Designing for Long-term Digital Data Management

Francesco Vitale
The University of British Columbia
Vancouver, British Columbia, Canada
fvitale@cs.ubc.ca

ABSTRACT

Digital data is a pervasive component of modern society, with people managing a growing number of data types across many devices. My research explores people's choices on what to keep over the long-term and aims to design personalized data management tools. In a first study, I characterized individual differences in data preservation behaviors. I plan to use interviews, a survey, and probing methods to further extend this characterization and define a design space for long-term data management. Then, I plan to build and evaluate a prototype that synthesizes findings from all my studies.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **User models**; **Interactive systems and tools**; • **Information systems** → Data management systems;

KEYWORDS

Personal information management, data management, individual differences, qualitative methods.

Research situation

I am halfway through my four-year PhD program in the Department of Computer Science at The University of British Columbia. I have completed all of my coursework requirements. I am currently writing my thesis proposal, that I plan to defend at the end of 2018. I expect to graduate in 2020.

Research goals

1) I want to investigate people's practices around managing and preserving digital data over the long-term, in order to design tools that better support their needs.

2) I want to extend previous characterizations of individual differences in data management and tailor a range of different design solutions to different user needs.

Study One sample and methods

We conducted a thematic analysis on semi-structured interviews with 23 participants (aged 21-64, varied backgrounds), asking them to show us and discuss their data on multiple devices (computers, tablets, smartphones) and cloud platforms (e.g., Dropbox, Google Drive). We also asked them to remember what they had kept or discarded through the years.

BACKGROUND AND MOTIVATION

Digital data is everywhere: files, photos, videos, texts, songs. All aspects of life, from work to domestic life, increasingly rely on digital data and technologies. Over the years, people produce, save, share, and interact with large amounts of data. They have to manage it within a set of changing technology standards and devices [9]. And they have to decide what to keep or discard over a lifetime. But common management tools rarely support a long-term perspective. To address this unmet user need, researchers in Personal Information Management (PIM) have called for an increased focus on studying and designing for long-term management [7], focusing on questions about *curation* [11], legacy [4], and automation [7]. This is the space where my thesis work is positioned.

The typical way people curate data consists of three phases: selecting, organizing, retrieving [11]. Improvements in search functions have made retrieving easier. Instead, selecting and managing still present challenges for users. In particular, users struggle when deciding what to keep or discard, often opting to keep as a default action [2]. But PIM studies also show that people's attitudes and management behaviors vary a lot. Individuals differ in how they choose to organize their data, from documents [6] to emails [5]. This variation points to a need for a personalized approach in design.

RESEARCH TO DATE

Study One: I have conducted an exploratory study to better understand how people decide what digital data to keep or discard [10]. Using an iterative analysis approach with two co-authors, we identified a spectrum of tendencies that inform people's decisions. At one end of the spectrum, participants tended to keep most of their data–we label this extreme "hoarding," using the words of some of the participants. "Hoarding" had both an emotional and practical role for participants: they kept most of their data because they were attached to it or wanted to keep a record of things. Some participants faced challenges in organizing large quantities of data. At the other hand, participants tended to discard as much data as possible–we label this extreme "minimalism," again using participants' words. Here, participants regularly cleaned up their data or tried not to download too much. Some participants expressed underlying tendencies in wanting to limit how much money or time they spent on technology and taking care of their data. We frame the tendencies as a spectrum because participants showed a nuanced attitude towards their data, with a general approach guiding their decisions and several exceptions based on different data types.

PROPOSED RESEARCH

Study One opened up several opportunities for design and additional research. In particular, I want to further understand data selection practices by focusing on how and when people discard data. I also want to further characterize how individual needs vary based on general attitudes and different data

Current and expected contributions

My doctoral research extends and brings together a range of previous studies about information management and digital possessions. My goal is to provide three key contributions.

1) A comprehensive and actionable characterization of individual differences in data preservation and long-term management behaviors. Study One in part covers this, proposing a spectrum of behaviours in data preservation. Study Two will further extends this characterization with additional data.

2) A design space for long-term data management. Study Two and Study Three's goal is to identify and propose a set of design dimensions for building long-term data management tools and then explore them within a range of potential solutions.

3) A case study of a long-term oriented data management tool, consisting of design, implementation, long-term deployment, and evaluation. Study Four will cover this.

types. This will help in building a range of design solutions. I plan to conduct three additional studies to meet my research goals.

Study Two: Additional semi-structured interviews and broader survey

In Study Two (currently in progress), I am focusing on data deletion practices, to complement the focus on data preservation that framed Study One. Previous literature shows that deleting data is a challenging task [1]. I wanted to better understand the barriers people face in deleting their data and their strategies when they attempt to do so. I decided to take a grounded theory approach [3] to extend the findings from Study One and conduct a smaller set of semi-structured interviews in participants' homes. I asked seven participants how they get rid of both physical and digital items. Including physical items was helpful in pointing to generative insights for design and gaps in the way digital tools support discarding practices. In addition, comparing attitudes to physical and digital items is a recurring way of investigating management practices in the HCI literature [8]. I have also conducted a broader survey, focused on teasing out differences in keeping or deleting decisions across different data types. This helped me to further understand how individual needs vary. I used the interviews and the survey to extend the characterization of user behaviors from Study One, proposing a set of user archetypes. The archetypes synthesize several research findings into an actionable description, highlighting different user needs and attitudes towards data management.

Study Three: Design space exploration

In Study Three (planned and currently starting), I will take a *research through design* approach [12] to explore a design space for long-term data management. Using insights from the previous two studies and related work, I will conceptualize a range of design concepts tailored to varying user needs. I will create video scenarios for a set of four to five design concepts. Each concept will differ on a set of design dimensions, with some being more provocative than others (for example, one concept will include automatic deletion of data, while another one will only provide a visualization of relevant data attributes, leaving the management choice open-ended). I will use the videos as probes in semi-structured interviews with participants as a tool to elicit reactions to the concepts and the design dimensions they represent. I will use thematic analysis to synthesize participants' reactions and impressions of the concepts, to understand people's perceptions and boundaries in this domain.

Study Four: System building and longitudinal evaluation

After the design exploration, I will focus on the most salient design dimensions and integrate them cohesively in a high-fidelity prototype, following an iterative design process. My plan is to build a browser extension that will work on top of popular cloud platforms like Google Drive or Dropbox. After building the system, I plan to deploy it in a longitudinal study with a pool of volunteers for two

to four months. This will make it possible for participants to use and evaluate the tool with their own data, overcoming a common limitation of previous PIM studies (i.e., the use of artificial file systems). I plan to use insights from the user archetypes in Study Two to recruit a small but varied sample of participants who regularly use cloud platforms and vary in their long-term management needs.

REFERENCES

[1] Ofer Bergman, Ruth Beyth Marom, and Rafi Nachmias. 2003. The user-subjective approach to personal information management systems. *Journal of the American Society for Information Science and Technology* 54, 9 (2003), 872–878.

[2] Richard Boardman and M. Angela Sasse. 2004. "Stuff Goes into the Computer and Doesn't Come out": A Cross-tool Study of Personal Information Management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. ACM, New York, NY, USA, 583–590. https://doi.org/10.1145/985692.985766

[3] Kathy Charmaz. 2014. *Constructing grounded theory.* Sage.

[4] Rebecca Gulotta, William Odom, Jodi Forlizzi, and Haakon Faste. 2013. Digital Artifacts As Legacy: Exploring the Lifespan and Value of Digital Data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1813–1822. https://doi.org/10.1145/2470654.2466240

[5] Jacek Gwizdka. 2004. Email Task Management Styles: The Cleaners and the Keepers. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems (CHI EA '04)*. ACM, New York, NY, USA, 1235–1238. https://doi.org/10.1145/985921.986032

[6] Sarah Henderson and Ananth Srinivasan. 2011. Filing, piling & structuring: strategies for personal document management. In *System Sciences (HICSS), 2011 44th Hawaii International Conference on*. IEEE, 1–10.

[7] William Jones, Victoria Bellotti, Robert Capra, Jesse David Dinneen, Gloria Mark, Catherine Marshall, Karyn Moffatt, Jaime Teevan, and Maximus Van Kleek. 2016. For Richer, for Poorer, in Sickness or in Health...: The Long-Term Management of Personal Information. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 3508–3515. https://doi.org/10.1145/2851581.2856481

[8] William Odom, James Pierce, Erik Stolterman, and Eli Blevis. 2009. Understanding Why We Preserve Some Things and Discard Others in the Context of Interaction Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1053–1062. https://doi.org/10.1145/1518701.1518862

[9] Janet Vertesi, Jofish Kaye, Samantha N. Jarosewski, Vera D. Khovanskaya, and Jenna Song. 2016. Data Narratives: Uncovering Tensions in Personal Data Management. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 478–490. https://doi.org/10.1145/2818048.2820017

[10] Francesco Vitale, Izabelle Janzen, and Joanna McGrenere. 2018. Hoarding and Minimalism: Tendencies in Digital Data Preservation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 587, 12 pages. https://doi.org/10.1145/3173574.3174161

[11] Steve Whittaker. 2011. Personal information management: from information consumption to curation. *Annual review of information science and technology* 45, 1 (2011), 1–62.

[12] John Zimmerman, Jodi Forlizzi, and Shelley Evenson. 2007. Research Through Design As a Method for Interaction Design Research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)*. ACM, New York, NY, USA, 493–502. https://doi.org/10.1145/1240624.1240704