

The Knowledge Archive as Convergence: Challenges of Scale and Sustainability for Scholarly Publishers, Libraries and Museums

Alex W. Barker, Director, Museum of Art and Archaeology,
University of Missouri

Museums and libraries are similar beasts. Both hold and offer access to growing amounts of information about objects—paper, digital, or dimensional—whose value is directly related to their accessibility or findability. Both also play key roles in the archiving, presentation and preservation of knowledge, what has been called the knowledge archive. That similarity of role is not accidental; some of the founding theorists of the American museum movement—men like John Cotton Dana—came from a library background, and envisioned both kinds of institutions as fulfilling the same societal role with differing kinds of objects. And that value was understood from the outset to be not simply educational or scholarly but social and economic—Melvil Dewey went so far as to name the first college of library science, established at Columbia in 1887, the School of Library Economy (Vann 1961:28).

As noted, they differ mainly in what kinds of objects they manage. Traditionally, at least, museums focused on specimens and artifacts, on art and dimensional ‘things’—while libraries focused on books, journals and documents, on paper and later film. Over the past several decades both kinds of institutions have converged, seeing their roles less in terms of solely managing the tangible items in their care and more in terms of managing the extant or potential information that those items convey, communicate or contain (Trant 2009; Bearman 2008).

For the most part that archive was formed through the efforts of scholars, who supported the production of knowledge through their publication of scholarship and creation of collections through research. Publication, largely via academic publishers, has long been viewed as both our primary means of dis-

seminating scholarship and simultaneously our surest and most effective method for archiving that knowledge. No single archive could systematize and collate all knowledge, and it is instead held in the back issues of thousands of different journals, held in multiple at myriad repositories and libraries, a distributed archive long before the internet made such concepts fashionable (European Bureau of Library, Information and Documentation Associations. 2009).

Museums similarly serve as archives for what we know about different aspects of the world. In natural history museums taxonomic type specimens—holotypes—backed up by the entire nomenclatural bestiary of secondary types—serve as a distributed archive across world museums of the organismal record of biological systematics, paralleling the published archive held by world libraries. In ar-

chaeological and anthropological museums, collections document cultural diversity and change, and through the careful comparison of objects and their seriation we came to define cultural sequences over broad areas and thousands of years long before the advent of radiocarbon or other chronometric techniques. And artistic traditions are founded and fixed on key museum objects perceived as central to or marking a fundamental change in the canon (Barker 2010).

Scholarship also played a key role in academic publishing, as the majority of journals were produced through scholarly societies, and their viability depended on both the volunteer labor of editors, assistant editors, reviewers, and others, and on subscription-based models funded by individual scholars and by institutions. The revenue models on which institutions like these were based are changing rapidly, and many once-solid foundations—of both the scholarly and financial kind—are eroding at a quickening pace.

Convergence of museums, libraries, and academic publishers (comprising memory institutions in the sense envisioned by Dempsey [2000]) is driven in part by the rapidly expanding scale of the archive. How quickly is it expanding? One recent study (Bornmann and Mutz 2015) found that the literature grew at less than 1% annually until the middle of the 18th century, by the period between the two world wars it had risen to 2-3% annually, and had risen to 8-9% by 2010. In practical terms, this means that the time needed for the total scholarly literature to double in size has become shorter and shorter. Now it doubles in volume

every decade—or less. Nor is this an outrageous result; previous bibliometric studies since 1965 have shown essentially the same trend (e.g., Price 1965, van Raan 2000; see also Riviera 2013). Using completely different methods, other researchers estimated that 1.35 million peer-reviewed papers were published annually by 2006, in line with a 2004 estimate by Elsevier to a parliamentary committee of 1.2 million peer-reviewed articles in science, technology and medicine (UK Parliament Select Committee on Science and Technology 2004).

In museums too, the growth of the archive is staggering. In the 1990s the federal Institute for Museum and Library Services estimated there were roughly 17,500 museums in the U.S., as of 2015 that number had doubled to 35,144. To put that number into perspective, today there are more museums in the U.S. than all Starbucks and McDonalds combined (Figure 1). And their collections are enormous. The National Park Service alone holds more than 63 million individual archaeological items, the larger Department of the Interior lists over 194 million as of February of this year, and the Smithsonian's website lists more than 138 million individually cataloged objects (Barker 2001, 2010).

And rates of growth are accelerating, as many academic publishers have moved from rejection rates (or selectivity) to citations (or impact) as the primary measure of quality, meaning there is increasing value to publishers in expanding the number of contributions published. Scale has become its own currency—the bigger the archive the more valuable—and as a result we see a trend toward fewer, much larger academic publishers,

as the more commercially viable presses either absorb or bankrupt smaller houses (Figure 2). The same pressures are at work in museums. For nearly two decades cultural heritage professionals have spoken of a “curation crisis,” (e.g., Childs 1995) as there are more collections being generated—both through research and compliance activities—than there are museum shelves to hold them, or curators to make sense of them.

The very range of anthropological holdings in museums—as a single apposite example—illustrates the complexity of the challenge, including everything from saliva samples held in freezers to pollen and phytolith samples documenting climate change over tens of thousands of years—and accompanying objects documenting how human societies dealt with those changes. They include prehistoric artifacts in their millions and unique objects like the intact Confederate submarine Hunley, or Ötzi the ice man, representative of the hundreds of thousands of human remains in world museums—even if he’s aged better than most. And work on any one of those myriad objects generates new objects to be tracked across all three kinds of archive (e.g., Barker 2001).

Another trend, now formally mandated by White House policy, is a shifting emphasis from the published results of research toward an equal concern with the datasets on which research is based, partly to promote synergistic or secondary research, partly to promote validation of data and conclusions, and partly to promote greater transparency and accountability. In 2013 the Office of Science and Technology Policy (OSTP 2013) di-

rected each federal agency funding research at significant levels to require researchers to provide detailed data management plans ensuring long-term access to all data resulting from sponsored research. In addition to preservation of and access to data from research projects, these data management plans often preserve metadata regarding the workflows associated with each dataset.

So where once we could differentiate between collections of objects on which research was done, processes of preparing, reviewing and publication of research, and access to and archiving of that research through libraries, all of those lines have now blurred. The raw data itself is now being archived, and in some cases published through peer review. OpenContext is an online initiative which aggregates, reviews and publishes research data and datasets rather than publications (Figure 3). Unlike online repositories like tDAR, OpenContext is aimed at the review and publication/dissemination—rather than long-term dark archiving—of research data, much of it based on physical objects in museums, or analytical datasets derived from such objects. Such programs further subvert and transform the boundaries between production, consumption and preservation.

Another obvious trend is the demand for open access. EU, the UK and White House directives mandate different levels of public access for any publicly funded research (e.g., OSTP 2013). While open access has gained many champions in recent years, it poses significant challenges to most forms of academic publishing, since to date it has only proven financially viable in a small number of

highly monetized disciplines, and instead of leveling the playing field for all scholars it has instead shifted the locus economic inequity. Instead of making it harder for less-advantaged scholars to access the archive (because of price), open access in its current forms effectively makes it harder for less-advantaged scholars to contribute to that archive (because of cost)—hardly a step forward (Figure 4).

Several implications emerge from these trends. First, because of reader overload (if naught else) we are likely to see a massive increase in curated content across academic publishers, libraries and museums. This includes both human-curated content—a trend we already see, largely supported through blogs, social media and other informal content providers—and algorithm based curation predicting other texts based on current selections. This also suggests a shift in authority from editors—who helped establish an intellectual space and voice for a given journal by grouping papers—to readers, who are increasingly able to select their own content, or select content based on curation by other individuals, aggregating content seamlessly across a range of titles.

This is made easier by another trend—the erosion of the traditional, legacy forms of aggregation. Print-based workflows aggregated content hierarchically by title, volume and issue, and groups of papers (selected by editors) appeared together in a single issue. While those trappings remain, articles are generally published online, separately, as soon as they've been accepted. Increasingly they're consumed not as groups of articles in a given issue, but as standalone

works retrieved by search, and major titles are moving from issue-based publication to continuous publishing models.

A further indication of this convergence is apparent in peer review. Once an early or initial element in the production of publications, it is increasingly expanding and becoming part of the archive of published research as well. Commercial publishers have long benefited from the volunteer labor of academic reviewers, but this too is becoming a revenue stream. Elsevier recently filed for a patent for its waterfall system of peer review, causing considerable disquiet among academics and proponents of more open systems (Blumenstyk 2016). Services like Rubriq (<https://www.rubriq.com/>) (Figure 5) provide presubmission, fee-based peer review, and some services like AXIOS Review (<https://axiosreview.org/>) offer pre-submission review and cascading submission programs which seek to place articles in multiple journals, ranked in preferred order. Peerage of Science (<https://www.peerageofscience.org/>), a free service, also allows pre-submission peer review, and some journals may accept such submissions based on external review in lieu of peer review by the journal itself.

The *Journal of Atmospheric Chemistry and Physics* (<http://www.atmospheric-chemistry-and-physics.net/>) (Figure 6) further blurs the boundary between production, consumption and archiving of knowledge through an iterative review process, where papers are submitted, referees comment, authors respond, the scientific community at large offers brief comments, the author revises, and the paper is published in journal format, with

all parts of the process being part of the public archive.

But while these trends are transforming the cycle of knowledge production, what of research completed before this integrative approach developed? Some initiatives are trying to reach back and salvage parts of the knowledge production cycle which would otherwise be lost. One example is GRSciColl, (Figure 7) an outgrowth of an older biodiversity collections registry, which is currently seeking to register all scientific research collections, including cultural collections. Part of its purpose is to provide unique collection identifiers that publishers could treat as authoritative pointers to the objects on which published papers are based. The initiative is community-curated, and anyone can add records and register collections. While crowd-sourced data of this kind present certain challenges in validation, they are already offering unexpected benefits. One longstanding problem in the knowledge production cycle has been that many collections are created by individual scholars, and when they die or retire the collections are orphaned, and may be lost. Through GRSciColl collections can be registered even if they aren't held by a museum or repository, and thus the Registry also serves an unintended function as a clearinghouse for information about collections which may be at risk, and a solution for departments inheriting orphaned collections when a faculty member retires or dies.

As external pressures like these force convergence between publishers, libraries and museums, new synergies are

emerging, and old distinctions borne of legacy print-based workflows are blurring. These synergies may yet dispel the old idea of forgotten specimens in museum basements, and dusty journals on creaky shelves remote from current intellectual discourse. But only time will tell how robust those synergies will prove to be, and in the meantime they offer fertile ground for both scholarly inquiry and implications for research infrastructure.

FIGURES

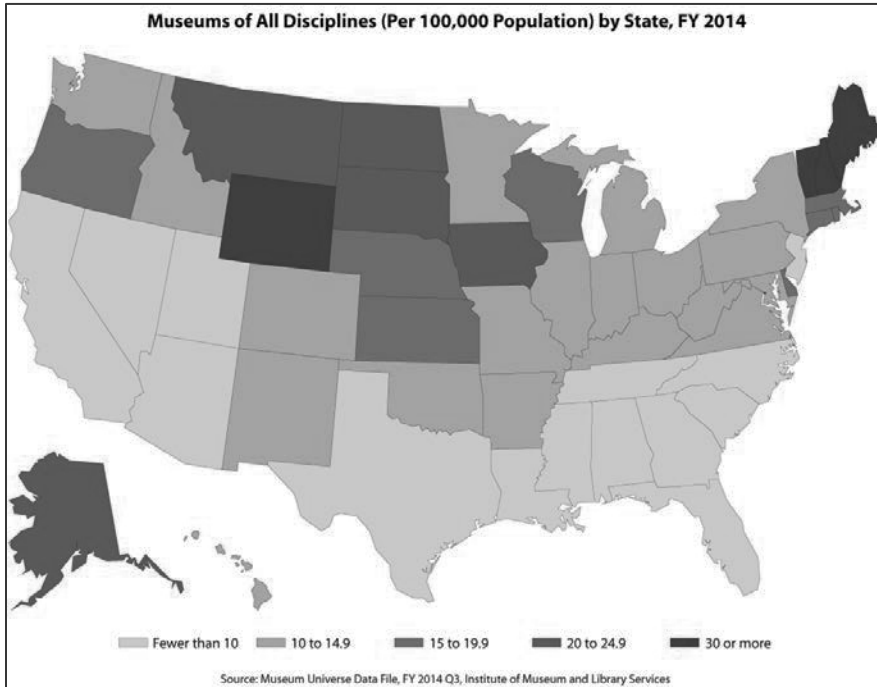


Figure 1: Density of Museums by Population, FY2014

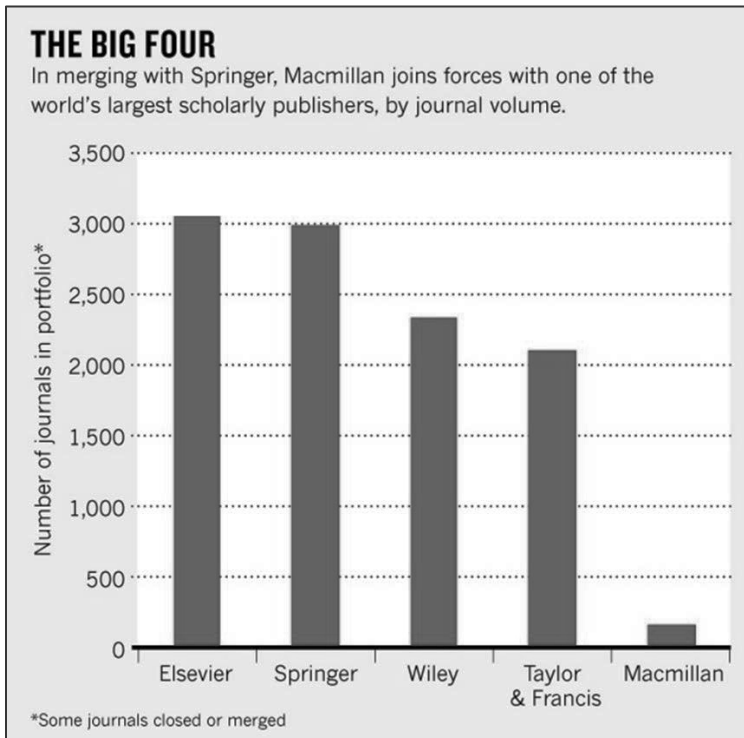


Figure 2: Fewer But Larger Academic Publishers (source Nature.com; <http://www.nature.com/news/nature-owner-merges-with-publishing-giant-1.16731>)

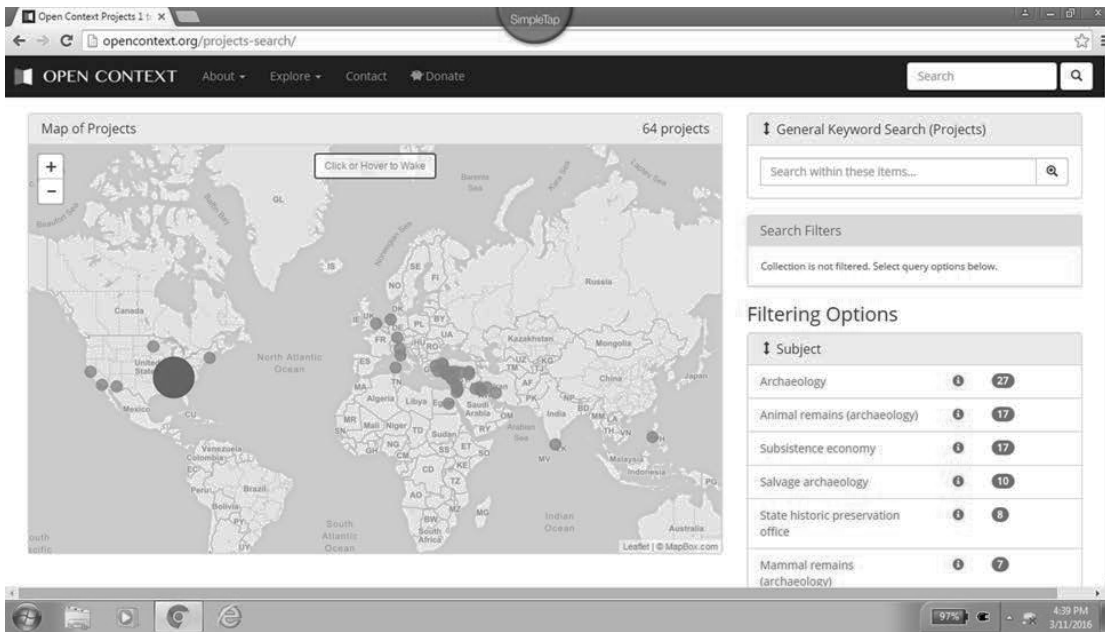


Figure 3: Open Context, an online publication venue for peer-reviewed data sets rather than publications in the more traditional sense.

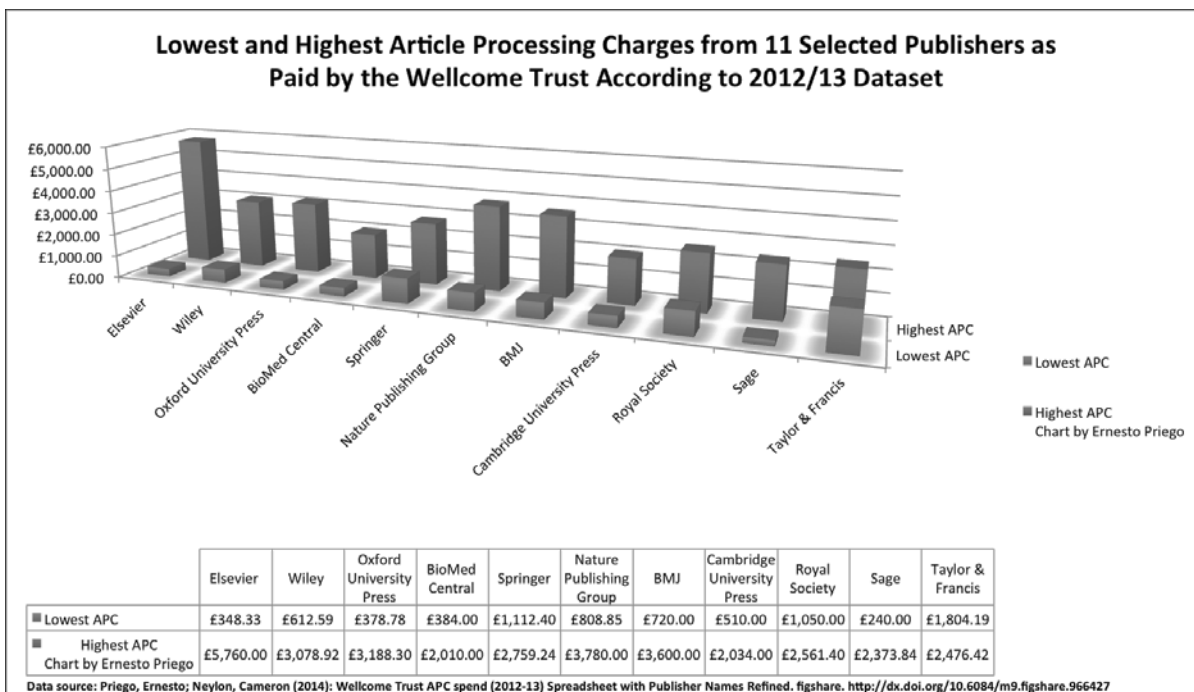


Figure 4: Range of Author Processing Charges (APCs) by Open Access Publishers

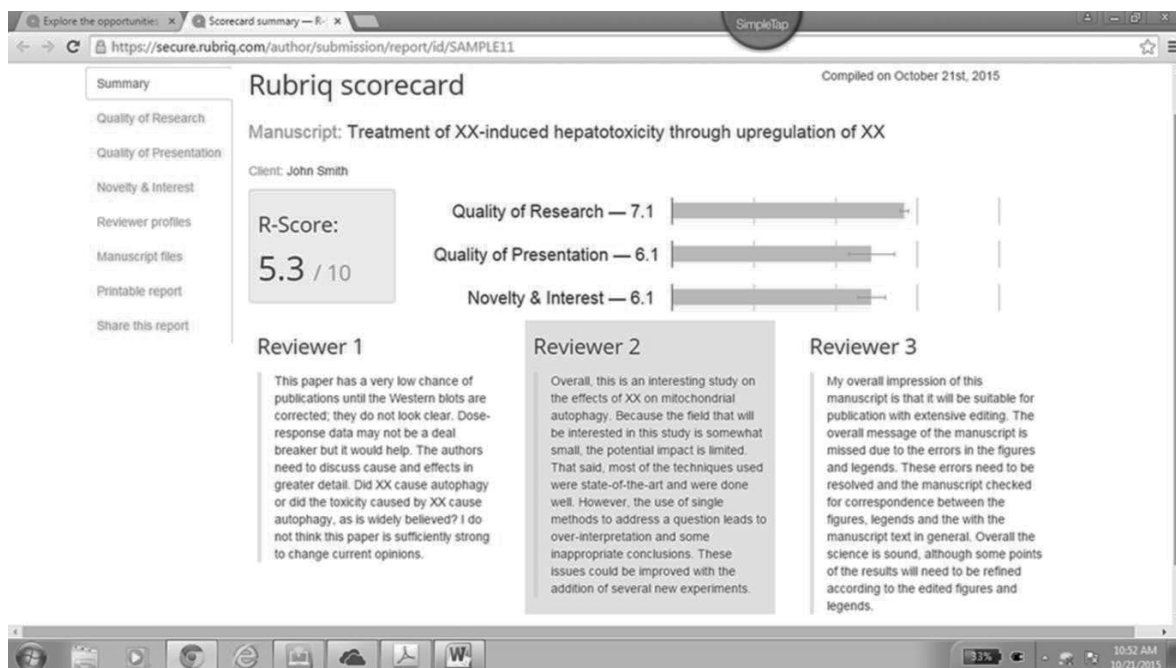


Figure 5: Rubriq, an online pre-submission peer review service

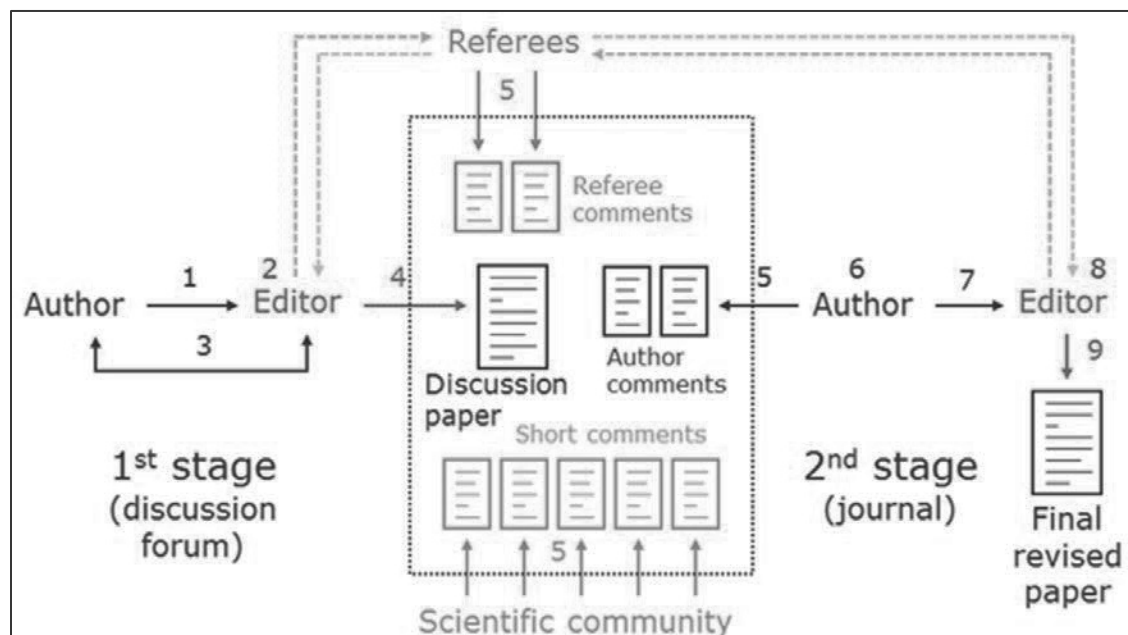


Figure 6: Journal of Atmospheric Chemistry and Physics Review Cycle

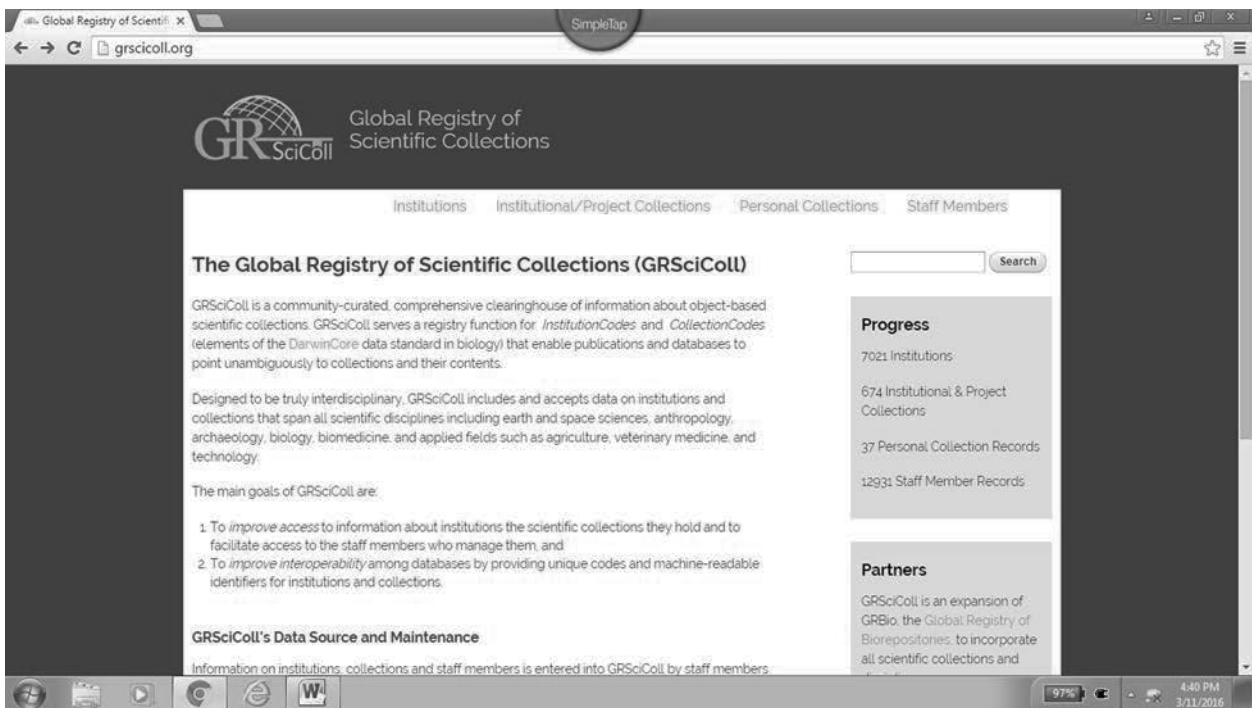


Figure 7: GRSciColl: The Global Registry of Scientific Collections. An online registry attempting to uniquely identify all scientific collections held in universities, museums or other repositories.

References

- Barker, Alex W.,
2010 Exhibiting Archaeology: Archaeology in Museums. *Annual Review of Anthropology* 39:293-308.
- 2001 Core Databases; Anthropology in Museums. *International Encyclopedia of the Social and Behavioral Sciences*, Neil J. Smelser and Paul B. Baltes, eds.,. Oxford: Elsevier Science, Ltd.
- Bearman, David
2008 Representing museum knowledge. In *Museum Informatics: People, Information, and Technology in Museums*, P. Marty and K. Jones, eds. New York/London : Routledge.
- Björk, Bo-Christer, Annikki Roos, and Mari Lauri
2009 Scientific journal publishing: yearly volume and open access availability. *Information Research* 14(1)(paper 391)(<http://InformationR.net/ir/14-1/paper391.htm>) (accessed 9/5/2016)
- Blumenstyk, Goldie
2016 Elsevier's New Patent for Online Peer Review Throws a Scare Into Open-Source Advocates. The Chronicle of Higher Education, http://www.chronicle.com/article/Elsevier-s-New-Patent-for/237656?cid=wc&utm_source=wc&utm_medium=en&elqTrackId=f4aabfe396fd47faa64097569f84c042&elq=198031dfb0304e56b8b2cd77d157fab9&elqaid=10545&elqat=1&elqCampaignId=3974 (accessed 9/6/2016).
- Bornmann, Lutz, and Rüdiger Mutz
2015 Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology* 66(11):2215-2222.
- Childs, S. Terry
1995 The Curation Crisis. *Federal Archaeology* 7(4):11-15.
- Dempsey, Lorcan
2000 Scientific, Industrial, and Cultural Heritage: A Shared Approach. A research framework for digital libraries, museums and archives . *Ariadne* no. 22. <http://www.ariadne.ac.uk/issue22/dempsey/> (accessed 9/5/2016)
- European Bureau of Library, Information and Documentation Associations
2009 Digital Agenda for Europe: Digital Libraries Initiative. <http://www.eblida.org/news/digital-agenda-for-europe-digital-libraries-initiative.html> (Accessed 9/5/2016).

- Office of Science and Technology Policy (OSTP)
- 2013 Increasing Access to the Results of Federally Funded Scientific Research. Memorandum for the heads of executive departments and agencies from John P. Holdren, Director and Advisor to the President, 22 February 2013. https://www.whitehouse.gov/sites/default/files/micro-sites/ostp/ostp_public_access_memo_2013.pdf (accessed 9/6/2016)
- Price, Derek de Solla
- 1965 Networks of scientific papers. *Science*, 149(3683), 510-515.
- Riviera, E.,
- 2013 Scientific communities as autopoietic systems: The reproductive function of citations. *Journal of the American Society for Information Science and Technology*, 64(7), 1442-1453.
- Trant, Jennifer
- 2009 Emerging Convergence? Thoughts on Museums, Archives, Libraries and Professional Training. *Museum Management and Curatorship* 24(4):369-387.
- U.K. Parliament, Select Committee on Science and Technology
- 2004 *Scientific publications: free for all? Tenth Report of Session 2003-04. Volume 2. Oral and written evidence. Appendix 46. Memorandum from Reed Elsevier. London: Parliament.*
- van Raan, A.F.J.,
- 2000 On growth, ageing, and fractal differentiation of science. *Scientometrics* 47(2):347-362.
- Vann Sarah K.,
- 1961 *Training for Librarianship Before 1923. Chicago: American Library Association.*