

Reconstructing the Digital Past: A Case Study of the Reconstruction of the Lost Pittsburgh Project

Bernadette G. Callery
School of Information Sciences
University of Pittsburgh,
Pittsburgh, PA 15260 USA
bcallery@sis.pitt.edu

ABSTRACT

The Web-based Pittsburgh Project, aka Functional Requirements for Evidence in Recordkeeping, was administered by the University of Pittsburgh's School of Information Sciences between 1992 and 1996. The site disappeared in 2000 when the School switched servers. Although partial versions of the project could be recovered through the Wayback Machine, graduate students in the School's 2008 Digital Preservation course reconstructed the entire site, and added documentation on the reconstruction process. The reconstructed site is now available at <http://www.sis.pitt.edu/~bcallery/pgh/index.htm>. This case study discusses educational strategies used in the reconstruction process, particularly the introduction of issues of the completeness and authenticity of the restored site, and considers the effectiveness of collaborative tools in the management of a group project.

Categories and Subject Descriptors

K. 3. 2. Computer and Information Science Education. K. 4. 3. Organizational Impacts

General Terms

Management, Documentation

Keywords

Pittsburgh Project, Functional Requirements

1. INTRODUCTION

One of the major concerns with the maintenance of digital documents is the ease with which they can be deleted or altered. While some web-based records can be recovered through the use of salvage sites such as the Wayback Machine, users should then question whether the files so retrieved actually represent the final version intended by the authors. In collaborative sites, these questions become more complex, as users are then dependent

This work is licensed under the Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported license. You are free to share this work (copy, distribute and transmit) under the following conditions: attribution, non-commercial, and no derivative works. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>.

DigCCurr2009, April 1-3, 2009, Chapel Hill, NC, USA



upon the availability of multiple inter-related sites.

This case study examines the reconstruction of the Web-based Pittsburgh Project, or Functional Requirements for Evidence in Recordkeeping, funded by the National Historical Publications and Records Commission (NHPRC) between 1992 and 1996, led by Richard J. Cox, University of Pittsburgh, School of Information Sciences. The purpose of the original Pittsburgh Project was to investigate possible techniques for determining the authenticity of electronic documents. Specific outcomes of the Project's research included functional requirements for recordkeeping systems, production rules, metadata specifications and the concept of a literary warrant which reflected the "professional and societal endorsement of the concept of the recordkeeping functional requirements" [1].

The resultant Pittsburgh Project website brought together the working papers, progress reports, a bibliography and a list of project participants. The original site included only citations to external publications related to the project and produced by the participants, but not the text of those publications. The project concluded in 1996 after which no further updates were made to the website. The loss of the entire site occurred in 2000 when the project files were not transferred when the School switched servers. Prospective users searching for the site were referred to the Internet Archive's Wayback Machine in order to locate a version of the site as captured in one of its Internet crawls.

2. RECOVERY PLAN

Rather than simply rebuild the lost site, the students were advised that the restored original site should appear as the core of a new website, surrounded by documentation about the process of reconstruction. This design decision illustrated the need to distinguish the recovered parts of the original site from the additions of updated links and new material. As the recovery team noted in their documentation, "We felt that it was more authentic to note what was there, rather than to force an inauthentic recovery."

The initial efforts of recovery of this site included searching the Internet Archive Wayback Machine for archived copies of the site. Versions of the site were captured in eleven separate crawls of 1998, 1999 and 2000, with the version captured in 1998 selected for this project as identified as being the most complete [2]. Difficulties with the intermittent access to and availability of those Wayback Machine sites reinforced the need for a stabilized reconstruction of the Pittsburgh Project site. Additional searches of the print and online literature retrieved external references to the project, its participants and its reception in the archives and

information science community. This updated list of references was incorporated into the restored site.

2.1 Editorial Decisions

As noted in the restoration project documentation, “The most challenging task was identifying what, exactly, we were recovering. Unlike traditional paper-based documents, websites change over time, are non-linear and contain external links” [3]. Although the 1998 Wayback Machine capture of the site was reasonably complete, it contained broken links and lacked the lists of functional requirements. To distinguish between which documents were present and which were represented only by broken links, a key document was created which listed the titles of the website’s documents in a hierarchical list which in turn reflected the structure of the original website. Missing items were indicated by a strike-through font. External links which remained unlocated at the conclusion of the project were noted as inactive and irretrievable in this key to indicate the need for future work.

Where broken links were included in the restored website, customized error messages indicated the current URL address if the content had been located elsewhere. Unresolved broken links noted the unsuccessful attempt to find the information and solicited the user’s assistance in locating the missing content.

2.2 Workflow Management

Strategies for successful project management evolved over the course of the project. The major challenge was to organize working groups that were sufficiently separate so that their work did not overlap. One successful approach was the formation of Inside and Outside Committees. The Inside Committee, which dealt with contributions by University of Pittsburgh faculty and technical staff, met with current representatives of those groups and searched for publications related to the project so that they could be linked to the reconstructed site. The Outside Committee identified external contributors (*i.e.* non-Pitt faculty), and sought their permission to re-publish their work in this new context. As there had been no special agreements made for the transfer of copyright, it was presumed that copyright continued to reside with the original authors, or, in the case of authors associated with the governmental agencies, with those agencies.

2.3 Collaborative Tools

The online course management and delivery tool used for this course was Blackboard, whose version 7.3 includes a number of Web 2.0 collaborative tools. A project wiki was built to maintain documentation on the project and to co-ordinate the efforts of the various members of the project team. The wiki was used to collect the reports of the various committees, meeting minutes and versions of documents under construction. Other sections of the wiki maintained the ongoing task list, comments from the instructor and a status report as required by the assignment.

While the project team attempted to use the wiki as a means of coordinating their activities, reporting progress, and sharing and editing documents, it proved difficult to use as a current awareness tool. The principal objection was that it was difficult to determine what pages or other content had been newly added. In time, this information was communicated to the group members via email, as too much time was spent in the project meetings making sure that all the participants were up to date with the content on the website.

2.4 Comments from the Original Authors

Several of the original contributors who responded to requests for copies of their documents from the original project noted that they had published elsewhere on the project and therefore there was little need to reconstruct the site. In reporting their interview with project director Cox, the Inside Committee noted that “he feels that anything useful about the project was written in the essays...He has little nostalgia about the project and mentioned that when it was lost, his attitude was that the project had ended anyway.”

2.5 Distinguishing the Original

As in conservation of museum objects, the restoration must be clearly distinguished from the original object. The original website, embedded in the restoration site, retained its 1996 look and feel. The 2008 wrapper website is quite distinct in its overall design, with the dark blue background providing a strong contrast to the simpler original site. The home page of the restoration provided background on the significance of the original project, and gave an overview of the restoration process, including the addition of specific information to complete the broken links.

3. CONCLUSIONS

In this project, students acknowledged that the ideal of a complete restoration was not achievable and focused on stabilizing what could be located and identifying supplementary information. Comments from the site’s original creators indicated that not all agreed that a permanent record of this manifestation of the Pittsburgh Project was necessary or desirable. These responses from the creators raise the question of whether or not all lost digital projects should be found, reminding us that appraisal decisions continue to apply in the digital realm.

4. ACKNOWLEDGMENTS

My thanks to the students in the 2008 Digital Preservation course at the University of Pittsburgh, School of Information Sciences who worked on this project: Jessica Benner, Cathy Clevenger, Sara Deegan, Ross Griffiths, Sarah Hackett, Brandi Liskey, Jenny Lockerby, Erin Mackin, Abigail Middleton, Miranda Nixon, Ann Retzinger, Erin Thompson.

5. REFERENCES

- [1] Cox, Richard J. 1997. Electronic systems and records management in the Information Age. *B. Am. Soc. Inform. Inf.* 23 (June/July 1997). The bulk of this issue was devoted to the Pittsburgh Project and related efforts.
- [2] Cox, Richard J. ed. 1996. *Functional Requirements for Evidence in Recordkeeping*. 2 Dec. 1998. 30 Sep. 2008. Internet Archive. <http://web.archive.org/web/19981203042506/www.sis.pitt.edu/~nhprc/>.
- [3] The Pittsburgh Project Group. 2008. “Recovery Phase: Recovery Project Strategies.” http://www.sis.pitt.edu/~bcallery/pgh/index_files/Page431.htm.