Vinton G. Cerf

DOI:10.1145/3169085

The Role of Archives in Digital Preservation

IHAD THE pleasure of spending a week in Mexico City to participate in the annual meeting of the International Council on Archives^a hosted by its Latin American branch, the Association of Latin American Archivists^b (ALA). This was a large and wide-ranging conference that stimulated many thoughts and ideas that affected my view of the digital archiving challenge. It is clear there are many archival institutions pursuing the problem out of necessity. They are receiving or collecting artifacts created or rendered in digital form. Many of these items are static in the sense they could be rendered and preserved using older techniques such as print or even microfilm. There is a concept of *variable fidelity* in which some aspect of the digital artifact is imperfectly captured. For example, a text file might be retained in readable but not editable form. This does not work well with objects such as spreadsheets whose value is in the interactive use of the computations captured in the spreadsheet. But one might be willing to give up some resolution if it meant the difference between capturing a portion of versus all of a collection of digital images.

The independent operation of archives may not produce the kind of mutual reinforcement that would arise from inventing and adopting standards for representation and transmission of content, metadata, and other information establishing the origins and history of the artifact. If such standards could be developed and implemented, it might make it possible for one archive to ingest the content of another, should circumstances demand it. Moreover,

such standards would make it possible for one archive to reference the content of another—making it possible to achieve a kind of global discovery of relevant content. Some readers of this column will know there are already a number of standards in place, one of the most important of which is the Open Archival Information System reference model^c (OAIS). This is a significant specification of the desirable functionality of any archive and has the potential to establish interoperability among archives. This property, interoperability, will require that the standards for transmission of digital archival information include descriptors of format and semantics of the nature and structure of the information, its metadata and provenance, intellectual property considerations among other things. Another powerful example is the Digital Object Architecture^d developed by the Corporation for National Research Initiatives.

Some readers may wonder whether the World Wide Web is already an example of an archive. It is not because there is no assurance of long-lasting storage or accessibility of the content. To be sure, the WWW content is a candidate for archiving and the Internet Archive^e is attempting to do exactly that.

A significant test of any specification is that multiple, independent implementations can be shown to interwork. That is, information from one archive can be successfully accessed and transferred to another archive without losing critical archival information. Establishing such standards and allowing archives to be networked could create an ecosystem with substantial resilience.

Multiple instances of content could be found in distinct archives. The impact of institutional failure of an archive (For example, lack of funding, physical destruction) could be mitigated by storing information in multiple locations; an instantiation of the LOCKSS (lots of copies keeps stuff safe) principle.

During this meeting it also became apparent that archivists of digital content will need to be prescriptive about the structure and encoding of digital artifacts they can process into the archive successfully. A proactive stance could improve the likelihood that digital content can be acquired and preserved. Again, standards can help and the software industry can assist by working toward the creation of applications that produce archive-ready content. Since some digital artifacts require software to be used or rendered, it stands to reason that software itself must also be archived and executable over the long term. That includes applications and operating systems and perhaps, detailed functional descriptions of hardware to the extent that the instruction set of the computer can be emulated at need.

It should be clear by now that archives will have to have privileges to acquire and execute applications and the necessary operating system(s) so that acquired artifacts can live for long time periods. Moreover, a long-lived archive must rely on a business model that matches the desired longevity that might extend 100 years or more. Think of archives housing 1,000-year-old vellum codices, for example. In addition to the extensible standardization of digital representation, then, we will need legal and financial frameworks that assist long-term preservation. We must pursue these objectives lest our digital history be lost in a mist of uninterpretable bits.

Vinton G. Cerf is vice president and Chief Internet Evangelist at Google. He served as ACM president from 2012–2014.

Copyright held by author.

c <http://bit.ly/2zSkXb6>

d <http://bit.ly/2ArYa5c>

e <http://archive.org/>

a <https://www.ica.org/en>

b <http://bit.ly/2i4myzg>