



Jack Tseng at the University at Buffalo in New York uses 3D scans to simulate stress patterns during biting, in skulls from fossils and extant species.

# THE FIGHT FOR CONTROL OVER VIRTUAL FOSSILS

*Palaeontologists have been urged to share 3D scans of fossils online, but a Nature analysis finds that few researchers do so.*

BY DYANI LEWIS

A decade ago, palaeontologist Jack Tseng set out on a treasure hunt. Not the typical boots and pick-axe affair you might imagine, but one that is relatively common in his field. From his base at the Natural History Museum of Los Angeles County in California, Tseng visited museums around the world to examine the skulls of carnivores in their collections. And whenever he encountered one, he asked whether he could take away 3D scans of the specimen. Tseng's own institution housed skeletons from striped hyenas, cheetahs, jackals, aardwolves and mongooses, as well as skulls from extinct hyenas and dogs. But Tseng,

then a doctoral student, needed even more exotic fossils for his research on how carnivores evolved the ability to crush bone. "I was looking for exceptionally complete skulls," he says.

And so, he travelled. To New York, Washington DC, Beijing, London, Uppsala, ticking off items on his palaeontological shopping list as he went. One place Tseng did not need to visit was the National Museum of Natural Sciences in Madrid, even though it holds an unusually near-complete skull of a large extinct hyena. A carnivore specialist at the museum, Manuel Salesa, had already scanned the fossil and sent the data to Tseng directly.

Salesa's generosity left a lasting impression on Tseng, who now heads his own evolutionary-biology laboratory at the University at Buffalo in New York. He still travels to see far-flung collections, but increasingly relies on 'virtual fossils' for his studies. And when he has published his findings, he uploads any scans he has made to an online database, ready for other researchers to download. "It's the most obvious way to pay it forward," he says.

It's now common for palaeontologists to scan their fossils in 3D: not just to view their surfaces, but also to deduce internal structures using X-ray computed tomography (CT),

which can reveal the contours of a skeleton embedded in rock, the dimensions of a skull's braincase or the internal pathology of a stegosaurus's bone tumour. Researchers also share 3D models of excavation sites and footprints, generated from 2D photographs using a technique called photogrammetry. Tseng, who now builds biomechanical models of animals' chewing mechanics, says that virtual fossils are indispensable for his work.

With this trend, terabytes of images — such as digital facsimiles of Neanderthal teeth, sabretooth-cat skulls and pteranodon wing-bones — are filling online repositories. But not everyone embraces Tseng's idea of paying it forward. Despite nearly two decades of exhortations to share 3D data, only around one-third of the most popular palaeontology studies involving 3D imaging over the past two years uploaded their scans online, according to an analysis by *Nature* for this article (see 'Who shares 3D scans?'; data at [go.nature.com/virtualfossils](http://go.nature.com/virtualfossils)). Even so, more than half of the non-sharers said that they supported open sharing of data — in principle.

Fears of relinquishing control are rife. Researchers are loath to forfeit first dibs on potentially years' worth of publications describing specimens that they collected or were first to scan. Museums are apprehensive about loosening their grasp on data generated from fossils in their care, sometimes citing loss of income streams — or simply the desire to control how the research community uses their specimens.

In one sense, the field of palaeontology — its researchers, professional societies, museums and journals — is just another academic discipline grappling with the fast-moving norms of the open-science movement. Compared with other research fields, however, its difficulties are particularly acute: fossils are often rare or entirely unique physical specimens, closely guarded by scientists and museums, which makes their 3D data unusually valuable.

Times are changing. In the past year alone, multiple museums have rewritten policies on the sharing of 3D fossil data, and professional societies are formalizing statements on what is expected from palaeontologists when it comes to sharing — although not everyone agrees on what that should be. "We're at this transition point where the technology is there and now people's attitudes have to catch up," says Anjali Goswami, a palaeobiologist at the Natural History Museum in London who supports open access in science.

### PALAEO-PLATFORMS

The world's most popular website for virtual fossils, MorphoSource, holds in excess of 62,000 data sets from more than 7,300 species. "The larger the reservoir of data that people have access to, the more sophisticated and powerful analyses they can do," says its creator, Doug Boyer. Just as important is the idea that access breeds integrity. As the number of researchers with access to data increases, "the

more repeatable the science that's generated from the data becomes," he says.

Boyer had the idea for a community repository nine years ago, when, as a postdoc studying evolutionary biology at the University of Helsinki, he was asked to build a platform to archive his group's 3D data sets and computational models. In 2012, he set up his own evolutionary-anthropology group at Duke University in Durham, North Carolina, and pushed to build a site everyone could use. Duke bankrolled the initial development, and in 2013, MorphoSource was soft-launched: functional, but in need of data with which to fill its digital vault.

That changed in 2015, when word of Boyer's passion project reached Lee Berger, a palaeontologist at the University of The Witwatersrand (Wits) in Johannesburg, South Africa. The two agreed that MorphoSource should host data for the soon-to-be-published remains of a newly discovered species of early human that Berger's team had unearthed near Johannesburg. When *Homo naledi* was announced

*"We're at this transition point where the technology is there and now people's attitudes have to catch up."*

to the world in September that year<sup>1</sup>, data for 86 virtual specimens were simultaneously unlocked on MorphoSource.

The trickle of contributions became a steady stream almost overnight. The site now employs three full-time developers, and has secured more than US\$2 million in funding from the US National Science Foundation (NSF) and Duke to support operations until at least 2025. It is free to browse and to download and upload files, although high-volume users are asked to contribute to the site's storage costs.

While Boyer was putting the finishing touches to MorphoSource, Goswami, then at University College London, was driven by frustration to build her own platform — Phenome10K — to house a stockpile of virtual skulls. "We were generating these huge amounts of data and then only doing one or two things with them," she says. "Then they just sit on a hard drive for the rest of eternity until we forget what's on that hard drive." Phenome10K, which is also free to use, now houses more than 2,200 of her group's surface scans, one-quarter of them fossils, as well as surface scans and CT data from other academics.

Smaller repositories are also cropping up. Some host substantial collections — the Smithsonian Institution in Washington DC, for instance, has a public-facing data portal that it is upgrading this year to provide access to large CT data files — whereas others are bare-bones websites with only a handful of specimens.

General-purpose research repositories, such as Figshare, Dryad or Zenodo, are other popular choices. (Figshare and *Nature* have a common owner, the Holtzbrinck Publishing Group.)

In 2017, Boyer, Goswami and dozens of palaeontologists, anatomists, anthropologists and other purveyors of digital specimens came together to recommend best-practice guidelines for sharing digital morphology data<sup>2</sup>. Sharing data in repositories is crucial, because it is not enough for researchers to say in their papers that data are "available on request", argues palaeobiologist Phil Donoghue from the University of Bristol, UK, the lead author on the recommendations. Some scientists simply don't respond to requests, says Donoghue — and if they move on to other institutions, the data can be lost.

### MUSEUMS IN CHARGE

But many bureaucratic barriers dissuade researchers from openly sharing fossil scans online. A big problem, says Tseng, is that fossils are usually housed in museums — which often keep a tight rein on their specimens and the scans made of them. Major US and European museums, for instance, such as the Field Museum in Chicago, Illinois, and the Natural History Museum in London, insist that they are assigned ownership of data from scans of their fossils, in part so that they can track how their collections are being used<sup>2</sup>. Posting these data to an online repository, or even passing them to a colleague, without the museum's explicit permission contravenes these agreements.

Still, there are many examples of museums granting permission for scans to be shared. MorphoSource's vast cyber-crypt houses the skull of an extinct sea turtle from the Natural History Museum in London, extinct devil frogs from the Field Museum and pterosaur vertebrae from the American Museum of Natural History, for instance. Multiple museums told *Nature* that they have been formalizing ad-hoc data-sharing arrangements over the past year.

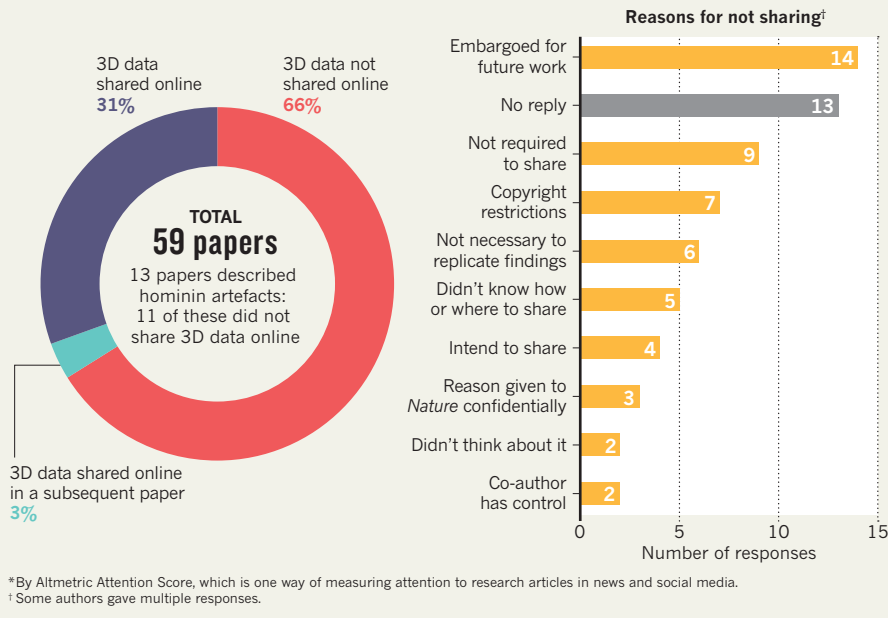
But museums still often want to control who can download their data. Wits allows scientists to upload 3D scans to MorphoSource, but a university committee has to grant access to the highest-resolution data; and the committee might deny this if it encroaches on a Wits student's work, says Bernhard Zipfel, who is the curator of fossils and rock collections at Wits and is involved in decisions about access. "We treat these raw data like we do the original fossils," says Zipfel. "We don't let these data freely out of our hands without due process."

And although the Field Museum doesn't want to restrict distribution unnecessarily, it does want users to sign permission forms demonstrating that they understand how to credit the data they're using, says Bill Simpson, the museum's head of geological collections and collections manager for fossil vertebrates.

At MorphoSource, Boyer allows museum curators to build in these restrictions. They can choose to release data only after a user requests permission, say. (Museums cannot yet

## WHO SHARES 3D SCANS?

*Nature* analysed the 200 most popular\* palaeontology papers across 2017 and 2018. Of 122 that described 3D objects, just under half — 59 papers — collected 3D imaging data. But only 31% of these shared those data online. See [go.nature.com/virtualfossils](http://go.nature.com/virtualfossils) for more-detailed results.



charge for their data sets, although Boyer was not ruled this out in future.) Meeting museums halfway like this is important, “particularly as we transition”, Boyer says. More than half of the data on the site are open access; the rest require approval by the data set’s owner before they can be downloaded.

To some advocates of open sharing, these access controls are unnecessary barriers to speedy science. “What’s wrong with multiple teams working on the same data and, indeed, the same question at the same time?” says Donoghue.

But other palaeontologists say that they don’t want anyone else working with their scans, even after publication. “It is not clear that academics who have garnered the resources to acquire these data now suffer a moral imperative to share the raw scan data,” says vertebrate palaeontologist Michael Caldwell at the University of Alberta in Edmonton, Canada. Other researchers, he says, can always borrow specimens themselves and create their own CT scans. In *Nature*’s analysis, the most frequent explanation given by those who hadn’t shared 3D data online was that they didn’t want to jeopardize their ongoing research.

Caldwell didn’t share 3D CT scans alongside a paper<sup>3</sup> last year that described the first-known example of a fossilized snake embryo, which was preserved in amber; although, in that case, he says, access to the data was controlled by his Chinese co-authors. Xing Lida, the lead author on the study, told *Nature* that the CT-scan data will not be published in a repository — although they will be made available to researchers on request — because the private citizen who found and donated the specimen to the Dexu Institute of Palaeontology in

Chaozhou, China, plans to make 3D-printed metal replicas for the museum to sell.

### REVENUES DISRUPTED

Abiding by museum policies is especially important in collections drawn from the fertile fossil beds of Africa. Throughout the continent’s poorer nations, museums bolster their meagre budgets by selling replica casts of important specimens to academics and other museums around the world. The casting lab at the National Museums of Kenya in Nairobi “provides an important service and at the same time supports a lot of Kenyan families”, says palaeoanthropologist John Kappelman at the University of Texas at Austin. “I’d feel terrible if somebody started giving away data and disrupting the revenue stream from a casting programme.” (Museums all over the world also charge researchers ‘bench’ fees to come and work on fossils.)

At Wits, Zipfel says he is particularly angered by researchers who distribute virtual fossils without proper acknowledgement of the countries and institutions that own them. That, he says, reeks of colonialism. “This is essentially white people from other countries swaggering around using our heritage to further their own careers,” he says.

It’s not impossible to devise workable data-sharing arrangements, says Jean-Jacques Hublin, a palaeoanthropologist at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. In 2013, he negotiated one solution with the Ditsong Museums of South Africa, when he created a repository on his institute’s servers to share surface models — and, on approval by the curator, raw CT data sets — of hominin fossils in the museums’ collection<sup>4</sup>.

Similarly, Kappelman gained approval from Ethiopia’s government and its National Museum in Addis Ababa to make models freely available for others to 3D print a small selection of bones from ‘Lucy’, the famous 3.2-million-year-old *Australopithecus afarensis* skeleton discovered in Ethiopia in 1974. He set up the site eLucy.org in 2016 for a study in which he and his colleagues argued from bone analysis that Lucy probably died falling from a tree<sup>5</sup>. He has now proposed that his university set up a site for Ethiopian authorities to sell access to higher-resolution 3D scans. Yonas Desta, the director general of Ethiopia’s Authority for Research and Conservation of Cultural Heritage, which controls access to Ethiopia’s fossils, says that he is considering the proposal.

“Sharing doesn’t mean it has to be absolutely scot-free,” Kappelman says. Researchers can use grant money to pay to acquire scans. “That’s what I see as the model coming down the line.”

But open-access advocates are dead set against that idea. “There should be no bar to people accessing the data,” says Donoghue.

### HUMAN ORIGINS

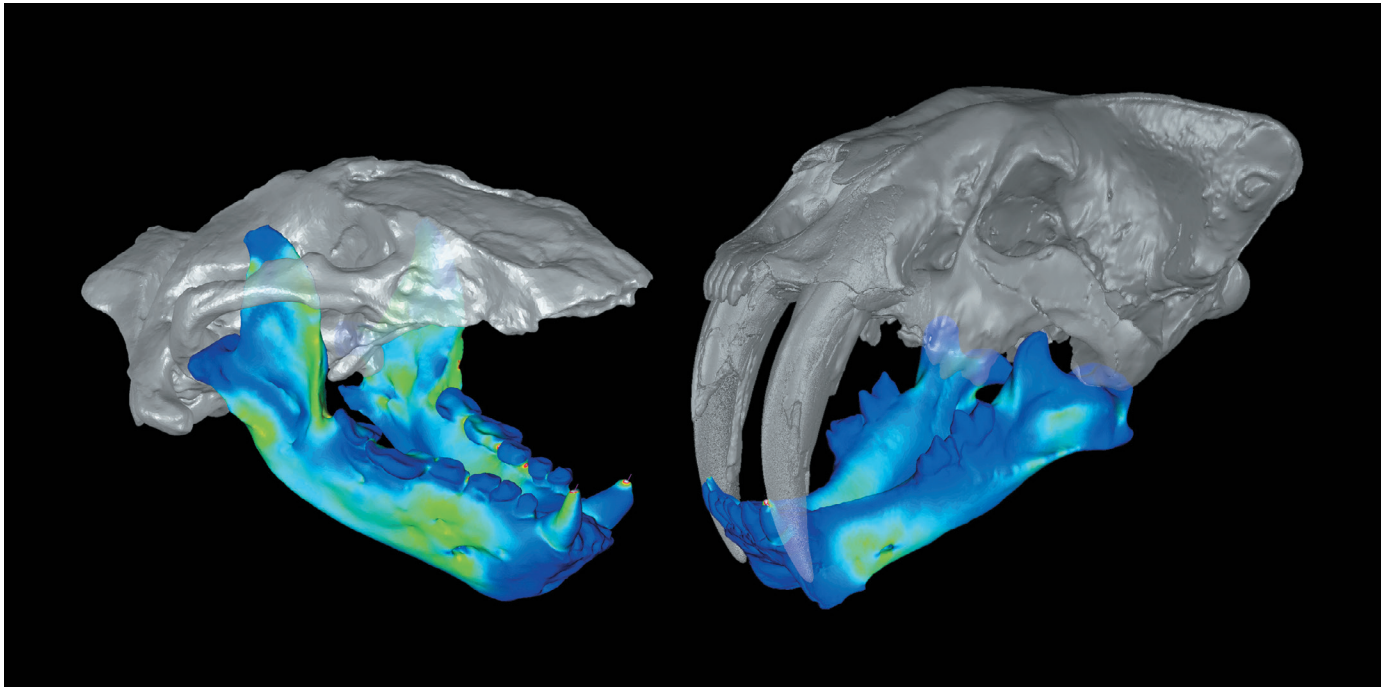
The restrictions of museum policies are felt most keenly by those working in the high-stakes field of human-origins research. Palaeoanthropology is notoriously secretive: in some cases, researchers have been denied access to precious fossil specimens for years or even decades. One example is a 7-million-year-old thigh bone discovered in Chad in 2001 that is said to belong to a species called *Sahelanthropus tchadensis*, which is claimed to be the earliest-known hominin on the basis of skull analysis. The bone could establish whether this species walked on two feet, but it has not yet been described in detail in a scientific publication.

Digital specimens are often closed off, too. In *Nature*’s analysis, scans were not shared online for 11 out of 13 papers that involved hominin specimens. Corresponding authors for nearly half of these studies did not respond to *Nature*’s query — but two who did blamed museums’ copyright policies.

The sharing of human specimens is always a sensitive issue, because some remains can be traced to living indigenous communities that do not want scans to be made public. But in other cases, says Donoghue, researchers simply monopolize rare fossils because they can. It’s a case, he says, of “I have access to the fossil and you don’t, so that’s what I’m going to build my career on”.

Specimens from living species or other non-hominin artefacts are more likely to be shared online, Boyer says. “There really is a divide between curators that focus on extant remains versus those who work on fossils,” says Boyer. “The mammalogy and the herpetology departments are very eager for open access, and the palaeontology departments are very cautious about it.”

The NSF-funded project oVert, for example,



Fossil scans show how an extinct marine mammal (*Kolponomos*, left) bit in a similar manner to an extinct sabretooth cat (*Smilodon*, right).

is a Herculean effort, launched in 2017, to make CT scans of more than 20,000 vertebrate specimens held in 16 US museum and university collections. The data will be openly available on MorphoSource. No project of that scale for palaeontology has yet been proposed.

#### CHANGE FROM THE TOP

Some researchers say it is up to funding agencies, journals and the professional societies that publish them to push researchers to share data openly.

“I think journal policies can be really powerful,” says Andy Farke, a dinosaur specialist and curator at the Raymond M. Alf Museum of Paleontology in Claremont, California. Increasingly strict journal policies, he notes, have helped to stamp out the practice, now considered unethical, of publishing research on fossils in private collections.

In one case in 2016, a journal did change a museum’s policy. Lynn Copes and Lynn Lucas, two PhD students then at Arizona State University in Tempe, had collected more than 400 micro-CT (very high resolution) scans of primate skulls in the collection of the Harvard Museum of Comparative Zoology for use in their dissertations, and uploaded their data to MorphoSource. But the museum, based in Cambridge, Massachusetts, was reluctant to allow the data to be shared openly; its policy required approval for all third-party uses. Then a journal, *Scientific Data*, rejected a paper on the material because, editors said, there was no good justification for access to the data to be restricted. “We came back to the museum and said, ‘Look, this is good publicity for the museum, this is a great resource. But it’s not going to be published if you insist on sticking with this more restrictive policy,’” says Boyer.

The museum decided to change its policy, and the paper was published<sup>6</sup>. (*Scientific Data* is published by Springer Nature, which also publishes the Nature family of journals; these typically prefer large data sets to be deposited online, but do not mandate it. *Nature’s* news team is editorially independent of its publisher.)

The Harvard museum now encourages researchers who scan specimens — fossils included — to upload their data to MorphoSource, says its director James Hanken. “They do a much better job of making these data sets available than we could,” he says.

Some journals mandate open sharing, but do not always enforce their policy — and are rarely explicit about how extensively 3D imaging data should be shared, sometimes saying simply that they follow community standards. In 2017, for instance, a paper<sup>7</sup> that contained CT scans and 3D reconstructions of what its authors contended might be the oldest-known hominin published flat images and measurements, but not the scans themselves, stating only that: “All relevant data are within the paper and its Supporting Information files.” The report appeared in the journal *PLoS ONE*, which mandates sharing of data in a repository. But corresponding author Madelaine Böhme from the University of Tübingen in Germany says that the journal did not require sharing of the 3D data. A spokesperson for PLOS pointed to the journal’s data policy, which states that “authors do not need to submit the raw data collected during an investigation if the standard in the field is to share data that have been processed”.

In Farke’s view, spreadsheets of measurements from scanned specimens do not provide enough information to verify CT data. Researchers need to go back to the original

scans, he says, because they might interpret them differently.

In the absence of clear community standards, many journals, funders and societies do not go beyond ‘encouraging’ data sharing. Some require data-availability statements — but, again, *Nature’s* analysis found examples in which papers had been published in such journals without data-availability statements. The NSF requires that researchers outline how data will be managed and published in grant applications. But, in practice, those requirements lack teeth, Tseng says. “If the funding agencies and journals actually enforced what they recommend and encourage people to do, we would be making a lot more progress than we’re seeing right now,” he says.

Ultimately, says Tseng, sharing 3D images online has to be something that palaeontologists want to do. When reluctant colleagues argue with him about why they should share their data, he points to citations. “In the world of academic promotions, that is a real currency,” he says. Above all, however, he wants his colleagues to see his point of view: that “open sharing is the best model with which to accelerate the pace of science”.

**Dyani Lewis** is a science journalist in Melbourne, Australia.

- Berger, L. R. *et al.* *eLife* **4**, e09560 (2015).
- Davies, T. G. *et al.* *Proc. R. Soc. B* **284**, 20170194 (2017).
- Xing, L. *et al.* *Sci. Adv.* **4**, eaat5042 (2018).
- Skinner, M. M., Kivell, T. L., Potze, S. & Hublin, J.-J. *J. Hum. Evol.* **64**, 434–447 (2013).
- Kappelman, J. *et al.* *Nature* **537**, 503–507 (2016).
- Copes, L. E., Lucas, L. M., Thostenson, J. O., Hoekstra, H. E. & Boyer, D. M. *Sci. Data* **3**, 160001 (2016).
- Fuss, J., Spassov, N., Begun, D. R. & Böhme, M. *PLoS ONE* **12**, e0177127 (2017).