The Long Term Fate of Our Personal Digital Belongings:
Toward a Service Model for Personal Archives
Catherine C. Marshall*, Sara Bly**, Francoise Brun-Cottan***
*Microsoft Corporation (USA), **Sara Bly Consulting (USA), ***Veri-Phi Consulting (USA)

Extended abstract

We conducted a field study to shed light on the current state of personal digital archiving in practice. We sought to identify archiving challenges stemming not only from existing and proposed personal technologies, but also from the exigencies introduced by home computer use and the genres and collections of digital materials that people value. The ultimate aim of this effort is to design a service model for the long-term storage and access of personally meaningful digital artifacts by examining how personal archiving needs intersect with emerging digital archiving technologies such as institutional repositories, personal stores, implementations of specific preservation strategies (such as emulation, replication, or migration), and format policies (such as canonicalization).

While previous studies of personal archiving have taken into account specific types of material such as email or photos, we took a more holistic view of digital belongings. Participants in this study represented a wide range of ages and backgrounds, but they were all long-time computer users, many with multiple home PCs; all engaged in a variety of digital activities. As the consequence of a pilot study, we decided to focus on digital belongings; analogies between how people maintain physical and digital belongings may be misleading, if simply because digital belongings accumulate at a much greater rate than physical ones do and are much less apt to be thought of and managed as circumscribed collections.

Some of our findings are relatively straightforward: First, we verified that our informants are becoming increasingly sophisticated about creating digital media that they value. They not only used new recording means like camera phones and digital video cameras; they also created instances of new digital document genres such as blogs, web sites, chat logs, game characters, manipulated images, and bookmarks. Furthermore, they have accumulated additional meaningful digital belongings that they did not create themselves – they received documents, photos, and other media via email; they downloaded materials from the Internet; they moved data on and off of CDs (for example, music) and they engaged in the many emergent forms of information transactions such as file sharing and photo sharing services. But while consumers are becoming more sophisticated about specific devices, applications, and services, they may not be nearly as savvy about their computing environments as a whole, and they find the idea of long term storage of anything surprisingly daunting. In other words, they navigate among islands of understanding through rough seas of confusion.

What we observed overall was a clash of principles and practice. Ease of replication was recognized as a safety strategy for digital media, yet a surprising number of our informants either did not know how to make such copies nor did they have the media they thought was required on hand. In spite of the declining cost of digital media, our informants still adhered to the principle of keeping only what is necessary, although in practice, they let digital materials accumulate. In fact, they tended to express a marked ambivalence about the value of what they had and were uneasy about its permanence, mostly because they had all lost important files or had been unable to read external media (in fact, ¾ had lost files in computer disasters, and 2/3
had a “dead” computer at the interview site). They also reported ill-defined worries about the security of personal materials stored in an online repository, beyond simple identity fraud. There was a pervasive tendency to regard anything that had been downloaded as replaceable, either by purchasing it again, or by finding it again, even though most everyone had been surprised by Web sites that were no longer accessible.

We expected some of the barriers and challenges to keeping digital belongings at the outset. Predicting value was demonstrably difficult. Digital belongings were often distributed in on- and offline storage, on outdated media such as jaz drives, as email attachments, on Web sites, on multiple household computers (some dead), on other peoples’ computers, and in a growing number of unexpected places. Certainly much has been said about format transience; however, we observed additional problems related to format opacity since our informants knew little about the mapping between formats and applications. Because user interface designers have become increasingly agile at hiding complexity, home computer users are less apt to understand some of the more profound format differences (e.g. RAW files from their digital camera vs. the JPEG files they might send in email), important media properties (e.g. whether CDs have been written open or closed), or the consequences of saving a document in one format over another (e.g. why last year’s taxes might be better saved as PDF rather than in native Turbo Tax format).

What is more disquieting is the apparent breakdown of the desktop object-action metaphor. Many sensible archiving strategies rely on the ability to handle heterogeneous collections of files rather than worrying about individual items. Furthermore, informants neither labeled external media nor named file system constructs like folders with an eye to long-term intelligibility, instead relying on human memory to decipher storage. Desktop search does not appear to be a panacea for getting around the vagaries of human memory; while people seem to be comfortable with searching the Internet, it is clear that their query formulations and expectations are not applicable to finding (or remembering they have) materials, especially from a long-term perspective.

Yet people make do with the digital environments they have, approaching them with a mix of optimism, fatalism, and fear. They rely on ad-hoc IT support from friends and family and have no way of reconciling conflicting advice. Even in the best case, all of our informants had registry problems, software that was only partially installed, and inexplicable recurring dialogs that could not be resolved by the menu options offered: they experienced an aggregation of minor complaints that add up to a weak scaffolding upon which to build further storage applications. Finally, their vulnerability to malware – combined with folkways about how computers work – make it possible that we are entering what Terry Kuny has referred to as a “digital dark ages” in the personal realm, as well as in the institutional realm.

Given these difficult problems, can we see a path to a solution? If we examine the capabilities addressed by different kinds of archiving – records, scientific data, library and cultural heritage material, and digital arts – elements of a solution are beginning to crystallize. Coupling these elements with an understanding of current practice makes it possible to design a solution that fits consumer needs.