



KENNEALLY: In April, a forest fire descended Cape Town's Table Mountain, quickly reaching the University of Cape Town. Historic buildings on campus fell to the flames, including Jagger Library, home to rare collections of South African books and other literature, such as anti-apartheid pamphlets. Because some of the collection was digitized, however, losses to the UCT archives are thought to be limited.

Welcome to Copyright Clearance Center's podcast series. I'm Christopher Kenneally for Velocity of Content. University libraries around the world provide safe and secure homes to vast stores of published and unpublished materials in physical and digital form, from books and journals to illustrations and photographs. Digitized copies as well as data pose important and invisible threats to this effort.

In 1999, Stanford University librarians pioneered a new field of digital-first archiving when they created LOCKSS, an acronym for Lots of Copies Keep Stuff Safe. Virtual organization CLOCKSS, or Controlled LOCKSS, now maintains authoritative versions of 43 million journal articles and 240,000 book titles, as well as a growing collection of supplementary materials and metadata information.

Alicia Wise was recently appointed executive director of CLOCKSS. Wise has long been active on access issues for research information. Most recently, she worked as a consultant in scholarly communications, advising libraries, funders, and publishers on sustainable strategies for navigating the rapidly changing information landscape. Alicia Wise joins me now from Windsor, England. Welcome to the program, Alicia.

WISE: Hello there, Chris, and thanks for the invitation to join you.

KENNEALLY: Well, we're looking forward to speaking with you about this, because it's a topic that we sort of take for granted. We think something is digitalized, it's online, it is there forever and easily found. Such is not the case, really. So CLOCKSS is there to help maintain archives of digitized materials for today and for the future. Tell us a little bit more about where CLOCKSS is today. I gave the very quick origin story, but here in 2021, it's a very substantial effort.

WISE: Yeah, CLOCKSS and similar organizations that are doing digital preservation are really important. And the very sobering context you placed us in with the recent fire at the University of Cape Town just brings that message home. We're all entrusted with preserving the scholarly record so that scientists and scholars today can contribute to it and also build on the shoulders of those giants who came before them. And research libraries and publishers themselves have long taken this responsibility for the scholarly record very, very seriously.



So CLOCKSS is one of the trusted digital archives for this material. We're a collaboration between world-leading research libraries and academic publishers. We're a community-governed and supported not-for-profit organization with an active global community of more than 300 libraries and 300 publishing outfits. Our mission is to keep the scholarly record safe, and we do this by providing a dark archiving service. The technology we use to deliver these services was, as you said, developed at the Stanford University libraries, with whom we continue to have a close and special relationship. We also mirror all of that content in 12 major academic institutions spread around the world – in Asia/Pacific, in Europe, and in North America. That approach gives us resilience to threats from potential technological, economic, environmental, or political challenges that might arise.

**KENNEALLY:** Alicia, why is a dark archive the way to approach this challenge? What makes a dark archive a good digital archive?

**WISE:** So a dark archive means we preserve the content entrusted to us in a highly secure environment which is not accessible to people, because unfortunately, people can often be a point of failure when it comes to content and data preserved in computing systems. The only time the content in the CLOCKSS archive becomes accessible is when a trigger event occurs. If any of the content that we're preserving has disappeared from the web or is about to disappear, then that is triggered and opened. It's made available open access under a Creative Commons license, so it remains available and accessible to everyone in the world. So that's our mission.

There are lots of different approaches to preserving content. We're not the only archive. And this is a really important point, perhaps, to communicate. CLOCKSS provides a complementary service to other organizations – for example, LOCKSS itself, which is widely used by libraries to preserve special collections or digitized collections. There are also archives that are held by national libraries, Portico. The important thing here is that there's a lot of resilience. The scholarly record is so valuable for our species and for our future that it's important for this content to be held in multiple formats, in multiple systems. There's lots of redundancy and resilience in the preservation safety net that underpins this content.

**KENNEALLY:** What type of trigger events have happened? Are you able to discuss an example of one, so we get a better sense of just how this really is so important?

**WISE:** Right. So occasionally, publishers will go out of business, and they may have a journal that has published articles by authors over a number of years, and there's not a new home for that content. All of those scholars' contributions would disappear if that content had not been preserved by the publisher, if it were not available to be triggered by an archive like CLOCKSS.



KENNEALLY: And there are other aspects of your work, your very new work, Alicia Wise, there at CLOCKSS. There is a disappearing journals project. All of this sounds very mysterious. (laughter) But a disappearing journals project that you are undertaking with the Directory of Open Access Journals, DOAJ, and other partners. Tell us about that. And in particular, open access – well, that seems again to be something we take for granted. But open access, particularly for small-scale APC-free journals, can be a challenge in this case.

WISE: Exactly right, Chris. There was a great paper published this year, in February, by Mikael Laakso and colleagues. It was in the *Journal of the Association for Information Science and Technology*. They reviewed a wide array of journals, and they found that over 174 open access journals had entirely vanished from the web between 2000 and 2019. These journals were in all discipline areas. They weren't just humanities or social sciences or STEM. They crossed those fields. And they were published in every geographic region of the world.

So it is unfortunately a challenge to be confident that all journals and books will remain accessible. It's not only open access journals, either, that are at risk of disappearing. Journals published by the full spectrum of publishers, and perhaps especially the long tail of smaller publishers, are at potential risk. For this reason, CLOCKSS is really proud to be part of a project team aiming to create and provide cost-effective and really workable preservation solutions for the long tail of publishers. The vision for the project predates my involvement with CLOCKSS, and I really want to give a shout-out and credit to my predecessor, Craig Van Dyck, who really championed this idea.

But the team is starting with content that's indexed in the Directory of Open Access Journals, and the DOAJ is a partner, and collectively, we're aiming to provide a central hub where preservation agencies like CLOCKSS and others will be able to harvest consistent metadata and full text from this long tail of publishers. We're starting by focusing on diamond open access journals. These are journals that don't charge any sort of article processing fee, and they may, in fact, rely very heavily on only voluntary effort from academics to be published. There might not be any formal publishing organization underpinning them at all.

In addition to CLOCKSS and DOAJ, the other project partners are the Internet Archive, the Keepers Registry, and the PKP Preservation Network, which provides the OJS journal publishing software which is used by so many of these long-tail publishers.

KENNEALLY: And the work you're undertaking there at CLOCKSS, Alicia, is by its nature a coalition of actors. They are predominantly libraries and publishers. Tell us how someone who's listening could become involved.



WISE: There are lots of ways to become involved. For publishers, we really invite you to support archiving agencies, whether that's CLOCKSS or another one, and to be visible and vocal in your support with your authors, your editors, your customers, your peers about the importance of digital preservation. And it's brilliant if your content is preserved in at least one, but preferably more of these services and that your archival deposits are confirmed and that they will remain accessible in the long term as part of the scholarly record.

For research libraries, there is a strong need to champion the importance of digital preservation within your organizations and to make sure that your libraries have clear digital preservation rules and responsibilities outlined. You can join and support an archiving organization, or preferably more than one, and investment in these services helps you ensure your researchers will have long-term, continuing access to the journals and books that they value and metadata. For example, CLOCKSS archives the DOIs that are registered in Crossref, so some of the essential infrastructure that enables scholarly content to interoperate.

Libraries can also add tremendous expertise to help preservation agencies identify important titles or classes of content that might be at risk of loss, help us to prioritize our collection efforts, and increasingly, libraries are themselves publishers. They are uniquely positioned to make sure that content published on their campuses are systematically archived as well and that systematic consideration of the long-term access issues are considered at the point of either creating that kind of content or licensing content from other publishers off campus.

And a way libraries can do that, for example, is by asking publishers about their digital preservation plans during content negotiations, or if they're entering into transformative open access agreements or interacting with publishers in different ways. And just generally, advocating for the importance of digital preservation and long-term access to content.

KENNEALLY: And in your work, Alicia Wise, your career has taken you along various paths, but this is a return to the beginning, in fact, because I understand that your first professional job was in a digital archive.

WISE: It was. When I was a newly minted archaeology PhD, I had the tremendous privilege of being the first member of staff at the Archaeology Data Service. It's a digital archive for archaeological data based at the University of York.

In archaeology, digital preservation is really important, because when we excavate sites, we destroy the evidence left behind by the people who previously lived in those locations.



So in my discipline, we take preservation really seriously. All of our notebooks, all of our drawings of the archaeological record, the artifacts, our interpretations of those, elaborate diagrams of stratigraphy – all of that is preserved, and increasingly, of course, it's all born digital. So digital archives and services like the Archaeology Data Service are absolutely central to scholarship in my field.

KENNEALLY: Alicia Wise, executive director of CLOCKSS, thanks so much for joining me on the program.

WISE: Thank you so much, Chris. Really lovely talking to you.

KENNEALLY: Our co-producer and recording engineer is Jeremy Brieske of Burst Marketing. You can subscribe to the program wherever you go for podcasts and follow us on Twitter and Facebook. I'm Christopher Kenneally. Thanks for listening and join us again soon for another Velocity of Content podcast from CCC.

END OF FILE