The Principle of Digital Preservation

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1080/03615260802291212">http://dx.doi.org/10.1080/03615260802291212</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Taylor &amp; Francis</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Wed Dec 26 18:38:57 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/87033">http://hdl.handle.net/1721.1/87033</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
The Principle of Digital Preservation

Patsy Baudoin

77 Massachusetts Ave. (14S-230), Cambridge, MA, 02139

Published online: 12 Dec 2008.

To cite this article: Patsy Baudoin (2008) The Principle of Digital Preservation, The Serials Librarian: From the Printed Page to the Digital Age, 55:4, 556-559, DOI: 10.1080/03615260802291212

To link to this article: http://dx.doi.org/10.1080/03615260802291212

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or
The Principle of Digital Preservation

Patsy Baudoin

In the digital age, we need to consider ways that will help us continue to have a history and know it, to prove that we have made the discoveries and advances we claim to have made. We need to think about how to produce electronically with preservation in mind, so that what is created electronically can be reviewed, understood, re-used, and built on. We must heed what I call the principle of digital preservation.

As librarians we think rightly a lot about access, whether we work in public or technical services. We seek to provide equitable and easy access, and always with respect for individual privacy. Until recently this has meant maintaining collections for our communities: caring for volumes that occasionally fell into disrepair, controlling the environment enough to avert catastrophes, and planning for recovery in case of unavoidable disasters. Now, in the digital age, we need to learn to think like archivists, too. Archivists work with two principles, often referred to as the principle
of provenance and the principle of original order. Both help approaches to digital preservation, regardless of the strategies used to enhance the long-term availability of our digital wares.

**THE PRINCIPLE OF PROVENANCE**

Researchers, regardless of their disciplines, increasingly need to understand provenance and the chain of custody of the digital material they create, evaluate, analyze, reproduce, transmit, publish, and save. Scientists need to be able to reproduce their experiments accurately, to trust the data they analyze, and to publish data they know to be reliable. Historians, likewise, depend on primary sources: raw data must be what they purport to be, that is authentic. Understanding a file’s, a dataset’s, or an e-mail-correspondence’s provenance may seem easy on the face of it. Typically it is not.

Scientists need to know where their data have come from and where they have been. To begin with, how were the data created and gathered, using what instruments (read hardware, software, and instrumentation)? What migrations paths were followed—and how can they be documented? To have this information is to have begun to curate the data, to acknowledge the difficulties of retrieving this information once data have been manipulated. The data’s lineage, all the way to their provenance, offers the safest and surest measure of the data’s pedigree and certification of trustworthiness.

Without provenance the task of reproducing and reusing data—of applying the scientific method that requires experiments to be repeated in the same conditions—is practically impossible. The chain of custody, if it is documented from creation onward, provides the requisite audit trail and confers trustworthiness of the data. With little trust in the data and their authenticity, there are arguably fewer reasons to start or to continue migrating datasets for re-use, to new software and hardware platforms. Researchers of the future will continue to use data they trust, namely data whose provenance and lineage have been adequately documented.

Even a small set of e-mail threads—say a random day’s worth of e-mail correspondence with a dozen correspondents—with visible originators (“From”), will surely be a tangle by nightfall. And although everything will not unravel in one day, things get quickly worse as the days grow into a year, and the years into a lifetime. Historians will try, as they have always done, to piece together the contexts and threads of professional
and personal correspondences by following these threads to their origins, reconstructing their chains of custody, and weaving uncovered associations and breaks, to provide us with the historical narratives that help us know who we are. We can only provide future generations of scholars with the wherewithal to build on our work if we remain mindful of provenance when we originate our own electronic files and data sets.

THE PRINCIPLE OF ORIGINAL ORDER

When archivists process an individual’s or an organization’s collections, an effort is made to maintain the original order of the material (although not to the absurd extreme of preserving chaos, because the collection must be arranged to be usable). The idea is to respect the organization of the original creator. Why? Because the arranging of one’s working data, files, articles, correspondence, and so on provides a context within which records acquire their meaning and their value. The way in which scientists organize their data and datasets may, for example, speak volumes about their research intentions, plans, and conclusions. Librarians would do well to translate the spirit of this principle as they plan digital preservation projects, especially if no archivist is on the team to remind them of the principle of original order.

There are several reasons to uphold the spirit of this principle when we work with collections in digital libraries or with our repositories full of articles, datasets, educational, and other Web pages. For one thing, digital environments tend to fragment collections into their component items. A correspondence is not merely a collection of letters and other missives; in the digital world, the collection can be and typically is itemized, so that a single letter can be retrieved out of the context of the whole correspondence or a chapter of an e-book out of the context of its author’s fullest, book-length statement. Retaining collection-level documentation (in the form of metadata) matters: it frames the relationships among the collection’s items and provides a context for anchoring and associating items to original purpose.

THE PRINCIPLE OF DIGITAL PRESERVATION

The principle of digital preservation, I propose, is to think archivally—to be mindful of the two principles addressed earlier that have provided us to
date with interpretable materials by scholars whose work we build on and elaborate. This principle of digital preservation is set forth here to prevent the loss and corruption of digital information as well as to discourage practices that put data at risk at any point in their lifecycles. But it is more than an exhortation to do no harm. It is a call to think and behave archivally—to respect and interpret the two core principles of archival thought and practice—when creating new electronic materials or planning a sustainable future for our digital wares and information. Archivists since the end of the 19th century have discussed these two principles, their value, and their application in practice. In recent times, further debate has emerged to determine how these principles can and should be translated into our approaches to digital archiving—whether understood loosely as continued access, or construed more strictly as a long-term, preservation-motivated strategy. There is ample literature available to contribute your own many ounces of prevention.