

Digitizing the vast 'dark data' in museum fossil collections

September 17 2018, by Charles Marshall



Researchers must travel to visit non-digitized specimens in person, not knowing what they will find – if they're even aware of their existence. Credit: Smithsonian Institution, [CC BY-NC-SA](#)

The great museums of the world harbor a secret: They're home to [millions upon millions of natural history specimens](#) that almost never see the light of day. They lie hidden from public view, typically housed behind or above the public exhibit halls, or in off-site buildings.

What's on public display represents only the tiniest fraction of the wealth of knowledge under the stewardship of each [museum](#). Beyond fossils, museums are the repositories for what we know of the world's living species, as well as much of our own cultural history.

For paleontologists, biologists and anthropologists, museums are like the historians' archives. And like most archives – think of those housed in the Vatican or in the Library of Congress – each museum typically holds many unique specimens, the only data we have on the species they represent.

The uniqueness of each museum collection means that scientists routinely make pilgrimages worldwide to visit them. It also means that the loss of a collection, as in the recent heart-wrenching fire in Rio de Janeiro, represents an irreplaceable loss of knowledge. It's akin to the loss of family history when a family elder passes away. [In Rio, these losses included](#) one-of-a-kind dinosaurs, perhaps the oldest human remains ever found in South America, and the only audio recordings and documents of indigenous languages, including many that no longer have native speakers. Things we once knew, we know no longer; things we might have known can no longer be known.

But now digital technologies – including the internet, interoperable databases and rapid imaging techniques – make it possible to electronically aggregate museum data. Researchers, including a multi-institutional team I am leading, are laying the foundation for the coherent use of these millions of specimens. Across the globe, teams are working to bring these "dark data" – currently inaccessible via the web – into the digital light.



High-resolution photos are an important part of the digitization process. Credit: Smithsonian Institution, [CC BY-NC-SA](#)

What's hidden away in drawers and boxes

Paleontologists often describe the fossil record as incomplete. But for some groups the fossil record can be remarkably good. In many cases, there are plenty of previously collected specimens in museums to help scientists answer their research questions. The issue is how accessible – or not – they are.

The sheer size of fossil collections, and the fact that most of their contents were collected before the invention of computers and the internet, make it very difficult to aggregate the data associated with museum specimens. From a digital point of view, most of the world's fossil collections represent "dark data." The fact that large portions of existing museum collections are not computerized also means that [lost treasures are waiting to be rediscovered](#) within museums themselves.

With the vision and investment of funding agencies such as the National Science Foundation (NSF) in the United States, numerous museums are collaborating to digitally bring together their data from key parts of the [fossil record](#). The [University of California Museum of Paleontology](#) at Berkeley, where I work, is [one of 10 museums](#) now aggregating some of their fossil data. Together through our digitized collections, we are working to understand how major environmental changes have affected marine ecosystems on the eastern coast of the Pacific Ocean, from Chile to Alaska, over the last 66 million years.

The digitization process itself includes adding the specimen's collection data into the museum computer system if it hasn't already been entered: its species identification, where it was found, and the age of the rocks it was found in. Then, we digitize the geographic location of where the specimen was collected, and take digital images that can be accessed via the web.

The [Integrated Digitized Biocollections](#) (iDigBio) site hosts all the major museum digitization efforts in the United States funded by the current NSF initiative that began in 2011.



Team members entering information about each fossil into a centralized database. Credit: Smithsonian Institution, [CC BY-NC-SA](#)

Significantly, the cost of digitally aggregating the fossil data online, including the tens of thousands of images, is remarkably small compared with the cost it took to collect the fossils in the first place. It's also less than the expense of maintaining the physical security and accessibility of these priceless resources – a cost that those supposed to be [responsible for the museum in Rio apparently were not](#) willing to cover, with disastrous consequences.

Digitized data can help answer research questions

Our group, called EPICC for [Eastern Pacific Invertebrate Communities of the Cenozoic](#), [quantified just how much "dark data"](#) are present in our joint collections. We found that our 10 museums contain fossils from 23 times the number of collection sites in California, Oregon and Washington than are currently documented in a leading online electronic

database of the paleontological scientific literature, [the Paleobiology Database](#).

EPICC is using our newly digitized data to piece together a richer understanding of past ecological response to environmental change. We want to test ideas relevant to long- and short-term climate change. How did life recover from the mass extinction that wiped out the dinosaurs? How did changes in ocean temperature drive marine ecosystem change, including those associated with the isolation of the cooler Pacific Ocean from the warmer Caribbean Sea when the land bridge at Panama first formed?

To answer these questions, all the relevant fossil data, drawn from many museums, needs to be easily accessible online to enable large-scale synthesis of those data. Digitization enables paleontologists to see the forest as a whole, rather than just as a myriad number of individual trees.

In some cases – such as records of past languages or the collection data associated with individual specimens – digital records help protect these invaluable resources. But, typically, the actual specimens remain crucial to understanding past change. Researchers often still need to make key measurements directly on the specimens themselves.



Once digitized, information about a fossil is available worldwide, while the specimen itself remains available to visiting researchers to make crucial observations or measurements. Credit: Deniz Durmis, contract photographer for the Natural History Museum of Los Angeles County, [CC BY-NC-SA](#)

For example, Berkeley Ph.D. student Emily Orzechowski is using specimens being aggregated by the EPICC project to test the idea that the ocean off the Californian coast will become cooler with global climate change. Climate models predict increased global warming will lead to stronger winds down the coast, which will increase the coastal upwelling that brings frigid waters from the deep ocean to the surface – the cause of San Francisco's famous summer fogs.

The test she's using relies on mapping the distributions of huge numbers

of fossils. She's measuring subtle differences in the oxygen and carbon isotopes found in fossil clam and snail shells that date to the last interglacial period of Earth's history about 120,000 years ago, when the west coast was warmer than it is today. Access to the real-life fossils is crucial in this kind of research.

Understanding response to past change is not just restricted to fossils. For example, nearly a century ago the director of the Museum of Vertebrate Zoology, [Joseph Grinnell](#) at the University of California, Berkeley, undertook systematic collections of mammals and birds across California. Subsequently, the museum [re-surveyed those precise localities](#), discovering major changes in the distribution of many species, including [loss of many bird species](#) in the Mojave Desert.

A key aspect of this work has been comparison of the DNA from the almost hundred-year-old museum specimens with DNA of animals alive today. The comparison revealed serious fragmentation of populations, and led to the identification of genetic changes in response to environmental change. Having the specimens is crucial to this kind of project.

This digital revolution is not just restricted to fossils and paleontology. It pertains to all museums collections. Curators and researchers are enormously [excited by the power to be gained](#) as the [museum collections](#) of the world – from fossils to specimens from live-caught organisms – become accessible through the nascent digitization of our invaluable collections.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: Digitizing the vast 'dark data' in museum fossil collections (2018, September 17)
retrieved 1 May 2024 from

<https://phys.org/news/2018-09-digitizing-vast-dark-museum-fossil.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--