

# Curation of Personal Digital Archives Research and Development: Agenda for Research and Development

*Discussion Draft Circulated in Advance of: Stewardship of E-Manuscripts: Advancing a Shared Agenda (March 31, 2009, Wilson Library, Chapel Hill, North Carolina)*

Draft: March 25, 2009

Contributors: Cal Lee, Tom Hyry

The Stewardship of E-Manuscripts: Advancing a Shared Agenda meeting has been conceived as an opportunity to convene a group of archivists, librarians, scholars, and researchers who have been working to solve the myriad issues related to the archival management of personal records in digital form. The meeting has a two-fold purpose: for people to share their research, projects, and approaches to this set of issues and to create a common research agenda that fosters future collaboration. Towards this second end, we have developed this discussion document to give participants something to which they can react when considering a coordinated approach to development. The document proposes a course of action and set of associated approaches for dealing with personal digital archives.

1) Collect data to determine:

- a) Current collecting institution practices for dealing with personal digital materials, in terms of the Digital Curation Approaches for Personal Digital Archives (below)
- b) Specific reasons for adopting or not adopting given practices

2) Developing resources and services to support personal digital resource curation

3) Test and evaluate the resources

4) Develop, implement and evaluate strategies and materials for preparing digital curation professionals -- e.g. archivists, librarians, museum curators, records managers, data engineers, data librarians -- to make effective use of the approaches, resources and services discussed above (including professional education and development)

An essential theme running through all aspects of this agenda is **ethics** – both personal and professional. Whenever dealing with information of and about individuals, it is essential to address their interests, concerns and values from the start, rather than waiting to raise ethical questions until after systems have been designed, collections have been built, and professional practices have been codified.

## ***Digital Curation Approaches for Personal Digital Archives***

The following is an elaboration of three fundamental approaches (get, grab, guide) and one fundamental service responsibility (give) for digital curation professionals to adopt in support of personal digital collections. They are formulated from the perspective of professionals who are committed to ensuring that documentation of/from individuals is meaningfully perpetuated over time.

**GET** - dealing with removable media that arrive as part of collections.

- Creating forensic images
- Creating file-level copies
- Procedures for managing forensic images
- Using information on individuals' computer to identify places on the Internet to get further information about him/her, e.g. Facebook profile and login proposed by Simson Garfinkel (see Digital Lives conference paper from February 2009)

**GRAB** - pulling content that is of or about individuals from places on the Internet (anywhere that can be grabbed through TCP/IP)

- Defining clear selection criteria
- Operationalizing selection criteria based on some combination of two main technical options: querying or link-following (and setting various parameters for these, such as crawl depth, number of query hits, frequency of crawling)
- Monitoring crawl results over time (in VidArch terms, monitoring a campaign)
- Special case of someone who has died - when, whether and how to get their materials from privately-run services

**GUIDE** - fostering individual curation practices through development of tools, influencing software development by others, education, promotion, and advocacy

- Development of tools, guidance documents and services to help people to extract reusable copies of their content from online spaces (e.g. Bebo, BlackPlanet, Blogger, Classmates.com, del.icio.us, Facebook, Flickr, GMail, Google Docs, LinkedIn, LiveJournal, MySpace, Tagged, TypePad, Windows Live Spaces, WordPress, Yahoo! Mail, YouTube, wikis) – this is likely to involve general techniques (such as screen scraping), specialized interfaces (APIs provided by the services), and clarification of what is allowed under the end user agreements of the services
- Play an active role in shaping the direction of the software industry by making the case -- on behalf of consumer interests, rather than relying primarily on an argument for long-

term research interests – for software that better facilitates curation of personal digital collections (see discussion paper from Nick del Pozo, dated March 25, 2009)

**GIVE** - providing access, when appropriate, to personal electronic records that allow them to be searched, delivered over networks, and integrated with holdings in other formats

- Development of access systems that integrate aggregate-level archival description (e.g. finding aids) with finer-grained data, such as search over the content of files and associated metadata
- Promulgating principles and practices that properly balance access needs with necessary access restrictions (e.g. copyright, data confidentiality, protecting users from malicious software)
- Development of new professional practices based on the presumption that many personal collections will be directly accessible over the Internet, rather than requiring a physical visit to the archives