UC SANTA BARBARA

THE Current

March 26, 2014
Shelly Leachman

Here Today, Gone to Meta

A collection of etchings is printed on paper that starts fading on its first exposure to light. Embedded in those pages is a floppy disk (now defunct) housing a single poem programmed to self-encrypt and vanish after one use.

Such were the curious components of Agrippa, an influential and controversial book, published in 1992, that saw an abstract artist, a cyberpunk novelist and a publisher jaded by commercialism collaborate on what remains an intriguing examination of decay.

It is also an apt metaphor for an issue now vexing universities and research institutions everywhere: data curation. With technology advancing at warp speed and data proliferating apace, can the scientific and scholarly output survive into the future?

“Agrippa was designed specifically to challenge the idea of curation and preservation,” said UC Santa Barbara English professor Alan Liu, who has studied the book extensively and created a related digital archive. “You can either collect and preserve the object but never access it, or you can access it and read it but never be able to preserve it. That’s your perfect paradox for the problem of data curation.”

Enter Data Curation @ UCSB, an effort to address that very problem on campus, where new data is being created around the clock in disciplines from marine science to Middle Eastern studies. Now in its second year, the pilot project is using insights from a launch-year faculty survey to shape the inaugural iteration of a data curation

service to be based at the UCSB Library.

The endeavor is a collaboration of the library and UCSB's Earth Research Institute (ERI), with support from the Office of Research and the Executive Vice Chancellor. Its ultimate goal is to ensure that data created on campus stays on campus and, more importantly, can be discovered and understood now and for generations to come.

"That's what libraries do and have always done — help people organize and manage information," said university librarian Denise Stephens. "There is a real need, and a real opportunity, for the campus to leverage its experts and resources in the library and across various departments to ensure that the scholarship of our researchers persists and is preserved. It's really exciting, and it's an example of how research libraries in particular are becoming something different from what they used to be, while still doing all the things we have always done. There is no departure from our values, our purpose or the eternal mission of a library, but the flavor of the work is changing."

The 'Big Shift'

When it comes to data — from collection to curation — there are changes afoot for libraries and researchers alike. As the former grapple with issues of storage, organization and technological survival, the latter are faced with evolving expectations of the research community.

"There is now an expectation that if you are gathering data, it will be available," said Greg Janée, a digital library research specialist at ERI and Data Curation @ UCSB project lead. "It will be available right now, it will always be available and it will be available in a digital form where I can use it. And I will be able to get to it from the paper where I'm reading about your data. That's a big shift."

Those presumptions are increasingly being made by funding agencies as well. The suggestions that grant-seeking researchers create data plans for their projects are now becoming mandates that data be put into a specific format and submitted to a particular place. That's according to James Frew, an associate professor at UCSB's Bren School of Environmental Science & Management, who said the new requirements are ratcheting up the pressure to get curation plans in place.

"The iron is hot," said Frew, an environmental informatics expert and Janée's research partner on the curation project. "For people in the academic community,

the notion that we have to do something about this is becoming more widespread. I don’t have to explain digital curation so much anymore. It’s all part of the university’s work product. We traditionally tended to think of it as papers but that’s just not true anymore. They want the data too.”

As demand for the data behind every project and paper continues to grow — and datasets are more frequently seen as citation-worthy in their own right — the conundrum of where and how to keep it all is ever-more complex. Scholarly data today needs more than just a permanent home; it needs to come with clear directions for finding it and a set of keys dangling outside the door for anyone who wants to visit.

The Domesday Effect

As if matters weren’t complicated enough, to retain its research value the data must remain accessible, for centuries, like Darwin’s original manuscripts.

Or the Domesday Book.

An 11th-century land survey of Britain conducted at the behest of King William I in 1085, the hand-scripted tome survives to this day in The National Archives in Kew, England. Meanwhile, a 1986 Domesday reboot for the then-dawning digital age — a multimedia version stored on laserdiscs — was dead only 15 years later, when the computers capable of reading it were rendered obsolete.

Therein lies another illustration of the obstacle at hand.

“It’s ironic that digital information is in bits, which are abstract and last forever yet are much more fragile than paper,” said Janée. “The technology, our current hard drives, are simply not reliable enough. So preserving the bits is unsolved, which is a little scary because everything is digital now. We’re concerned with preserving campus research outputs, but this is a much larger problem. We really are talking about preserving our cultural heritage in the broadest sense.”

That’s So Meta

If curating data for a researcher is akin to curating artifacts for a museum, metadata can be thought of as the book that accompanies the exhibit. Having the who’s, what’s, when’s and where’s are as essential to understanding data as they are to appreciating art.

“In addition to deciding what’s going to be preserved, a big part of data curation is describing the data,” said Margaret O’Brien, an information manager for the UCSB-based Santa Barbara Coastal Long-Term Ecological Research project, where she is engaged in such work daily. “The better, more complete the metadata, the easier it will be to find that data in the future. If you don’t put much more than a name and a date on something, you don’t have much context. You might not even be able to tell whether you’re looking at observational data or an experiment, which is an important distinction.

“Even with research that is ‘born digital,’ the curation is essential,” O’Brien added. “If you don’t add metadata, you might as well put it on a piece of paper and shove it in a shoebox because no one is ever going to figure out what it is — maybe not even the original scientist.”

In that sense, a data curator also functions as something of an editor, putting proper words in place of nicknames that are commonplace in labs and replacing any jargon with terms more likely to be understood by a broad audience. Much like a writer so attached to a story to be blind to its flaws, scientists can be too far inside their research to discern the most effective way to describe the data for outsiders. Objective eyes, like O’Brien’s, can make all the difference.

A Moving Target

The library-based service that UCSB hopes to launch by June would provide faculty and other campus researchers exactly that brand of counsel, by way of specialists trained in all things curation who are also experts in specific subjects. Recruitment is already underway for a spatial data specialist, according to Stephens; parallel positions will eventually be created in the areas of humanities, social science and science.

“We want to make sure that we bring credible expert services, which means having enough subject matter knowledge to be a trusted partner to researchers, because this kind of service is intimate,” Stephens said. “A trusted partner understands what the data issues are, how a researcher uses the data they create, the problems they’re trying to solve and the way they actually conduct their research. The future will call all the way across the research library for us to be collaborators and partners in the intellectual efforts of others.”

Dwight Reynolds will welcome the help. The religious studies professor's extensive research on an Egyptian oral epic poem includes hundreds of cassette tapes and transcriptions and volumes of handwritten field notes from multiple visits to the Egyptian village where he witnessed the tradition firsthand. With support from a small grant, he began building a digital archive for his collection. But he's got a lot further to go.

"It crosses your mind: If I get hit by a bus, what happens to all of this?" asked Reynolds, whose grant required a pledge from the library to preserve the physical materials. (The digital archive's final resting place, so far, is undetermined.) "Anybody who does a collection of any sort is very concerned about its permanent home and that it be available. In the long run it makes no sense whatsoever to have stuff sitting in departmental offices. This problem is going to expand 100-fold over the next few years and the need will grow exponentially. The more groundwork that can be done now, the better shape we'll be in."

In short, the challenge is massive. Tackling it is imperative, which is why the number of stakeholders across campus and throughout academia are proliferating like new data. The UC's California Digital Library now has a separate curation center that recently launched a systemwide repository for digital content. The UCSB library will soon unveil its own digital repository that is likely to play a future role in the planned campus curation service.

Action on the task is not in short supply, though technology may forever make the final solution a bit of a moving target.

"Just the raw preservation will take a huge institutional effort way beyond building a building full of shelves — keeping bits alive is a much higher-tech problem than keeping books alive," said Frew. "Then there's the problem of making the stuff intelligible over a long period of time. Culture guarantees we'll be able to read English in 300 years, but nothing guarantees we'll be able to read a CD-rom in even 10 years. It's not just copying the bits but dragging along all the documentation to make sure you know what the bits mean. It's massive."

About UC Santa Barbara

The University of California, Santa Barbara is a leading research institution that also provides a comprehensive liberal arts learning experience. Our academic community

of faculty, students, and staff is characterized by a culture of interdisciplinary collaboration that is responsive to the needs of our multicultural and global society. All of this takes place within a living and learning environment like no other, as we draw inspiration from the beauty and resources of our extraordinary location at the edge of the Pacific Ocean.