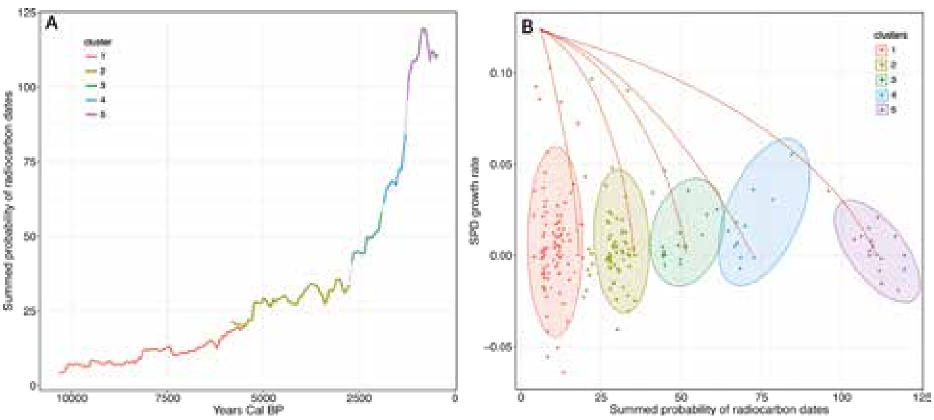


Figure 1: (A) Summed probability distribution of California, USA, radiocarbon ages against time for five demographic regimes. **(B)** Discrete time logistic growth models for each of the five demographic regimes for the California case study.

The workshop outcomes opened several new lines of research. Specifically, we will move forward by exploring relationships between archaeological and paleoecological datasets within the context of the VRSOS and logistic models discussed above. Of greatest interest to us are the ways that low-frequency, but high-amplitude, environmental events can prompt for societal adaptations resulting in either positive or negative demographic changes. Understanding what social and ecological factors conspire to condition the direction of demographic changes, and using this information to predict the direction and magnitude of change for specific archaeological phenomena, outline our next challenges – ones we aim to address in future studies.

AFFILIATIONS
[1]Anthropology Program, Utah State University, Logan, USA
[2]CONICET-IANIGLA/San Rafael, Technological National University (UTN-FRSR), Mendoza, Argentina
[3]Department of Ecology, Pontifical Catholic University of Chile, Santiago, Chile
[4]Department of Anthropology, University of Wyoming, Laramie, USA

CONTACT
David A. Byers: david.byers@usu.edu

REFERENCES

Anderies JM (2006) J Institut Eco 2: 133-155
Freeman J et al. (2018a) Radiocarbon 60: 453-467
Freeman J et al. (2018b) PNAS 115: 9962-9967
Lima Arce M (2014) Ecol Evol 4: 284-291

doi.org/10.22498/pages.26.2.95

Holocene abrupt climatic events and the environmental effects

Yu Liu[1], J.R. Dodson[1,2] and W. Cai[3]

Xi'an, China, 18-21 June 2018

This workshop was organized by the Institute of Earth Environment at the Chinese Academy of Sciences and attracted speakers from across China, as well as Australia, India, Japan, and USA. A total of 35 talks were presented, covering diverse topics from identifying abrupt events and their impact on past societies; interdecadal variability; modeling and dynamical forcing from the Atlantic, Indian and Pacific Oceans; the Sun; and aerosols. A large number of new high-resolution datasets were presented. A key feature of the meeting was the large number of presentations by young scientists, a sure sign of a healthy future for Holocene science.

In the past there have been conflicts arising from climate changes and even possibly dynastic collapses. Understanding how humans coped with changes in the past, evident in high-resolution time series, provides a template relevant for policy and planning issues. The role of humans as drivers of abrupt changes in the Holocene in various parts of the world was discussed, as well as the expression of abrupt changes such as the 9.2, 8.2 and 4.2 ka BP cooling and drying events, which are represented to varying degrees in different regions. Several analyses focused on modern process studies and on how these can help improve precision in interpretation of paleoseries. The recent international cooperative ocean monitoring program (led by the Ocean University of China, Qingdao) is an outstanding example of this, and despite its yet still short time series, it is already yielding new insights into ocean dynamics.

ENSO events now show evidence of enhancement by volcanic forcing, and appear to be stronger in the 20th century than any other period in the Holocene. The future expression of ENSO is of vital concern for many parts of the world. While no two ENSO events are the same, Figure 1 shows how a modeled future configuration will lead to increasing numbers of extreme events. New evidence is emerging on the frequency of droughts in eastern Asia and on how these are related to variability and intensity of the Asian Summer Monsoons, especially in the later Holocene. There is now a strong case emerging for the forcing of climate change in eastern Asia modulated by both the Atlantic Meridional Ocean Oscillation and the Pacific Decadal Oscillation. It was noted that much good science emerges from international cooperation, and at present this is taking place in good spirit and above regional geopolitical considerations.

Some time was spent identifying areas that are particularly vulnerable to future change in Asia. These include Central Asia, the Indian Monsoon region and North East China, and are likely to provide challenges for the people who live there and from a geopolitical aspect.

Young scientists were challenged to think beyond simple series comparisons when considering their own data, and to keep in mind that while some series do in fact show similarities, these sometimes break down due to phase shifts and altered forcing conditions. Systems become better understood when the full range of variability is revealed and studied.

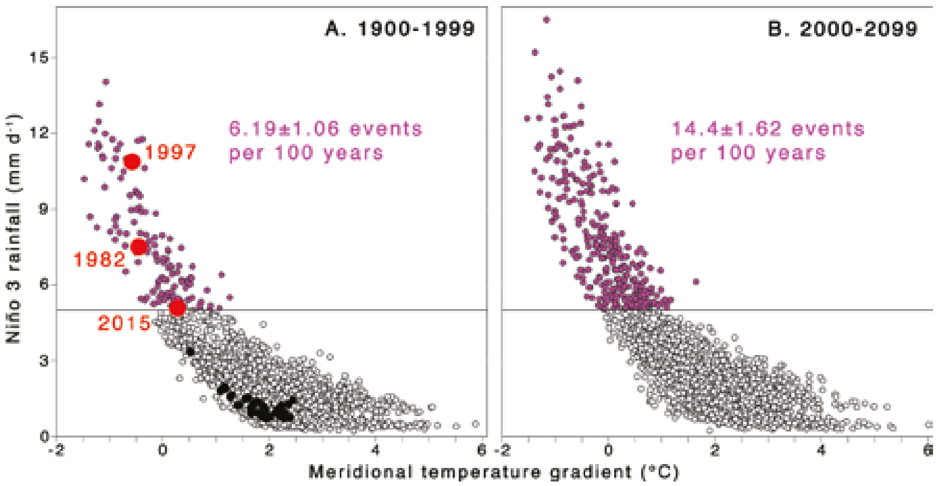


Figure 1: Relationship of meridional SST gradient (5-10°N, 150-90°W minus 2.5°S-2.5°N, 150-90°W) with December-January-February (DJF) mean Niño 3 (5°S-5°N, 150-90°W) rainfall, for **(A)**, Control Climate (1900-1999 CE) and **(B)**, Climate Change (2000-2099 CE) periods, respectively, aggregated over 21 selected CMIP5 models. Purple dots indicate modeled extreme El Niño events with a Niño 3 DJF rainfall greater than 5 mm per day and black circles indicate other modeled events. The corresponding average frequency of modeled extreme El Niño over the Control Climate period and Climate Change period is labelled in each panel, with the 95% confidence intervals based on a Poisson distribution. Red dots in (A) indicate observed 1982/83, 1997/98 and 2015/16 CE extreme El Niño events and black dots indicate other observed events since 1979. For more details see Cai et al. (2014).

AFFILIATIONS
[1]Institute of Earth Environment, Chinese Academy of Sciences, Xi'an, China
[2]Faculty of Science, Medicine and Health, University of Wollongong, Australia
[3]Centre for Southern Hemisphere Oceans Research, Hobart, Australia

CONTACT
John R. Dodson: john@ieecas.cn

REFERENCES

Cai W et al. (2014) Nat Clim Change 4: 111-116

CONTENTS

PAGES
PAST GLOBAL CHANGES

PAGES International Project Office
Hochschulstrasse 4
CH-3012 Bern
Switzerland

Telephone +41 31 631 56 11
Email pages@pages.unibe.ch
Website pastglobalchanges.org
Twitter @PAGES_IPO
Facebook PastGlobalChanges

Subscribe to PAGES magazine at
pastglobalchanges.org/products/pages-magazine

Parent program
PAGES is a Global Research Project of Future Earth.

Supporters
The PAGES International Project Office and its publications are supported by the Swiss Academy of Sciences and the US National Science Foundation.

Printed on recycled paper by
Läderach AG
Bern, Switzerland

Hardcopy circulation 2350

ISSN 2411-605X / 2411-9180

doi.org/10.22498/pages.26.2