Postprint: Research on Library Services for Personal Digital Data Archiving

Authors: Guobin Huang, Di Hongyang, Wang Shu, Wang Fengxuan

Date: 2023-08-26T00:00:00+00:00

Abstract

[Purpose/Significance] To define the scope and value of personal digital data, analyze its main types, storage methods, and associated risks, and explore library-based service models for personal digital data archiving and management, thereby providing ideas and references for Chinese libraries to develop customized personal information services in the mobile Internet era.

[Method/Process] Through comprehensive application of website survey and case analysis methods, and taking the Library of Congress, Campbell County Public Library, the British Library, and Library and Archives Canada as examples—institutions that have pioneered personal digital data archiving and management services with significant social impact—this study conducts an in-depth investigation and analysis of these libraries' service methods and characteristics in personal digital data archiving and management.

[Results/Conclusion] Chinese libraries can leverage their positioning as public welfare service institutions and their accumulated advantageous experience in organizing, storing, managing, and servicing digital information resources. By combining the characteristics of personal digital data across different user groups, privacy levels, and data types, they can develop more personalized and embedded archiving and management services for personal digital data in terms of service targets, content, methods, and maintenance management.

Full Text

Study of Library Services for Personal Digital Data Archiving

Huang Guobin, Di Hongyang, Wang Shu, Wang Fengxuan
School of Government, Beijing Normal University, Beijing 100875

Abstract: [**Purpose/Significance**]

This study defines the scope and value of personal digital data, analyzes its main types, storage methods, and potential risks, and explores library-based service models for personal digital data archiving and management. The aim is to provide insights and references for Chinese libraries to develop customized personal information services in the mobile internet era. [Method/Process] Using comprehensive web research and case analysis methods, the paper examines influential early adopters such as the Library of Congress, Campbell County Public Library, the British Library, and Library and Archives Canada, investigating and analyzing their approaches and characteristics in personal digital data archiving and management services. [Result/Conclusion] Leveraging their position as public service institutions and drawing on extensive experience in digital information resource organization, storage, management, and services, Chinese libraries can develop more personalized and embedded archiving and management services for personal digital data, tailored to different user groups, privacy levels, and data types, particularly in terms of service objects, content, methods, and maintenance management.

Keywords: personal digital information; library; preservation; service

---

2. Research on Personal Digital Data Archiving at Home and Abroad

2.1 Research Progress on Personal Digital Data Archiving Personal digital data archiving remains a relatively new research field, with scholarly attention beginning in 2007. Researcher C.C. Marshall approached the topic from traditional archival theory, arguing that personal digital data archiving must address not only fundamental issues such as data types, archiving methods, and long-term access, but also consider how to identify valuable personal digital data, integrate online and offline archives, handle contextual information (such as email attachments and web links), and manage privacy, intellectual property rights, security concerns, and archiving skills. Subsequent scholarship has explored personal digital data archiving from various perspectives, primarily in three areas: (1) Basic theoretical issues, such as D. Sinn's definition of personal digital data archiving as user activities to preserve personal digital files and create backups, and research on challenges including digital management, distributed storage, value determination, and retrieval in long-term preservation. (2) Specific problems, including data types for personal digital document archiving, key issues, and coping strategies, as well as cloud storage for personal digital data covering its advantages, products, and strategies. (3) Practical developments, such as the Library of Congress's initiatives to guide the public in personal digital information preservation through awareness-raising and storage guidance, and Microsoft Research's MyLifeBits project for personal digital data archiving, which focuses on system development, experimentation, and implications.

2.2 Practical Applications of Personal Digital Data Archiving Personal digital data archiving has gradually attracted public attention. The British Library, for instance, studies deficiencies in personal digital archiving through its "Digital Lives Research Project." Oxford and Manchester University libraries jointly launched "The Paradigm Project," providing operational guidelines for processing personal digital files and web content in various formats. The University of Virginia Library, in collaboration with Stanford, Hull, and Yale universities, developed the "Born Digital Collection: An Inter-institutional Model for Stewardship Project" to establish a framework for preserving born-digital content. Microsoft Research also established the MyLifeBits personal digital data archiving project, primarily developing systems to "store all information about a person's life in one machine." These examples demonstrate beneficial practical attempts by foreign institutions. In contrast, while some Chinese scholars have recognized the importance of this issue and published relevant research, domestic institutions have yet to launch practical explorations in personal digital data archiving.

Most existing theoretical research on personal digital data archiving focuses on archiving strategies and services for different data types, with few analyses from a library perspective examining service explorations in this domain. This study employs literature review, web investigation, and case analysis to categorize the main types and archiving methods of personal digital data, as well as the practical risks involved. Based on this foundation, it analyzes from a library perspective how to develop more personalized and embedded archiving and management services for personal digital data.

---

3. Main Types and Storage Methods of Personal Digital Data

3.1 Main Types and Characteristics of Personal Digital Data Based on the connotation of personal digital data and the characteristics of digital content, the scope of personal digital data archiving can be summarized as: personal digital photos (JPEG format images produced through photography or scanning); personal digital audio (various recording data); personal video data (files recorded by video software); personal email (including sent and received messages and related account passwords); personal digital documents (various original digital files and digital copies produced during work, life, and study); and personal website content (social media such as Weibo, WeChat, blogs, and related accounts and passwords).

Personal digital data archiving has three main characteristics: (1) The archiving objects belong to individuals or families, excluding data from companies, enterprises, and other organizations. (2) It emphasizes digital data recorded and preserved by ordinary individuals in their work, life, and study, with broad scope not limited to celebrities. (3) It focuses more on personal or family memory functions rather than utility beyond the individual or family.

3.2 Main Methods for Personal Digital Data Archiving Considering the importance of personal digital data, people typically choose different archiving methods based on storage needs and data types. These methods can be summarized into four categories:

3.2.1 System Backup
System backup is unsuitable for long-term storage and retrieval. Many choose to back up personal digital data within their current computing environment, but over time, local files become mixed with accumulated backup files, making them difficult to distinguish.

3.2.2 Personal Computer
For convenience, people often store personal digital data on personal computers or laptops, but this creates problems. Temporary files stored in user folders on the C drive, for instance, are difficult to preserve long-term and are not automatically backed up.

3.2.3 External Storage Devices
Low cost and portability make removable media such as CDs, DVDs, and flash memory popular, but these have limitations. Storage space is finite, often requiring multiple devices. Additionally, each storage medium has a lifecycle—CDs wear out, and hard drives fail.

3.2.4 Online Network Storage
With decreasing costs, online storage has become the primary choice. Personal documents, photos, music, movies, sensor data, and even medical and financial records are stored online. Provided by third parties, network storage enables access from any internet-connected device. However, reliability is a concern: file uploads may face network interruptions, service provider hardware/software failures, unexpected service termination, cyberattacks, natural disasters, or corporate events like bankruptcy, business transformation, or mergers that can render archives inaccessible, altered, or lost.

---

4. Risks in Personal Digital Data Archiving

Despite various archiving methods, personal digital data archiving still faces security risks, including "benign neglect," technical risks, and legal risks.

4.1 "Benign Neglect" Risk Personal factors, including failure to archive promptly or improper archiving methods, can lead to loss or inaccessibility. Accustomed to digital formats and online storage, people in the digital age often overlook the dispersed nature of their personal digital data storage, ignoring archiving issues altogether. This "benign neglect" manifests in three ways: (1) Believing personal digital data archiving is the responsibility of museums, libraries, or archives; (2) Managing personal digital data casually; (3) Recognizing the importance of personal digital data but unwilling to invest time

and effort in management. Additionally, varying levels of protection awareness due to differences in identity, knowledge background, and profession result in different degrees of "benign neglect."

4.2 Technical Risks Unlike traditional materials (books, paper photos, documents) that can be preserved in suitable environments, personal digital data such as audio, video, digital photos, and emails require specific hardware and software for access. Neglecting hardware and software management can cause data loss. Beyond the inherent disadvantages of the four storage methods mentioned, personal digital data faces technical risks from storage medium lifecycle factors:

(1) **Persistence**: Storage media have limited lifespans. Due to quality and construction differences, persistence varies across media types.

(2) **Obsolescence**: As storage technology advances, media become outdated. Obsolescence manifests as computers lacking drivers for outdated media, hardware connections for obsolete media, or software to read files on storage media.

(3) **Usage Process**: This includes frequency of use, storage environment, and quality of supporting equipment. More frequent use shortens lifespan; storage environments require appropriate temperature and humidity, free from fire and water damage; better quality card readers and hardware connections minimize damage to storage devices.

4.3 Legal Risks Beyond technical risks, inappropriate archiving methods can create legal risks, primarily privacy violations. Relevant privacy rights include: (1) The right to be informed about the scope, content, and purpose of personal digital data collection or disclosure; (2) The right to control data, including access, modification, and deletion; (3) The right to request data security measures. Personal digital data contains extensive personal information—identity, family, assets, health, social connections—whose leakage can cause significant harm.

---

5. Exploration of Personal Digital Data Archiving Services by Foreign Libraries

As the importance of personal digital data archiving grows, foreign libraries have actively explored services to raise awareness, guide users, and meet real needs. Representative examples include the Library of Congress, Campbell County Public Library, Stanford University Library, the British Library, and Library and Archives Canada.

5.1 Overview of Foreign Library Services The Library of Congress believes libraries have rich experience in institutional digital archiving, which shares similarities with personal digital data archiving: similar tools (management and storage), common concerns (preventing data loss), requirements for effective organization for retrieval, and needs for long-term preservation. Recognizing the contradiction between rapid growth in personal digital data and public awareness of its protection, the Library of Congress launched exploration practices to promote archiving significance and provide techniques. Influenced by this, public libraries like Campbell County Public Library have also initiated personal digital data archiving projects.

Unlike the Library of Congress's multi-type approach, the British Library focuses on web data, viewing UK institution-created web pages as vital digital records that frequently disappear inadvertently, making their preservation a national strategic priority. Library and Archives Canada concentrates on email, a crucial personal digital data type. According to Canadian surveys, government workers send and receive over 50 emails daily—10,000 per person annually, or 1 million for a 100-person agency. These emails contain substantial historical and cultural value. While many use institutional addresses, some use personal ones, raising questions about archiving emails related to public affairs that are restricted by personal email access.

5.2 Exploration Angles in Foreign Library Services 5.2.1 Classification Management of Personal Digital Data

Classification management involves applying different selection, organization, storage, and maintenance methods according to data types. This meets the technical and equipment needs of different data types while adapting to management policies and cost requirements. Personal digital data classification management shares many similarities with library digital collection management. The Library of Congress's recommended classification methods essentially represent a "miniature version" of library digital resource management, with suggestions for selection, organization, storage, and maintenance:

(1) **Selection**: The first and crucial step. Individuals must determine the scope of data for classification management. The process involves: (a) Centralized storage—copying or downloading all personal digital data from local media and network spaces to a single location; (b) Data selection—considering quality and value to eliminate low-value redundant data and reduce storage space needs.

(2) **Organization**: Effective organization ensures quick, accurate future retrieval. Naming and classification have no fixed patterns; individuals can follow personal habits while adhering to accessibility and usability principles. Common methods involve first classifying by data type, then by theme, creation date, or modification date. Practical tips include using "YYYYMMDD" formats for automatic chronological sorting.

(3) **Storage and Maintenance**: Measures to ensure security, integrity, and availability include: (a) Regular media updates (every 5-7 years) to prevent obsolescence; (b) Creating maintenance plans and informing trusted relatives or friends about storage locations, access channels, and passwords to ensure inheritability.

Different data types require different output methods. For small volumes, direct output is recommended; for large volumes, automated tools are suggested. The Library of Congress's classification management recommendations are shown in .

5.2.2 Utilizing Multiple Digital Storage Spaces

Digital storage spaces include local media (disks, optical discs) and network storage services. Exploring multiple storage options meets diverse needs while balancing security and economy.

The Library of Congress recommends using different media to avoid format-related inaccessibility and storing copies in different physical locations to prevent loss from natural disasters. The 3-2-1 principle is commonly followed: three data copies on two different media types, with one copy stored away from the individual's residence.

Purdue University Libraries considers the best strategy to be maintaining multiple copies for recovery if one is lost. They offer four options: (1) Laptop/desktop computers; (2) External hard drives; (3) Flash drives and solid-state media; (4) Cloud storage services. For cloud storage, they advise reading user agreements carefully to understand data processing procedures, especially with for-profit providers (Google, Amazon, Flickr, YouTube, Dropbox, Apple).

Cornell University Library suggests considering: storage space limits, data size, storage/upload/download costs, operating system restrictions, cross-device synchronization, and remote access limitations when choosing cloud backup.

The UK Economic and Social Research Council (ESRC) funds the UK Data Archive for social, economic, and demographic data, while the Natural Environment Research Council (NERC) and Science and Technology Facilities Council (STFC) have established data centers. Although not focused on personal data, their models provide valuable references.

5.2.3 Publicity for Personal Digital Data Archiving

Target audiences are ordinary citizens lacking awareness of personal digital data value and loss risks. Library publicity serves both educational and cultural heritage functions.

The Library of Congress promotes archiving through: (1) Official websites with instructional videos, manuals, and web pages; (2) Social media (Facebook, Twitter, blogs) and video channels (iTunes, YouTube); (3) Offline activities like Personal Archiving Day during Preservation Week. They find face-to-face offline communication most effective. Some libraries also hold conferences to discuss archiving issues within the library community.

5.2.4 Guidance for Personal Digital Data Archiving

While some recognize archiving importance, limited digital literacy and skills still pose risks. Library guidance helps users understand and adopt archiving services.

Campbell County Public Library provides: (1) Digital file management training courses; (2) Archiving equipment operation guidance through demonstration and hands-on practice. During a six-week test, about 15 users received guidance, with over half attending more than four sessions. Despite time requirements due to budget constraints, users preferred this to purchasing their own equipment.

Library and Archives Canada issued an "Email Management Guidelines Roadmap" for organizational managers, administrators, information specialists, IT staff, and users. The roadmap covers: background and application scenarios, plan development principles, collecting business-related emails, description and organization based on business scope, ensuring alignment with institutional vision, appropriate preservation strategies, and legal compliance. They also launched a compliance program addressing external environment, obstacles, costs, legal policies, and expected outcomes, with specific provisions for archiving attachments (audio, video, etc.) and systematic arrangements for stakeholders and management teams.

---

6. Implications from Foreign Library Practices

As discussed, library-based personal digital data services—whether training or direct archiving—constitute a systematic social project. Historically and culturally, personal digital data preservation benefits both present and future generations. However, practical implementation faces numerous IT, operational, service model, and legal issues. Consequently, even in developed countries like the US, libraries have not yet launched pure personal digital data archiving services. Nonetheless, training and guidance are emerging. Using explorations by US, Canadian, and UK libraries as starting points, several issues merit consideration for Chinese libraries.

6.1 Recognizing Obstacles to Library Personal Digital Data Services

Despite its significance, personal digital data archiving faces practical challenges: balancing privacy protection with cultural heritage, determining qualifications for institutions accepting archiving commissions, establishing access constraint mechanisms, and obtaining legal recognition for libraries to participate. Currently, libraries primarily serve as promoters and educators. Many libraries systematically collect, organize, and archive materials related to notable public figures connected to the library or its location as special collections. Transforming this practice into a service for ordinary citizens requires addressing public concerns about privacy and rights infringement.

6.2 Raising Public Awareness of Personal Digital Data Protection

Consciousness influences behavior. Since the 1990s, digitization has permeated work, life, and study, with Web 2.0 exponentially increasing personal information creation and collection. People habitually access data from providers and personal media, taking access for granted. However, without scientific storage, personal digital data loss can cause individual harm or prevent historical event reconstruction. Raising awareness is crucial for developing library-based archiving services.

6.3 Libraries' Unique Advantages in Personal Digital Data Storage and Management

Compared to network service providers and other cultural institutions, libraries have distinct advantages: (1) As public service organizations, centralized library storage prevents commercial misuse; (2) Libraries possess extensive experience in digital resource organization and management. Chinese libraries at all levels have experience in the full lifecycle of digital resources—from acquisition to utilization. The National Library of China, for instance, practices permanent storage of publicly available online information. This experience directly benefits personal digital data storage training or direct archiving services.

6.4 Adopting Multiple Service Approaches Based on Library Capacity

Digital information resources require adequate funding, technology, personnel, and institutional support due to their unique forms, storage media, and management methods, plus rapid software/hardware updates. Personal digital data, as a type of digital resource, faces similar challenges. Chinese libraries should adopt context-appropriate, capacity-based, progressive strategies. Depending on resources, they can provide training guidance or direct archiving services, with tiered and phased implementation in storage space, service objects, methods, and technical support.

6.5 Agreeing on Personal Digital Data Utilization Methods with Depositors

Storage aims ultimately at utilization. For library-stored personal digital data, usage terms must be pre-agreed, considering: library purposes for using data, privacy-level-based public access tiers and timelines, authorized accessors, handling when depositors face accidents without explicit authorization, and whether legal heirs can inherit 委托 rights to negotiate storage and usage terms. Libraries must respect depositors' right to know, providing multiple authorization options and necessary content review mechanisms. Individuals should not rely entirely on libraries or commercial providers but should actively manage their own archives using multiple preservation strategies.

In the information age, everyone should be their own personal digital data archiving manager. For the library community, facing massive personal digital data, libraries can play multiple service roles. Constrained by national and institutional conditions, personal digital data archiving may require a period of promotion and guidance, but this trend is historically inevitable. In the initial stage, libraries can learn from foreign practices by publicizing importance

through websites, social media, and on-site activities, helping the public establish archiving concepts and providing guidance. When conditions mature, libraries can offer complete archiving services. Direct services require establishing content review mechanisms to prevent illegal content while respecting individual freedom and convenience—a challenging but essential balance.

References

[1] Huang Guobin, Zhang Shasha, Yan Xin. Research on the concept and basic types of personal data[J]. Library and Information Service, 2017, 61(5): 41-49.
[2] Library of Congress. Special presentation: importance of digital preservation[EB/OL]. [2017-08-11]. http://digitalpreservation.gov/about/presentation.html.
[3] MARSHALL C C. How people manage information over a lifetime[EB/OL]. [2018-01-30]. https://www.researchgate.net/publication/248508325_{{{How}}}{{People}}{{{Manage}}}{{Pe
[4] SINN D, SYN S Y, KIM S M. Personal records on the web: who's in charge of archiving, hotmail or archivists?[J]. Library & information science research, 2011, 33(4): 320-330.
[5] MARSHALL C C. Rethinking personal digital archiving, part 1: four challenges from the field[J/OL]. [2017-08-11]. http://www.dlib.org/dlib/march08/marshall/03marshall-pt1.html.
[6] Wu Xiaoyi. On the protection of personal electronic documents[J]. Sci-Tech Information Development & Economy, 2008(13): 146-148.
[7] Zhou Yaolin, Zhao Yue. Research on digital archiving strategy based on personal cloud storage services[J]. Library Development, 2014(6): 21-24, 30.
[8] Liu Long. Measures of the Library of Congress to guide the public in personal digital information preservation[J]. Library and Information, 2015(1): 87-90.
[9] Microsoft Research. MyLifeBits[EB/OL]. [2017-07-30]. http://research.microsoft.com/en-us/projects/mylifebits/.
[10] Zhou Yaolin, Zhao Yue. Progress in foreign personal archiving research and practice[J]. Archives Science Bulletin, 2014(3): 79-84.
[11] Wang Haining, Ding Jiayou. Reflections on foreign personal digital archiving practice—taking MyLifeBits as an example[J]. Research on Library Science, 2014(6): 62-67.
[12] Internet Headline. Four hurdles for wearables[EB/OL]. [2017-08-11]. http://zhanglei.baijia.baidu.com/article/134955.
[13] Tao Juan. On the protection of personal data privacy in the network era[J]. Forward Position, 2007(5): 119-121.
[14] ASHENFELDER M. The Library of Congress and personal digital archiving[EB/OL]. [2017-07-29]. http://www.digitalpreservation.gov/documents/lc-digital-preservation.pdf.
[15] ASHENFELDER M. Digital archiving: making it personal at the public library[EB/OL]. [2017-07-29]. http://blogs.loc.gov/digitalpreservation/2014/05/digital-archiving-making-it-personal-at-the-public-library/.

[16] Library and Archives Canada. Email management guidelines roadmap[EB/OL]. [2017-08-14]. http://www.bac-lac.gc.ca/eng/services/government-information-resources/guidelines/email-management/Pages/roadmap.aspx#appb.
[17] Library of Congress. Preserving your digital memories[EB/OL]. [2017-07-29]. http://www.digitalpreservation.gov/personalarchiving/documents/PA_{{{All}}_{{brochure}}}.pdf.
[18] Purdue Libraries. Personal digital archiving: secure storage[EB/OL]. [2017-11-28]. http://guides.lib.purdue.edu/c.php?g=595151&p=4117382.
[19] Cornell University Library. Manage your digital research files: storage[EB/OL]. [2017-11-28]. http://guides.library.cornell.edu/c.php?g=31876&p=201794.
[20] UK Data Archive. About the archive[EB/OL]. [2017-08-11]. http://data-archive.ac.uk/about/archive.
[21] British Library. Archives into the future[EB/OL]. [2017-08-11]. http://www.bl.uk/events/archives-into-the-future.
[22] ASHENFELDER M. Digital archiving: making it personal at the public library[EB/OL]. [2017-07-29]. http://blogs.loc.gov/digitalpreservation/2014/05/digital-archiving-making-it-personal-at-the-public-library/.
[23] Huang Guobin, Liu Xinran. Research on personal data security risks caused by enterprises in mobile network environments[J]. Library and Information Service, 2017, 61(10): 57-67.

Author Contributions:
Huang Guobin: Topic selection, writing, and revision;
Di Hongyang: Writing;
Wang Shu: Research and writing;
Wang Fengxuan: Research, compilation, and writing.

Classification Number: G250.76

Note: Figure translations are in progress. See original paper for figures.