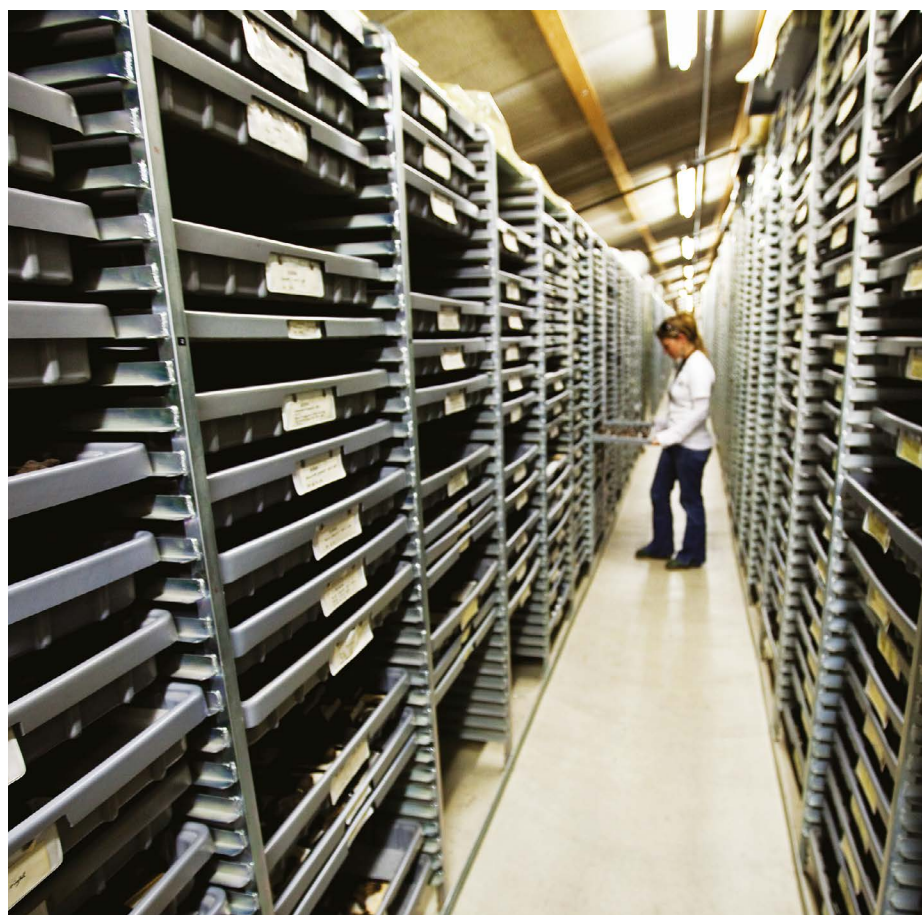


# Comment



A staff member views the fossil collection at La Brea Tar Pits museum in Los Angeles, California.

## Store and share ancient rocks

Noah Planavsky, Ashleigh Hood, Lidya Tarhan, Shuzhong Shen and Kirk Johnson

Geological samples must be archived for all if we are to solve the riddles of Earth's complex history.

**G**eologists think they know the basics of Earth's history. Liquid water has flowed on the planet for 4 billion years<sup>1</sup>. Tiny amounts of oxygen first gathered in the atmosphere about 2.3 billion years ago<sup>2</sup>. And the planet went through many periods of climatic upheaval, from freezing completely 700 million years ago<sup>3</sup> to warming so rapidly about 250 million years ago that more than 80% of marine species were lost<sup>4,5</sup>. It has had many more ups and downs.

This story can be reconstructed using data wrestled from ancient rocks. But as geologists

learn more, our planet's tale is getting muddier rather than clearer. Controversies have erupted in the past two decades over many aspects of the chemical record of the early Earth, including the evolution of life, environments and past long-term climate (see 'Contentious timeline').

For example, variations in carbon-isotope ratios in carbonate rocks have conventionally been interpreted as recording drastic global environmental changes, including huge episodes of volcanism or bursts of oxygen<sup>6</sup>. By contrast, some researchers suggest that these same records have been changed over time by local environmental processes, and that they do not provide information about Earth's ancient history<sup>7</sup>. This debate can be resolved only by applying a variety of geological and chemical tools<sup>8,9</sup> to the same samples used to generate the carbon-isotope results.

Attempts over the past decade to answer questions using better tools and larger databases have only amplified disputes. To make matters worse, too often, rock samples are not archived or shared. It is common for samples to be held by researchers in private collections instead of in accessible, curated institutional archives or museums. That's a problem, because different geoscience teams cannot check each other's work to test whether published results are robust and can be replicated.

We call on researchers, museums, funders, scientific societies and journals to ensure that all samples of sediment and sedimentary rock from which geochemical data have been produced and published are curated, archived and made available to members of the research community.

### Reproducibility crisis

Geological records are complicated and hard to interpret. It is easy to reach contradictory conclusions, most commonly for the following four reasons.

**Proxies and archives.** Several geochemical methods can be used to infer past conditions such as temperature. The same method applied to different sedimentary rock types can lead to inconsistencies. For example, the ratio of heavy to light oxygen isotopes in chemical precipitates (such as chert, carbonate or apatite) tracks the seawater temperatures under which these minerals formed. But even in the same piece of rock, the reconstructed temperatures can be different depending on whether they are measured in a fossil or in a bulk aggregate of the entire rock sample. This is because rocks are inherently combinations of different minerals, which might have

## Comment

formed during different stages of a rock's long geological history. The consequences for understanding past climates can be dramatic. For example, it is still not clear whether an interval of extreme heat killed marine organisms during the 'Great Dying' 250 million years ago. Sulfide toxicity, ocean acidification and carbon dioxide poisoning have also been proposed as possible mechanisms for killing off organisms at this time<sup>4</sup>.

Similarly, the question of whether oxygen levels were low enough to have delayed the emergence of animals for around 4 billion years – or most of Earth's history, thus addressing Charles Darwin's dilemma of why complex life appeared so late in the fossil record – depends on which rocks are studied and what analytical methods are used<sup>8</sup>. For example, an analysis of gas bubbles in sedimentary rocks<sup>9</sup> has suggested that atmospheric oxygen levels on Earth's surface would have been high enough to support animals as early as 2.6 billion years ago. However, this clashes with a compelling body of evidence indicating that atmospheric oxygen concentrations were vanishingly low at this time<sup>10,11</sup>. Refining such proxies is extremely challenging when different teams cannot work on the same samples.

**Geographical and temporal variation.** Rock samples that are used to tackle the same research question are often collected from different places, where the rocks were deposited at various times and in vastly different environments. This can result in completely distinct answers. For example, mercury enrichments in sediments are used as a tracer of large episodes of volcanic activity and their links to mass extinction events<sup>12</sup>. However, mercury enrichments can also result from wildfires or from local depositional conditions that lead to heavy-metal uptake by sedimentary organic matter<sup>12</sup>. Furthermore, diverse geographical settings can record mercury enrichments differently, depending on aspects such as water depth, dissolved oxygen concentrations, the rate of sediment deposition and the type and location of the volcanoes themselves<sup>12,13</sup>. All of this can lead to spurious correlations between volcanism and extinction events. It is difficult to disentangle signals of global changes in the Earth system through time from local environmental variability using only reported geochemical data sets.

**Analytical reproducibility.** Experiments can be hard to repeat even if rocks are pristinely preserved. Measurements are routinely checked against those of geochemical standard materials, the compositions of which are internationally validated. Yet there is always the possibility of errors during analysis. These can arise from differences in sample preparation (such as in rock-crushing techniques or in the type of acid used to prepare a sample)

and instrumentation (machine type, tuning) to variations in laboratory conditions. For instance, boron-isotope measurements on marine carbonates are one of the key tools used to reconstruct atmospheric CO<sub>2</sub> levels<sup>14</sup>. Various approaches to making such measurements can lead to CO<sub>2</sub> estimates that differ by more than 400 parts per million<sup>14,15</sup> – roughly equivalent to the total concentration of CO<sub>2</sub> found in the atmosphere today.

**Contamination and alteration.** As sediments become rocks, they undergo many processes that can alter the geochemical signals of where and how they formed. Sediments laid down on sea floors or lake bottoms can experience changes in water level or salinity, for example if they are flushed with meltwater. Hydrothermal processes and heat at depth might leach chemicals from the rock and alter the mineral composition.

Rocks collected near the surface can be altered by groundwater or other contaminants, such as oil used to drill cores. For example, organic remains in rocks once thought to be evidence for oxygen production by pioneering photosynthetic microorganisms 2.7 billion years ago are now acknowledged to be probable contamination from the modern petroleum products used to drill the rocks from the ground<sup>16</sup>. Similarly, debate is raging over whether the chemical composition of ancient rocks records microbial oxygen production extending as far back as 3 billion years ago, or whether those rocks have been compromised by contact with recent groundwater<sup>17</sup>.

### Precious prizes

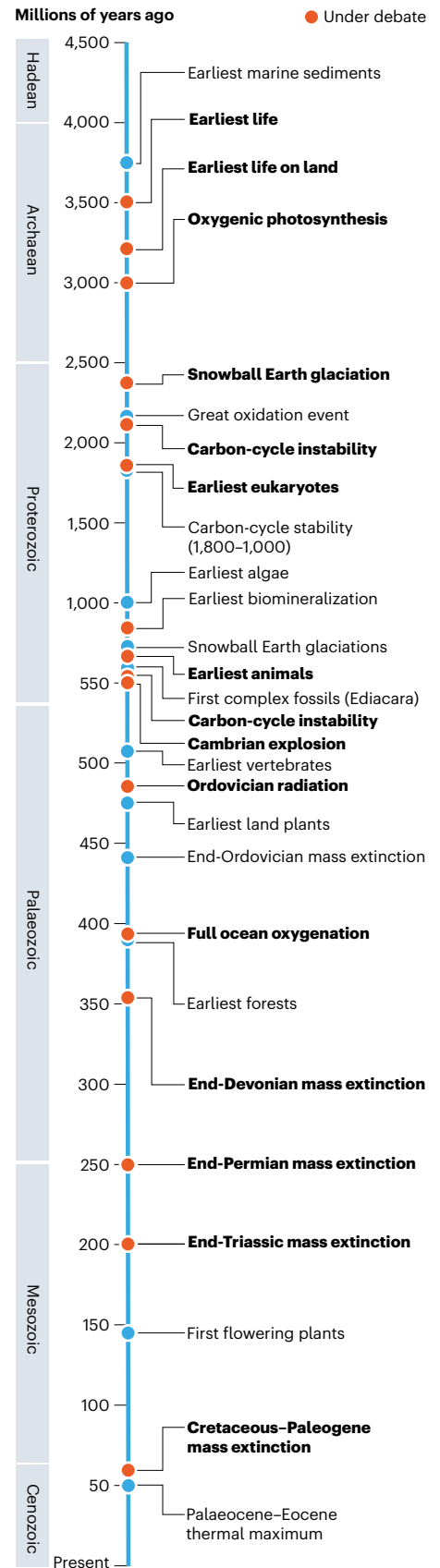
Without the ability to access and remeasure samples, it can be challenging to work out whether disparities in results and views stem from complexity in Earth's history, from sampling of rocks with different levels of alteration or from analytical issues. Yet sample archiving is not part of the standard protocol for inorganic or organic geochemical work, nor for some palaeoclimate work (other than, for example, ocean drill-core or ice-core samples, which are stored).

Why has this situation arisen? Many scientists are reluctant to share samples they have struggled hard to collect. After all, there are high costs associated with fieldwork on outcrops and drilling programmes. Research groups might want to perform multiple geochemical studies on a single set of samples, and this takes time. Large geochemical studies that use unconventional isotope systems can take several years to extract a data set<sup>18</sup>.

Other obstacles to archiving samples include how to fund archiving, where to store samples and how they are to be managed. Clearly, no single museum can hold all geological and geochemical samples. Museums would need to increase staff, space and funds

## CONTENTIOUS TIMELINE

Earth's environmental history has been reconstructed with data wrestled from ancient rocks. Some events remain hotly debated. Archiving and sharing of rock samples enable published work to be tested and replicated.



for such collections.

Initiatives for archiving materials in other fields could serve as models. These include the Global Genome Initiative, a shared data protocol for frozen tissue repositories (see [go.nature.com/3f4erur](http://go.nature.com/3f4erur)), and the Integrated Digitized Biocollections project for biological digital data ([www.idigbio.org](http://www.idigbio.org)). Global databases of these sample archives and their accessory information, building on initiatives such as the International Geo Sample Number (IGSN; [www.igsn.org](http://www.igsn.org)), will also be needed to assign unique identifiers and maintain inter-collection records.

Some Earth-science fields already deposit samples in publicly accessible museums. For instance, palaeontologists have been required to do so for samples formally described in scientific publications for more than 150 years. Likewise, museums hold type specimens of fossils, meteorites and biological samples. Well-funded drilling projects also have strict archiving policies and well-curated core libraries, such as that for the International Ocean Discovery Program (see [go.nature.com/2xoumhh](http://go.nature.com/2xoumhh)).

The FAIR data initiative offers strict guidelines on data archiving and has been adopted by many journals that publish Earth- and environmental-science research, including *Science* and *Nature* (see [go.nature.com/2wv2jxd](http://go.nature.com/2wv2jxd)). Although the recommended best practices of this initiative already include sample archiving, this is not yet strictly implemented as a formal requirement for publication.

### Forging consensus

Together, researchers, natural history museums, journal editors, scientific societies and funding agencies must develop and implement standardized archival policies. We recommend that the following steps are taken.

Geochemical researchers should routinely send their samples to museums. To encourage buy-in, we suggest an embargo period for delaying new studies by other research groups on each set of samples from which geochemical data have been published. Geochemists must also work with museums to broaden the conventional definition of collections to include a range of different materials, from fist-sized specimens to rock fragments, powders and mineral grains. Geochemists should work with custodians of protected lands to encourage the inclusion of archival policies and procedures for geochemical samples collected under research permits.

Natural history museums should broaden their mission to archive and curate geological samples. They should assign unique identifiers that can be logged in digital databases. Curators must decide how much of a sample can be withdrawn, because geochemical tests are destructive. Where resources are tight, museums will need to evaluate the

spatial, financial and scientific capacity of collections, and determine which samples are most essential to curate.

Scientific societies must tackle the question of what constitutes an acceptable repository. For instance, the Meteorite Society's Committee on Meteorite Nomenclature does this. Scientific societies such as the Geochemical Society in Washington DC and the European Association of Geochemistry in Aubière, France, should begin to recommend suitable institutions.

Recent decades have demonstrated that rapid changes in data archiving are possible when clear guidelines – and editorial mandates – are in place. So we would like to see journals go further in supporting the FAIR data initiative, by making requests to archive samples and assignment of database unique identifiers mandatory for publication.

Many scientific journals regulate data archiving using a checklist. We recommend that this practice be implemented for sample archiving, and that repository-issued sample identifiers (as well as unique identifiers assigned by inter-institutional database efforts such as the IGSN) be included in each paper. All major changes to a field take time to develop, and changes at the editorial level

**“We estimate that roughly 200,000 new sedimentary geochemical samples are analysed each year.”**

can help to nudge them along. Journals could implement these policies on a relatively short timescale, as long as exceptions are initially made to the archiving mandate when requests for sample deposition are declined.

Funding agencies should require that researchers' grant proposals include sample archival procedures and that budgets include curation fees. Critics might argue that archiving will decrease the money available for other scientific endeavours. In our view, a sample stewardship plan should be viewed as equivalent to budget-line items for data archiving, publishing fees or institutional overhead costs that support other essential components of the research workflow.

We strongly recommend against setting universal fees. Samples will vary widely in nature and size, from kilogram-scale samples to micrograms of separated minerals. So the cost to museums will likewise depend on institutional resources and expertise. However, we have confidence that museums, working with funding agencies and researchers, will ensure that fees are self-regulating.

Collections of palaeontological samples provide an analogue for the practices needed. They also show that large-scale archiving is possible.

The Invertebrate Paleontology Division of the Yale Peabody Museum of Natural History in New Haven, Connecticut, for instance, holds about 4.5 million specimens and takes in more than 2,000 samples a year, on average. As well as its curatorial researchers, the division is supported by two full-time staff members, one of whom handles the new acquisitions.

We estimate that roughly 200,000 new sedimentary geochemical samples are analysed each year. We therefore reiterate that curation fees – even modest ones – should be incorporated into the budgets of research-grant proposals. Regardless of the current availability of space and curatorial support in individual museums, extra funds will be needed to meet the demand for archiving sedimentary geochemical samples.

The guidelines we offer will need to be discussed and revised by the community and institutions. Nonetheless, all best practices must rest on a shared commitment – to ensure that scientific data are not divorced from scientific samples.

### The authors

**Noah Planavsky** is an associate professor in geochemistry at Yale University, New Haven, Connecticut, USA, and assistant curator of mineralogy and meteoritics at the Yale Peabody Museum of Natural History. **Ashleigh Hood** is a lecturer in sedimentology at the University of Melbourne, Parkville, Australia. **Lidya Tarhan** is an assistant professor in palaeontology and sedimentology at Yale University, New Haven, Connecticut, USA. **Shuzhong Shen** is a professor in palaeontology and stratigraphy at Nanjing University, Nanjing, China. **Kirk Johnson** is the director of the Smithsonian's National Museum of Natural History, Washington DC, USA. e-mail: [noah.planavsky@yale.edu](mailto:noah.planavsky@yale.edu)

1. Cavosie, A. J., Valley, J. W. & Wilde, S. A. *Earth Planet. Sci. Lett.* **235**, 663–681 (2005).
2. Farquhar, J., Bao, H. & Thiemens, M. *Science* **289**, 756–758 (2000).
3. Hoffman, P. F. et al. *Sci. Adv.* **3**, e1600983 (2017).
4. Fan, J.-X. et al. *Science* **367**, 272–277 (2020).
5. Sun, Y. D. et al. *Science* **338**, 366–370 (2012).
6. Nutman, A. P., Bennett, V. C., Friend, C. R. L., Van Kranendonk, M. J. & Chivas, A. R. *Nature* **537**, 535–538 (2016).
7. Allwood, A. C., Rosing, M. T., Flannery, D. T., Hurowitz, J. A. & Heirweh, C. M. *Nature* **563**, 241–244 (2018).
8. Cole, D. B. et al. *Geobiology* **18**, 260–281 (2020).
9. Steadman, J. A. et al. *Precamb. Res.* **340**, 105722 (2020).
10. Luo, G. et al. *Sci. Adv.* **2**, e1600134 (2016).
11. Lyons, T. W., Reinhard, C. T. & Planavsky, N. J. *Nature* **506**, 307–315 (2014).
12. Grasby, S. E., Them, T. R., Chen, Z., Yin, R. & Ardakani, O. H. *Earth-Sci. Rev.* **196**, 102880 (2019).
13. Percival, L. M. E. et al. *Am. J. Sci.* **318**, 799–860 (2018).
14. Foster, G. L. et al. *Chem. Geol.* **358**, 1–14 (2013).
15. Henehan, M. J. et al. *Proc. Natl Acad. Sci. USA* **116**, 22500–22504 (2019).
16. French, K. L. et al. *Proc. Natl Acad. Sci. USA* **112**, 5915–5920 (2015).
17. Albut, G. et al. *Geochim. Cosmochim. Acta* **265**, 330–353 (2019).
18. Isson, T. T. et al. *Geobiology* **16**, 341–352 (2018).