

Some Implications of Storing Everything Personal

*Gordon Bell, Jim Gemmell and Roger Leuder
Microsoft Bay Area Research Center
San Francisco, California, USA*

Within five years, our personal computers can store everything we read, write, hear, and many of the images we see including video. Vannevar Bush outlined such a system in his famous 1945 Memex article. For the last four years we have worked on MyLifeBits (www.MyLifeBits) that holds personal articles, books, email and written correspondence, photos, video files, web pages visited, telephone calls. While this has implications for future computers and users, the implications for imaging and archiving are less clear, but should be of uppermost concern to archivists within organizations, libraries, and museums as they answer: "why bother?"

Just when we are "catching up" and able to store all published information, including a scan of the Library of Congress' books, the appearance of a new great wave of bits...information emanating from individuals appears on the horizon. This wave arises from greater volumes of material brought about by personal computers and tools such as MyLifeBits that will capture more. Other similar research efforts include DARPA's LifeLog Project <http://www.darpa.mil/ipto/programs/lifelog/index.htm> and the UK's Grand Challenge http://www.nesc.ac.uk/esi/events/Grand_Challenges/proposals/Memories.pdf.

Archiving the works of a few notable scientists of the 20th century e.g. NIH's Profiles in Science <http://profiles.nlm.nih.gov/> hold on the order of a few hundred thousand pages and cost about \$0.20 per page for meta-data and scanning. Before we get into the requirements of having to retain individual personal computer archives of 21st

century lifetimes held on their Terabyte hard drive let's note several problems of today and the relationship to longer term archiving.

The "Dear Appy"¹ problem is most unsettling to archivists and computer professionals –and has to be solved. Just navigating the stored life of individual would at first glance appear to take almost a lifetime to sift through. While we are making progress in the capture of less traditionally archived content e.g. meetings, phone calls, video automatic interpretation and index of voice still lags. MyLifeBits is currently focused on retrieval including the *hopefully automatic*, addition of meta-data e.g. document type identification, high level knowledge. Such data is essential for the archivist, because without such higher level knowledge and concepts, the vast amount of raw bits from individuals will be completely unusable. The most cited problem of personal archives is the control of the content including personal security, together with joint ownership of content by other individuals and organizations. In many corporations, periodic expunging of documents is the standard. Similarly, one's life that is not available in publicly available documents, is owned by the organization and all documents may have to be expunged when an individual is no longer part of the organization.

¹ Dear Appy, My application, or platform, or media left me unreadable. Signed, Lost Data.