

Keeping and Discarding Personal Data: Exploring a Design Space

Francesco Vitale¹, William Odom², Joanna McGrenere¹

¹ University of British Columbia, Vancouver, British Columbia, Canada

² Simon Fraser University, Surrey, British Columbia, Canada

fvitale@cs.ubc.ca, wodom@sfu.ca, joanna@cs.ubc.ca

ABSTRACT

Today, most technology users have to deal with a growing amount of personal data across many devices and online platforms. There is a growing need for tools that can help people decide more intentionally what data to keep or discard. We created five design concepts in the form of video prototypes to probe on alternative design strategies for supporting users. *Automation*, *aggressiveness*, and *temporality* were key dimensions we explored. We conducted interviews with 16 participants using the concepts as a starting point for discussion. Participants had a range of reactions: some wanted to retain full control over keeping and discarding decisions, while others welcomed more automatic tools. We identify common ground in the need for a contextual and nuanced approach in design. We use these results to outline and reflect on possible future design directions for personalization, automation, new keeping or discarding actions, and privacy.

Author Keywords

Personal data management; data selection; keeping data; discarding data; research through design

CCS Concepts

• **Human-centered computing** → Empirical studies in HCI;
• **Information systems** → Data management systems;

INTRODUCTION

A day in your personal digital data. You wake up. You open Facebook. A dozen notifications, a few messages. Google Photos reminds you that last year on this day you were on holiday. You enjoy your daily dose of nostalgia and quickly move on. Send an email, share a file, download a pdf. A quick photo of your dinner. Reboot and repeat. It's been about ten years since your first smartphone. Your personal data, once limited to a personal computer, is now part of an ecosystem of devices and cloud platforms that make daily management challenging [72] and encourage ongoing accumulation [36]. You are not alone.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

DIS '19, June 24–28, 2019, San Diego, CA, USA.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-5850-7/19/06 ...\$15.00.

<http://dx.doi.org/10.1145/3322276.3322300>

The process of managing personal data over time can be framed as a “curation” cycle with three stages: “keeping, managing, retrieving” [79]. We look specifically at the *keeping* stage, where users acquire and select data over time. We define *selection* as the process of intentionally deciding what data to keep or discard. Since before the advent of cloud computing, researchers in Personal Information Management (PIM) have argued that selecting data over the long-term is necessary for “emotionally viable” archival systems: choosing what to keep and what to let go is important so that you can derive value from your personal data over time by keeping things that matter to you [52]. But as the amount of personal data grows, it is nearly impossible to decide what to keep and discard [9]. This ongoing selection process and the challenging decisions around keeping or discarding data are the focus of our work.

Recent studies on the value of digital data [26] and its longitudinal management [41, 73] point to a growing need for tools that can support keeping and discarding decisions. They also consider how individuals differ in their daily practices and keeping preferences (with some preferring to keep most data and others trying to limit their collections) [73]. This is the space that our work aims to inquire into and help expand. Research in the HCI community has shown it is unlikely that a single solution could satisfy all users [41, 73]. Yet, it is still largely unknown what specific design strategies might be viable in supporting individuals’ ongoing selection practices. With these issues in mind, the core research questions guiding our inquiry are: How can technologies be better designed to support people’s decisions around what data to keep or discard? In particular, what different design approaches might be viable to different users and in different situations?

Using a Research through Design approach [86], we created five design concepts as a way to probe people’s reactions, attitudes, and perceptions on the role of technology in supporting personal data selection practices. The concepts intentionally emphasize different design dimensions stemming from related work in PIM. The five concepts are: *Patina*, a visualization of temporal aspects of data (e.g., age or number of interactions), *Data Recommender* (a recommender system that suggests data to take care of using machine learning), *Temporary Folder* (a folder with an expiration date), *Temporary App* (a mobile application with an expiration date), and *Future Filters* (a mobile application to create advanced filters for deciding what to do with data in advance). For each concept, we created a short

video sketch [85] as a prototype that primarily illustrates how it works. Then, we conducted one-on-one interview sessions with 16 participants with varied data management approaches. The interview sessions touched on the potential benefits and drawbacks of the concepts, with a range of reactions. We identified contrasting attitudes towards the systems we presented, with the tension between automation and control informing the need for context-based solutions. Drawing on the interview analysis, we critically reflect on our results and outline future design directions to further open the design space.

Our work makes three key contributions. First, we outline four design dimensions (*selection regime*, *automation*, *aggressiveness*, and *temporality*) to define and broaden the design space around keeping and discarding decisions. These can be used as a generative resource for creating new solutions. Second, we offer five alternative design concepts that we used in an elicitation study with a diverse sample to probe and explore the space, showing where people's key boundaries around control and automation lie. Third, we discuss future design directions for supporting keeping and discarding decisions focusing on personalization, automation, defining new actions, and targeting data privacy.

BACKGROUND AND RELATED WORK

We draw on work by Whittaker [79] to define the focus of our exploration. Whittaker frames PIM actions as a “curation” process in three stages: keeping, managing, retrieving. While Whittaker's focus is broad, we focus on the keeping stage. Following Vertesi *et al.* [72], we use the term *data* in place of *information* to cast a broad focus that includes digital items spanning many devices and platforms. Our concepts focus on data that people capture or create explicitly (e.g., files, folders, images, apps). In some cases we use data captured automatically by technology (e.g., metadata on frequency of use) to support the selection process. We do not delve into lifelogging technologies (i.e., Personal Informatics systems [19, 45]) but our ideas can apply to this space, where people produce massive amounts of data and need some form of selection [80, 81].

In the following sections, we first touch on key user practices and challenges in selecting data. Then, we review design efforts around personal data management from prior research projects. We use this review to outline a set of design dimensions and approaches to probe on.

User practices and challenges in selecting data

Work on personal data management highlights how deciding what data to keep or discard is a key challenge for users [4, 5, 10, 73, 79]. The ongoing process of selecting data is challenging because it often involves anticipating the future [9, 79]. Many users end up choosing to keep all digital data as a default [10]. Bergman & Whittaker [9, 79] refer to *prospect theory* [39] to explain why. People are averse to risk and perceive a potential loss from discarding data as more substantial than any gain. (i.e., The possibility of losing information that turns out to be useful outweighs the possibility of finding data more easily.)

The prevalence of keeping as a default decision, however, creates an accumulation cycle: the more data, the harder it becomes selecting, organizing, or finding things [9, 73]. An

emerging body of research points to feelings of stress about the accumulation of data [69, 71, 73]. The increasing popularity of cloud platforms further complicates keeping decisions. People often perceive their digital data as being an undefined collection of items without knowing exactly what they possess and where things are. As Odom *et al.* [60] remark, “the role of curator can become complicated if one does not know what one is curating.” Yet, there is a desire for some form of selection. Previous studies highlight that a meaningful personal archive should focus on “the remarkable” items [47] and that selection is one of the ongoing practices that inform what an archive is [40], although there are considerable differences in how much people tend to keep or discard [73]. Thus, our goal is not to force people to discard data, but rather support any level of selection that each individual deems appropriate to emphasize the value of things that matter to them (whether those things are many or just a few).

Studies on people's physical “decluttering stages” [15] or “selection regimes” [36] report useful accounts of people's selection practices and discuss how they can inform digital systems. For example, people often look at items individually as “best examples” of a broader collection, or instead they consider a mass of items all together for later processing [36]. However, the HCI community still lacks a systematic attempt at defining a design space around keeping decisions. To define its possible dimensions, we now turn to related design work.

Existing and proposed design approaches

Augmenting data management interfaces

Two common user interface paradigms to manage data are: 1) *document-centric*, with the most dominant desktop metaphor of files and folders (common on personal computers and cloud storage platforms), 2) *application-centric*, with the application acting as a bundle for data (common on mobile devices and social media platforms) [1, 78]. Both paradigms have benefits and drawbacks. The file systems community, for example, has criticized the desktop metaphor organized around folders for being too rigid and inadequate as the amount of data grows [70]. But despite their faults, folders still dominate management platforms because they provide valuable functions: they help people control, organize, and structure their work [78]. Enhancing them, rather than replacing them, might be the best design approach [78].

Several projects propose augmentations or alternatives to folders using for example metadata [77] or annotations [74]. Other projects choose an alternative activity-centric approach [1], exploring “time-ordered streams” of documents [20], flexible desktop organizations [18, 75], or new metaphors based on places, time, and data provenance [48]. While exploring radical alternatives can help push forward design, our concepts largely focus on augmenting current interfaces so that participants can better relate them to their own experience.

Using automation to complement selection practices

Most of the design projects reviewed so far focus on data organization or retrieval, with little attention paid to keeping decisions. An exception is work by Bergman *et al.*, with several related projects addressing the keeping stage of the curation cycle: *GrayArea* [8], *DMTR* [7], and *Old'n Gray* [6].

These systems use the “demotion” principle [4], an intermediate action between keeping and deleting (which is the most common discarding action afforded by user interfaces). *Demotion* makes items visually less prominent or hides them in a separate area of the interface. This is a valid compromise between keeping and deleting: unnecessary items do not distract when they are demoted, but they are still there in case they are ever needed. Although *discarding* data can mean more than deleting (e.g., demoting) our concepts focus on deleting as a key action in order to elicit more powerful reactions from participants and understand where their boundaries lie.

The examples by Bergman *et al.* also highlight two distinct design approaches to solve the “burden of curation” [36]: in GrayArea, users rely on direct manipulation to demote items by dragging them into a separate area of a folder, where in DMTR and Old’n Gray the process is automatic. The tension between automation and user-control is at the centre of many investigations in HCI [21, 24, 34, 65]. Within PIM, the discussion around automating user management strategies goes back to key studies about email [3, 82]. Some examples of automation or semi-automation focus on selecting photos [46, 57] and audio [55], or on the process of passing down digital data [27]. Jones [37] discusses automatically archiving information that is no longer useful, while *Vanish* [22] introduces the idea of self-destructing data. Bergman *et al.* [5], among others, argue for finding a balance between automation and user control in PIM interfaces. Yet, this is a question rarely explored in the specific context of keeping decisions. Understanding which “curating” actions can be automated and which should not is an ongoing open question in PIM research [38]. Our work uses the tension between manual actions and automation as a key design dimension to explore potential new directions.

Using metadata to build awareness of digital items

Most design efforts discussed so far focus on improving data management tasks. A different strand of design work by Odom and colleagues, instead, offers a counter-perspective, and focuses on reflection, reminiscence, and enjoyment [58]. This approach uses metadata for rediscovering kept data through everyday objects [59, 64]. Sas *et al.* [67, 68] also use a similar, reflective approach by proposing “rituals” for letting go of sentimental digital items. We use this work as inspiration to add possible design choices and incorporate an open-ended, reflective dimension in some of our concepts. While prior work focuses on tangible interactions, we explore using metadata in graphical user interfaces to resurface old digital items or build awareness of their accumulation over time.

Exploring the potential of prospective decisions

Finally, we narrow the focus to email, with work by Gwizdka [28–30] providing additional inspiration. In studying and designing email management tools, Gwizdka introduces the idea of *prospective* information to support task management by anticipating future needs. Today, several email management tools apply a similar concept through reminders and snoozing functions. Empirical work on more general data selection decisions points to temporal dimensions that consider both the past and the future [36]. For example, Kim [42] mentions one participant having a folder called “to delete,” while Brewer

et al. [11] discuss prospective memory for digital reminders. Thus, we see the challenge of anticipating future needs with prospective decisions as a major opportunity to expand the design space. While most data management tools are retrospective, there might be space for prospective decisions and we use some of our concepts to explore this area.

METHODOLOGY

Research approach and design dimensions

Our review of related work shows how designing to support keeping and discarding decisions is a largely under-explored territory, with different potential directions to follow. This multitude of possibilities makes a design-led exploration ideal. Thus, in our work, we took a Research through Design approach [86]. Our inquiry can be seen as parallel to Gulotta *et al.*’s [27] use of a similar approach to investigate the space around data curation, legacy, and memory.

We started by clustering and mapping insights from prior work into four key design dimensions to probe: *selection regime*, *automation*, *aggressiveness*, *temporality*.

Selection regime - The first dimension considers possible “selection regimes” [36] that people use when evaluating data: whether they consider one item at a time or a collection of items all together. We created variations along this dimension to encompass both individual items and collective categories, probing on the differences in support needed for both.

Automation - The second dimension focuses on the tension between user-initiated selection and automation. This is a key issue to explore given the increased potential for automatic data management tools, thanks to machine learning and artificial intelligence. We contrasted concepts that took automation to its extremes with others that have more user influence.

Aggressiveness - The third dimension is about the level of aggressiveness of the system. We contrasted concepts that are more open-ended (i.e., they only inform of data to take care of, whether the user notices it or not, letting them decide what to do and when) with others that are more forceful and push the user to decide whether to keep or discard something.

Temporality - The final dimension represents the temporal user mindset in selecting data: either *retrospective*, looking at items based on past use, or *prospective*, looking at items based on future use. Variations along this dimension probed on the largely underexplored area of prospective decision making to see whether this might be a viable direction compared to the retrospective nature of many traditional management tools.

We created five concepts that differ along these design dimensions. For each concept, we created a short video prototype or sketch [85], an illustration of how it works. The videos (available as supporting materials to the paper) use a mix of descriptions and user scenarios, depending on what we felt best illustrated each concept. The concepts take inspiration from existing work or systems, as we detail in their respective descriptions. However, in order to have more control in our elicitation study, we decided to design our own set of video prototypes instead of using existing systems. By creating our

own concepts and videos we were able to push the design dimensions in specific directions, often exploring their extremes in new combinations. All together, the concepts synthesize a mix of disparate ideas into a cohesive collection, applying existing and proposed design approaches in new contexts. In the videos, we tailored the user scenarios around key research questions we were interested in exploring with participants. The videos, similar to experience prototypes [12], frame the concepts in a way that offer glimpses into possible futures in order to provoke and open dialogue with participants about perceived benefits and consequences of each design [51]. Our approach is inspired by and shares similarities with prior work on *User Enactments* [61] and *Speed Dating* [17]. These related approaches argue for exploring the potential roles, values, and social boundaries of emerging near-future technology by using more than one design vision. They encourage participants to imagine future interactions and react to them by drawing on their own experiences.

Design concepts

We now describe the five design concepts, pointing to sources of inspiration. Because of space, we do not detail their design process (i.e., technical implementation, visual choices). Instead, we focus on positioning them within the design space.

Patina

The first concept, Patina, is a visualization on top of data in the geometric form of a spiral. It is inspired by a tree's growth circles and symbolizes temporal qualities of data. In the video for Patina, we show two different options for the spiral: on a desktop, it represents the age of folders (Figure 1); with a set of music playlists, instead, it represents the number of interactions over a period of time (Figure 2). Music playlists provide a good contrast to folders because they are “meant to be enjoyed repeatedly and grown over time.” [49]

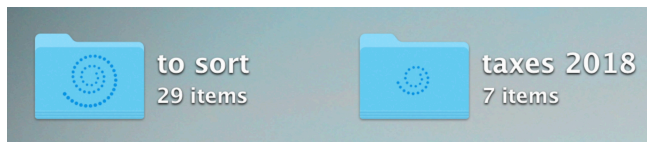


Figure 1. Patina showing the age of desktop folders: the bigger the spiral, the older a folder is. (Each dot stands for a set time amount, e.g., a week.)

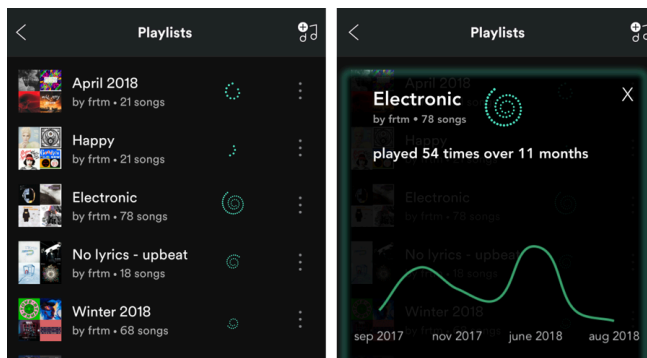


Figure 2. Patina showing the frequency of use for music playlists: the bigger the spiral, the more times a user has played the playlist.

Patina's video leaves some aspects as intentionally ambiguous and unexplained (e.g., How is the age of a folder calculated? How is the interaction period determined?). We wanted to encourage user interpretation and discussion. This choice also emphasizes the open-ended nature of Patina in the design space: it invites reflection and builds awareness, but does not suggest any specific action to take on data. We use this concept to probe on the viability of open-ended designs that leave users in charge of initiating any selection action and on what metadata attributes might be useful for doing so.

The idea of a *patina* take inspiration from prior studies that mention its potential [62] or use it with physical [23, 43, 44] and digital objects [35, 54]. The frequency of use in Patina is inspired by Hurst *et al.* [35] and Matejka *et al.* [54]. The idea of aging, instead, is informed by Giaccardi *et al.*'s work on “traces of use” for daily objects [23]. Our work, however, takes place in a different context and uses a different approach designed with data selection in mind. There are also commercial products with visualizations for used space on hard disks (e.g., Daisy Disk (<http://daisydiskapp.com>), Disk Inventory X (<http://www.derlien.com>)) but these are separate from the data and only use one type of metadata (size). Instead, we tie the visualization to the data and use two types of metadata.

Data Recommender

The second concept, Data Recommender (Figure 3), notifies users and provides recommendations on data that might need attention, using metadata like last access, creation date, or size. Users can decide to trash items, archive them in a central archive, move them in a specific folder, or be reminded of them at another time. Data Recommender will use machine learning to learn from their actions and provide new recommendations. This concept is the closest to existing products. For example, Google Photos (<https://www.google.com/photos/about/>) provides recommendations on photos to archive, while Files on Android (<https://files.google.com>) gives recommendations on how to free up space.

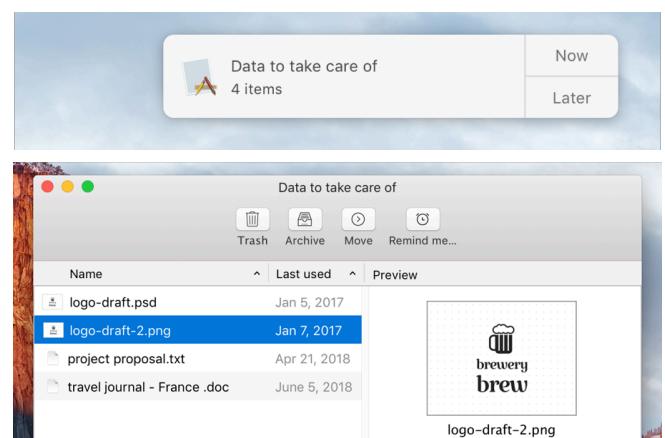


Figure 3. Data Recommender notifies users when they have some data to take care of (top) and provides a list of items (bottom): users can choose to trash, archive, move, or be reminded of items again.

Based on prior work [63], recommender systems might be a good approach to offload the work of selecting items. Data Recommender is in the middle between human-driven activity

and automation, following a *mixed-initiative* approach [13,33]. Using this concept we want to probe on the link between data and context, the viability of different selection actions, and the attributes that make items good candidates for disposal.

Temporary Folder and Temporary App

The next two concepts come as a couple: Temporary Folder and Temporary App. In this case, we created two videos on two different platforms. The first, Temporary Folder, takes place on a desktop computer (Figure 4): it acts like a standard folder, but users can decide to set an expiration date for it. After the expiration date, the folder will be automatically deleted. The second, Temporary App, takes place on a smartphone (Figure 5). In this case, users can install a mobile application temporarily (e.g., for two weeks). At the end of the preset period, the application will be automatically uninstalled.

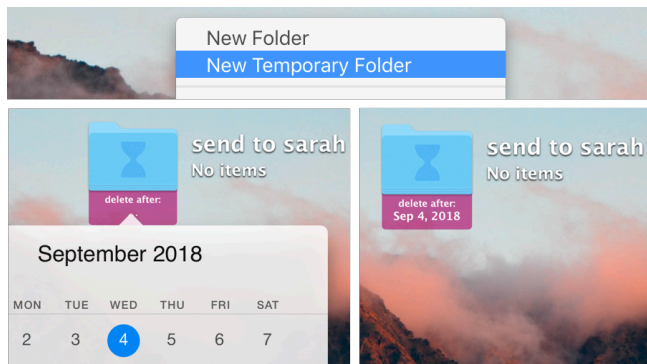


Figure 4. When creating a Temporary Folder (top) users can pick an expiry date for it (bottom, left). After the expiry date (bottom, right), the folder is automatically deleted.

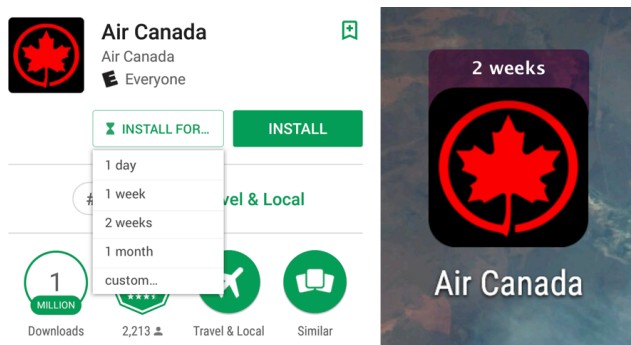


Figure 5. When installing a Temporary App users can pick an expiry date for it. After the expiry date, the app is automatically uninstalled.

These temporary concepts fall within the prospective side of the design space, unlike the previous two. Their radical take on automatic deletion was meant to stimulate discussion among our participants about their perceived social acceptability. Several commercial products already use automatic deletion in specific contexts. As an example, the messaging application Telegram (<https://telegram.org>) allows users to set an expiry date for photos, videos, and other files exchanged with contacts. If they do not access them for a set period of time they are removed from the device (they are still on Telegram’s cloud though, so they are not completely deleted). Also, the notion of different information lifespans goes back to an early study

of desktop usage that identified three types of information [2]: *ephemeral*, *working*, and *archived*. Temporary Folder and Temporary App explore this idea in two specific contexts.

Future Filters

The final concept, Future Filters, is a mobile application that allows users to decide what to do with data in the future creating set of rules or filters. For example, “delete selfies and downloads that are older than two months when my free space is below 20%,” (Figure 6) or “archive shared documents not looked at in 2 years,” and so on. Filters use a set of actions (e.g., delete, move, archive, remind me), criteria (size, use, number of copies, source of data, copied on the cloud, etc.), and triggers (a new update available, free disk space is below a certain amount, etc.).

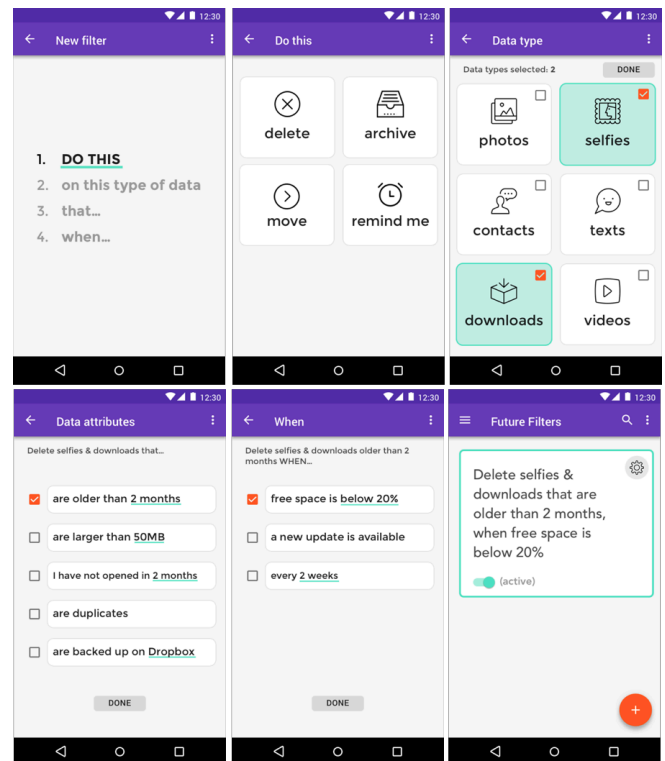


Figure 6. Future Filters is a mobile application that lets users create data filters based on actions, data types, data attributes, and triggers.

This final concept has a strong emphasis on prospective decisions and mass processing of items, with a certain degree of automation. It takes direct inspiration from *If This Then That* (<https://ifttt.com>), a platform to create cross-application rules based on triggers, and other products or features that use automatic filters (e.g., File Juggler (<https://www.filejuggler.com>), Hazel (<https://www.noodlesoft.com>), Gemini (<https://macpaw.com/gemini>), or email filters in Gmail). We use Future Filters to further explore the viability of prospective decisions, probing on what actions might be more acceptable when considering automation.

ELICITATION STUDY

We used our set of design concepts in an elicitation study with 16 participants. We showed them the videos of the concepts

during one-on-one interview sessions that also touched on their general data management practices.

Recruitment and participants

We used purposive sampling to recruit a diverse sample of participants. We advertised the study on a university listing and on Craigslist in Vancouver, Canada. We used a screening questionnaire (see supporting materials) to select participants based on their age, occupation, technical familiarity, and general approach to data curation. In a parallel study focused on building a model of individual differences in PIM, we identified a taxonomy of user archetypes or approaches to personal data curation. In short, these approaches differ based on how much data people tend to keep or discard (a lot, a little, in the middle), their general organization practices (structured, unstructured, a mix), and their feelings about their current data management state (happy, unsatisfied, unconcerned). We used brief descriptions of these general approaches in the screening questionnaire as a closed-ended question.

We received 177 responses to the screening questionnaire. We contacted 36 respondents and 16 agreed to take part in the study. We stopped recruiting when we reached a diverse set of participants and reactions. Nine participants self-identified as female, six as male, one as gender non-conforming. They were aged 23–71 (average: 36). Occupations included administrative assistant, engineer, HR specialist, journalist, photographer. In terms of general approach to data curation, four tended to keep most of their data and be happy about it, five tended to keep most but were not happy about it, three tended to only keep necessary things, four tended to keep some, delete some without big concerns. We note that this was the general self-reported approach, but during the interviews participants elaborated on their approach, displaying some differences among data types and more nuanced behaviors. Our diverse sample is not meant to be statistically representative but instead generative. Similarly, we look at digital data with a broad lens, including a diverse set of data types, based on what “stuff” participants consider to be their own data. This approach is in line with recent studies on data management [72, 73].

Procedure

The study sessions consisted of 1) a short introductory interview on general data management practices, 2) a main elicitation section going over each of the design concepts, 3) a final, longer semi-structured interview discussing and comparing all the design concepts and the ideas behind them. In the introductory interview, we asked participants to discuss how they organized and selected their data over time and on different platforms or devices, asking them to show us examples where possible. Then, for each concept, we first gave a short introduction and then showed the video. After each video, we asked whether something was not clear, providing printouts of the concepts. Then, we probed on participants’ first impressions, asking them how they felt when watching the video and what they felt about different aspects of the concept. Following Odom *et al.*’s approach [61], in the final interview, we asked participants to reflect on and across their experiences of all concepts. We asked them to pick the most or least valuable for them, discuss the most positive or negative aspects across

all the concepts, and elaborate on the ideas behind them based on how they would fit their needs and experience. Participants used the printouts of the systems to compare and contrast them. One member of the research team conducted all interviews, in English, at our university. Interviews lasted between 37 and 70 minutes (on average: 49 minutes) and were audio-recorded. Participants received \$15 as compensation.

Data analysis

To analyze the data, we used Braun and Clarke’s approach to thematic analysis [16]. We transcribed participants’ answers and started analyzing them inductively using open coding. Then, we grouped codes into categories and identified themes across categories. One member of the research team coded the data and discussed the themes and interpretations with other authors during multiple meetings.

THEMATIC ANALYSIS RESULTS

In this section, we present the results of our thematic analysis. In general, participants appreciated the idea of getting help in selecting data—they saw it as an important but often challenging task. However, there were striking differences in how they reacted to the concepts. For example, reactions to Temporary Folder ranged from enthusiastic (P15: “*I like that one, a lot!*”) to perplexed (P16: “*Why would one want a temporary folder?*”) and terrified (P1: “*I would be terrified to put something in a folder that’s going to be deleted!*”). In the first two themes of the analysis, we contrast diverging opinions on what role should technology have in supporting selection practices: some participants preferred to retain full control of the process (Theme 1). Others, instead, welcomed automation and felt comfortable in offloading selection tasks to technology (Theme 2). Then, we synthesize a middle ground between the two different stances in participants’ overall desire for a contextual approach and how this leads to a new perception of keeping and discarding actions (Theme 3).

Theme 1: Selecting data is a personal responsibility

The first theme captures opposing reactions towards support in selecting data. Some of the participants were generally against automation, mass processing of items, or aggressive systems. Their reactions to the design concepts highlighted a need for control, a sense of responsibility for their data, and a strong desire for doing things in a specific way (the “right way” [72]), all on their own.

Wanting full control

For instance, P11, a professional photographer who managed thousands of photos between her phone, laptop, and an external hard drive, made it clear that fully automatic tools would not work for her because they crossed an important boundary. She felt that having full control over data and the selection process was essential. Her career depended on properly managing digital data, with no room for mistakes: “*I’ve heard the horror stories of photographers not backing up data properly and losing up a whole shoot, and, yeah, it will pretty much just ruin your reputation.*” She explained how automatic tools felt intrusive and undermined her sense of control: “*For data, and maybe this is my personality or work, but you don’t ever want somebody coming in and telling you ‘this is mine’. Or, ‘get*

rid of it'. You make the work, you wanna have control over it. That's why I wouldn't want something like Future Filters going through my files. [...] I really I don't appreciate that."

Thinking independently

Similarly, P13, who used to work as a "programmer of sorts" in a medical imaging company, highlighted that thinking independently and taking care of items without the help of technology was important to feel in control. In her reasoning, she drew a parallel between some of the concepts and older recommender systems in Word processors (i.e., Clippy): "It's almost like, you know—there were Word processors that tried to think for you. 'Oh, it looks like you're writing a document, let me do such and such.' And I'm like so mad. I did a lot of Word processing. And I know what I want, I know the spacing I want, I know the editing I want. I know what I'm trying to do and this nuisance tries to think for you. I don't like that. [...] Personally, I would like more power and control myself." After seeing Temporary App, she added that automatic tools made her feel lazy, hinting once again at the idea that selecting data is a personal responsibility: "It's kind of like, you feel lazy. Because how hard is it and throw it into the delete stuff? Are we all that busy that we need [this]?"

Distrusting technology

Underlying many statements from participants there was a sentiment of distrust towards technology. For example, participants questioned how the machine learning from Data Recommender would work and whether it would learn the "wrong things." Similarly, they feared that any function where they did not have full control would eventually go wrong. Thus, a lingering feeling of uneasiness. "I tend to like the ones that remind you or prompt you vs automatically doing it for you," explained P7, an HR specialist who did not trust any cloud platforms for personal data management preferring to do things on "her own." She articulated her preference for Patina and Data Recommender in terms of trust and comfort: "I wouldn't feel comfortable putting in parameters and just having the technology determine for me. I'd prefer to have them notify me or go through them and choose. I liked [Data Recommender]. I'd feel much more comfortable with that vs having things automatically deleted."

These preferences were not always a direct reflection of differences in general curation approaches (e.g., tending to keep a lot vs. keeping little). Consider the case of P15, working as an administrator in a government agency: she self-described as a "very organized" person who "doesn't keep a lot of things," limits her technology use, and deletes photos, videos, files. She was onboard with discarding data, but trusted herself more than a tool: "I'm not someone who keeps everything. So I'm not at all reluctant to delete things at once, I know people who are. But at the same time, I don't know if I would trust the computer to delete things if I haven't reviewed them and made sure I want to delete them." This explains why she was enthusiastic about Temporary Folder but not Future Filters: "I have control over what I'm putting in the folder. [With Future Filters], you don't really know exactly what it's deleting, you're just trusting that you're putting things in the right place so I feel there's more potential for errors with [Future Filters]."

Changing idea and having the final say

The need to have full control and think independently informed a strong preference towards always having the final say in all keeping or discarding decisions. Being in control meant seeing recommendations as nothing more than suggestions that need approval and leaving space for changing ideas. When pondering prospective decisions, participants who were generally against automation felt uneasy and wondered what would happen if they changed their mind: "I think I like the concept of [Future Filters] but it's such broad categories that you might end up deleting data and regret it." (P8). Anticipating regret, some participants said that keeping everything "just in case" [71, 73] seemed a better approach, while others saw a safeguard in the possibility of controlling decisions and having a final say. For example, P9, a journalist who used to deal with large amounts of data but who wanted to be "more organized" and "have less," explained that "Across the board, the review process is really important: before something gets deleted I should know it's getting deleted, it should not get deleted without me knowing. And I should have the physical option of choosing to delete or not. [...] I should have the final say. [...] Sometimes you change your mind."

Theme 2: Selecting data is a chore

The second theme captures reactions towards the concepts that contrast those explored in Theme 1. In this case, participants welcomed automation and generally expressed the need for tools that would take care of things for them, freeing them from the weight of selecting data. They felt tired of the responsibility of the selection process, something they put off or were not good at dealing with, and were happy to offload the process to technology.

Being tired of taking care of data

Participants who had positive views of automation were happy about tools taking care of selecting data, a process that they perceived as tiring and relentless. For example, P14, an HR specialist, said she "seriously" loved Future Filters and explained that it would help her deal with things she was tired of: "I am tired of organizing my information and taking care of it every 2-3 months because of the space limit. It bothers me a lot, so if I can set the filters just once for the majority of things that bother me—and it would be pictures, videos, and music, that's the main problem—that would be just perfect."

Needing a push to feel the urgency

Participants noted how they needed a "push" to attend to data: "I thought the most useful [aspect of the system] was that it actually popped up. So you have to actually take action on all of the items, because that forces you to decide what to do with them. I thought that was very useful." (P10). Selecting items was a task they wanted to engage in, but often put off: "I feel this could be a cool way to do it for me, because it's something I put off. The last time was probably a year ago or something, so having an app do it for me would be great." (P2) More automatic and prospective systems felt useful in creating urgency: "I think Patina [was the least valuable]. Even though it sets an indication, it doesn't create an immediate urgency" (P6).

Desiring a proactive system

Some participants were satisfied enough with recommendations, but several had preferences for stronger intervention, expressing the desire for a proactive system that would think for them: *“It’s perfect if the program can think for me in advance. [...] The [Data] Recommender is going to bother me for sure. It means the program advises me to think about something and I want the program to think in advance, give me some kind of solution.”* (P14) This preference might have come down to personal style. This is how P10, a student in Education who reported constantly running out of space on his laptop, related tools like Future Filters to his self-described “lazy” selection style: *“I think it depends how organized you are. Future Filters automatically deletes without telling you what. Patina and Data Recommender remind you and you decide what to do. So, if you feel like you need that reminder and you can delete yourself, I think they would be nice. If you feel you’re too lazy or not organized enough, Future Filters takes care of it for you. I am less organized, that’s why Future Filters is a really good option for me.”*

Deciding in advance to not worry later

The enthusiasm for simplifying selection extended also to prospective decisions, with participants relating the idea to their own practice: *“There’s definitely things I know I don’t need. You know, pictures from the internet you want to send someone and they stay on your desktop.”* (P4). Several participants preferred to decide in advance and not worry later: *“I think it’s a good idea in terms of coming up with some parameters for things you know you’re not going to need in the future and it’s better to just automatically delete it and not worry about it”* (P7). They perceived such options as a way to limit the constant input required for selection, a process they compared to daily chores: *“[Future Filters] might be better because I’ve made a decision and then it will happen and it’s not being dependent on me being... you know, it’s like cleaning or doing dishes. [...] Your input is at the beginning and then it automatically takes care of itself.”* (P13) However, as the next theme shows, these decisions still needed to have some safeguards in place or otherwise respect the context of data.

Theme 3: Context is key

In Themes 1 and 2, we have described a range of reactions to the concepts, with two general contrasting stances. The final theme highlights how these reactions were different but never completely polarized, because the context of data played a key role. Participants noted important differences between data types based on their nature (a document you take a lot of time to create vs. a movie or an app, that you can always download again), their context (work data being generally more important and critical than personal data), and the device under consideration (computers being for serious stuff and smartphones being for less critical stuff). They perceived data as being always somewhere out there, in the cloud or on a device: this had both positive and negative consequences, and informed what we call a *post-cloud* perception of selection.

Selection decisions are contextual

A recurring thread in participants’ reactions, whether more positive or negative, was that keeping or discarding decisions

are highly contextual. Thus, a concept that worked for one type of data, might not have worked for others. For example, many participants drew distinctions between work and personal “stuff,” saying that they tended to be more organized and less selective with what to keep at work: *“I keep everything for work related, for personal it’s different.”* (P7) Similarly, they regarded data on phones as easier to discard and often less important. They perceived smartphones and mobile devices as “fluid, temporary, and more accessible,” compared to laptops, that were “serious, demanding,” and with more places where to hide things away. The difference in storage capabilities between the two types of devices was also an important factor to consider: *“For the phone I periodically delete pictures I don’t want and apps I don’t use anymore. But on the phone it’s partly due to storage problems. Which is not that much of a concern with my Mac”* (P3).

Exceptions to general decisions

Participants also remarked that digital data being old or unused did not necessarily mean that they would have liked to get rid of it, as some concepts suggested: *“I don’t like the fact it says you haven’t used it because it might say, you haven’t used it in six months, get rid of it. But that’s not a good idea because sometimes you save files for you future situations. Deleting files because they are older, is that a good idea? Maybe there’s a reason they should be kept.”* (P16) They wanted to define exceptions to general decisions and have the option to instruct the system about any item that they might want to keep: *“Maybe there’s an option to exclude certain things. Like, all photos that are older than 2 weeks, except these three. That’d be a good option to have, to be able to create exceptions.”* (P10). These reactions point to the importance of marking items to keep explicitly, an action often absent from data management tools.

The cloud is as big as the universe

The contrast between different contexts and storing places was particularly evident when comparing Temporary Folder and Temporary App. Several participants explained that automatic or prospective actions were more acceptable with mobile applications because applications are not unique and are not the result of time or effort: *“There’s no big risk, you can install the app again”* (P5). When expanding the focus within their data ecosystem [72] and discussing the cloud, with places like Facebook, Google Drive, or Dropbox, participants noted how the change in context changed their attitude, hinting that selection would be less necessary in these storing places: *“I never delete [from Facebook] because I imagine their storage is as big as the universe.”* (P6)

Data is always somewhere out there

The key role of cloud platforms and the interconnected nature of data ecosystems informed a perception of data as ubiquitous and perennial. Participants in Gulotta *et al.*’s study [26] saw deleting as being against the nature of digital data. Participants in our study perceived data as never truly deleted because there will always be a copy somewhere, out there. Perhaps it is a copy on an external hard drive, maybe it is a backup on Facebook or Google, but data never really disappears unless you want it to: *“We’ve talked about deleting apps on my*

phone regularly and things and as long as they're backed up somewhere, if they're deleted it's not a big deal.” (P2). This post-cloud conception of data makes keeping and discarding decisions take on a different meaning: deleting means removing data only from one specific device or removing a specific instance, while having a copy somewhere else: “I don't think I will put anything in Temporary Folder that I don't have a backup for, so it's fine.” (P8). And archiving really means moving or hiding data within a device ecosystem: “I wonder how the archive... maybe it's like moving, it sounds like a similar function. I [archive] with my emails, but I think it goes to... it's the same idea as moving.” (P6). Suddenly, automatic or prospective decisions are more acceptable: “Especially in the selfie scenario, you probably already posted it on Instagram or Snapchat or whatever, so there's a copy of it already in the world, so removing it from your device it's not a big deal.” (P5). The selection process then becomes a matter of moving data back and forth within an ecosystem and the cloud is the ultimate storage utility.

DISCUSSION AND FUTURE DIRECTIONS

Moving the design space towards personalization

The range of reactions we gained from participants supports the idea that decisions around what data to keep or discard are highly personal [73]. As we expected, no single solution was able to resonate with most participants. But these results show how branching into potentially controversial or radical areas of the design space can be fruitful. By inquiring into concepts that appeared risky, we were able to get a better idea of where people's boundaries lie, what it means to cross them, and how different people may have different boundaries. The design dimensions we explored can now be used as a generative resource to work towards new solutions. Some of the design dimensions and concepts we explored could be remixed (e.g., providing a list of filters that can feed into recommendations) and modulated to support different user attitudes. There is an opportunity for future work to further investigate this emerging space through designing, developing, and studying more personalized solutions (e.g., customizing default keeping and discarding actions or criteria for recommendations). In the following sections, we articulate some more key directions for future efforts.

Finding a space for automation

In our analysis, we were particularly fascinated by the contrasting attitudes towards automation. Do the strong reactions against some of the concepts mean that we failed as designers to support user needs? We think the key here lies in identifying the underlying threads of such negative opinions and leverage them to move towards a more nuanced approach. As highlighted in the related work, the tension between automation and user control is a long-standing issue. A key contribution of our work is revealing that some keeping decisions, under specific circumstances, can likely be automated. In particular, promising initial contexts to pursue automation in design interventions are mobile devices with limited storage space, media files, and distributed data (i.e., data that is not unique or otherwise re-accessible). In other cases (e.g., different devices and types of data), there still is a space for automation, but

only with proper safeguards. This principle extends to other dimensions of the design space, as we outline next.

Safeguarding automatic and prospective decisions

Our findings reveal that there can be a space for both *retrospective* and *prospective* actions, *manual* and *automatic*, and *open-ended* or *specific*. However, it is essential that future design interventions synthesize these extremes so that any action is reversible and any potential risks are mitigated in advance. This suggests an opportunity to investigate how to design effective safeguards for automatic and prospective actions. The easiest way to design safeguards would be to provide reminders before automatic or prospective actions, something that several participants asked for. Another approach would be to simply promote softer actions over the more radical concept of deletion (e.g., moving, trashing, hiding). Yet another opportunity would be to see the perceived risk and anticipated regret that come up in participants' words as explicit components of the decision process. For example, systems could visualize a history of prospective or reverted decisions (e.g., how many times a document was marked for trashing and then reverted, or how many times a mobile app was uninstalled and then re-installed over the course of a period of time). Similarly, efforts along the other dimensions of the space could focus on letting users explicitly define potential risks and regrets and then evaluate them at a later point in time.

Rethinking keeping and discarding actions

When reacting to the concepts and the actions they afforded, some participants struggled with understanding what an archive is. For others, archiving was the same as moving or hiding. Similarly, participants reported deleting practices rooted in the importance of context and the availability of a multitude of platforms and devices. Sas *et al.* [68] argue that “deletion is a crude binary process,” while Ramokapane *et al.* [66] highlight how cloud platforms in particular provide poor deletion models. We agree that a binary representation of data as either present or deleted does not reflect the majority of our participants' mindset. This argument ties to *prospect theory* [39] and can further explain why keeping decisions are so challenging: if deletion is a binary process, it is more difficult to balance risks and gains. Yet, in most tools deleting is the default discarding action. Although our results show that crude deletion is welcomed in some cases (e.g., with mobile applications), moving towards a mitigated process of deletion might be the way forward to support different contexts. This idea resonates with work by Harper *et al.* [31] who argue for rethinking actions about owning, copying, and deleting data.

Based on these implications, we see two possible directions to follow. The first is a design-focused effort in the line of work by Lindley *et al.* [48] to explore and define a new grammar of actions around keeping decisions. Harper *et al.* [31] propose “eradicating” or “withdrawing” files from the cloud, while Bergman *et al.* [6–8] show examples of “demoting.” This set of actions could be extended to include *mirroring* (for storing a copy of data from a central repository only temporarily), *distributing* (to disseminate copies of data around many storing places), *warranting* (to authorize automatic tools to act only on specific items), *locking* (to mark items as protected

by any discarding action). Providing options for discarding actions to be granular rather than binary will also allow for more personalized solutions. For example, some users might choose “crude” deletion, while others might opt for demoting or distributing items as their primary selection action. A second, research-focused effort would be to further disambiguate actions as used in interfaces and perceived by users with a taxonomy. This would follow the example of Watkins *et al.* [76], who categorize very specific different types of digital “collections” and collecting practices. As researchers, we keep referring to the old metaphor of an archive to study and discuss people’s practices. It seems like apt timing to take more seriously the question of what is an archive in the post-cloud age and is this the best metaphor we can use?

Taking steps towards active data privacy protection

A tangential but important issue that came up in our exploration was the topic of privacy and security. This was not our focus, but we inevitably touched on it. Participants discussed how the concepts could work for managing privacy and security, both on their devices and in the cloud. Their attitudes varied. In general, they perceived the concepts as acceptable if they came from trusted brands, were officially part of the operating system, or were looking at data on devices more than on cloud platforms (possibly because cloud platforms are “curated through use” [84]). But often participants noted how keeping decisions are more delicate and consequential when it comes to privacy. Nudges, reminders, and prospective actions could prevent unwanted issues or “leaks” of sensitive data.

These results resonate with the public’s desire for more control over data [50] in the face of recent data scandals [14, 25]. The advent of the cloud has imposed a centralized data management model where a few corporations (Amazon, Apple, Google, Facebook, Microsoft) aggregate the large bulk of people’s data. But as Mortier *et al.* argue [56], this approach is “fundamentally flawed” and as the importance of digital data continues to grow, it faces increased scrutiny. We argue that a contextual and user-driven approach to keeping decisions in the cloud can be a concrete step in protecting privacy. In particular, our ideas around safeguarding and rethinking keeping decisions can be extended to a privacy-oriented mindset to provide more control to users. There is an opportunity for future work to target similar ideas exclusively around data privacy management and better explore people’s attitudes. This approach would fall in line with recent work on “design workbooks” for privacy [83]. Another possibility in this domain is to study how to use similar concepts for data created *about* people (e.g., advertising data). Opportunities include allowing users to create temporary advertising profiles or review and discard any tracking data that companies have on them.

Reflecting on the broader impact of our work

Finally, we reflect on the broader impact of our work for individuals and society [32]. On one side, we hope to inspire positive change in the space of data management and selection, pushing towards a nuanced approach and truly user-centred designs. At the same time, we see how some of the concepts we propose might lead to unintended consequences and be abused to further centralize data management and restrict

users’ freedom. For example, Temporary Folder and Temporary App could be used to restrict users’ access to their own cloud-stored data, imposing a subscription model to items that they perceive as their own. Indeed, we see this trend already emerging in several software applications, as P11 lamented when discussing her use of the Adobe Creative suite. Business needs drive these decisions, yet, this approach may contribute to eroding people’s sense of ownership and agency in relation to their data. This, in turn, feeds their general distrust for technology. We argue for an alternative, de-centralized data management model, where users’ control is key, privacy comes first, and management actions are context-based. New regulations like the European Union’s GDPR (General Data Protection Regulation) provide a first step towards more ethical practices in the space of personal data, but regulatory efforts need to be complemented by design changes. It is up to us, as researchers and designers, to ensure that the needs we discuss around personal data are met and that people’s boundaries are respected.

LIMITATIONS

Some limitations in our study point to additional future work. First, while our sample is meant to be generative and varied in terms of occupations and data management styles, participants had a predominantly Western background. This limitation is an opportunity for future work to focus on participants from different cultures, to see if and how attitudes change. We also did not screen participants for more general attitudes around decision making or psychological traits, because finding correlational links was not the goal of this study. But there might be individual differences in risk aversion and risk seeking that inform people’s attitudes in deciding what data to keep or discard. A previous study by Massey *et al.* [53] links personality traits to differences in file management behaviors. Similar efforts along this line can complement our work.

CONCLUSION

Drawing on previous work on personal data management, we created five concepts to explore a design space around keeping and discarding decisions. By probing on different design dimensions, we elicited contrasting attitudes about the role of technology in supporting decisions, finding a common ground in the need for nuanced and contextual support. Our work opens possibilities for new tools that, with proper safeguards, have the potential to help users better select what personal data to keep or discard. We see this as a critical step in addressing our *post-cloud* world that is overflowing with data. We hope that our work will inspire change and move future research and practice towards new approaches that will improve the management of personal data.

ACKNOWLEDGMENTS

We thank the anonymous reviewers, Paul Bucci, Sabrina Hauser, Michael Opperman, and Hanieh Shakeri for their feedback. We created the concepts using Sketch and InVision with icons from the Streamline library (<https://streamlineicons.com>) and The Noun Project (<http://www.thenounproject.com>) (by Abdo, AR Ehsan, Dino, Hare Krishna, and Landan Lloyd). This work was supported by the grant NSERC RGPIN-2017-04549 “Highly personalized user interfaces.”

REFERENCES

- [1] Jakob Bardram, Jonathan Bunde Pedersen, and Mads Soegaard. 2006. Support for Activity-based Computing in a Personal Computing Operating System. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 211–220. DOI: <http://dx.doi.org/10.1145/1124772.1124805>
- [2] Deborah Barreau and Bonnie A Nardi. 1995. Finding and reminding: file organization from the desktop. *ACM SigChi Bulletin* 27, 3 (1995), 39–43.
- [3] Victoria Bellotti, Nicolas Ducheneaut, Mark Howard, and Ian Smith. 2003. Taking Email to Task: The Design and Evaluation of a Task Management Centered Email Tool. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, New York, NY, USA, 345–352. DOI: <http://dx.doi.org/10.1145/642611.642672>
- [4] Ofer Bergman. 2012. The user-subjective approach to personal information management: from theory to practice. In *Human-computer interaction: the agency perspective*. Springer, 55–81.
- [5] 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. DOI: <http://dx.doi.org/10.1002/asi.10283>
- [6] Ofer Bergman, Oded Elyada, Noga Dvir, Yana Vaitzman, and Adir Ben Ami. 2015. Spotting the Latest Version of a File with Old'nGray. *Interacting with Computers* 27, 6 (2015), 630–639.
- [7] Ofer Bergman, Andreas Komninos, Dimitrios Liarokapis, and James Clarke. 2012. You Never Call: Demoting Unused Contacts on Mobile Phones Using DMTR. *Personal Ubiquitous Comput.* 16, 6 (Aug. 2012), 757–766. DOI: <http://dx.doi.org/10.1007/s00779-011-0411-3>
- [8] Ofer Bergman, Simon Tucker, Ruth Beyth Marom, Edward Cutrell, and Steve Whittaker. 2009. It's Not That Important: Demoting Personal Information of Low Subjective Importance Using GrayArea. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 269–278. DOI: <http://dx.doi.org/10.1145/1518701.1518745>
- [9] Ofer Bergman and Steve Whittaker. 2016. *The Science of Managing Our Digital Stuff*. MIT Press.
- [10] 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. DOI: <http://dx.doi.org/10.1145/985692.985766>
- [11] R. N. Brewer, M. R. Morris, and S. E. Lindley. 2017. How to Remember What to Remember: Exploring Possibilities for Digital Reminder Systems. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 38 (Sept. 2017), 20 pages. DOI: <http://dx.doi.org/10.1145/3130903>
- [12] Marion Buchenau and Jane Fulton Suri. 2000. Experience Prototyping. In *Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS '00)*. ACM, New York, NY, USA, 424–433. DOI: <http://dx.doi.org/10.1145/347642.347802>
- [13] Andrea Bunt, Cristina Conati, and Joanna McGrenere. 2007. Supporting Interface Customization Using a Mixed-initiative Approach. In *Proceedings of the 12th International Conference on Intelligent User Interfaces (IUI '07)*. ACM, New York, NY, USA, 92–101. DOI: <http://dx.doi.org/10.1145/1216295.1216317>
- [14] Carole Cadwalladr and Emma Graham Harrison. 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach | The Guardian. (17 March 2018). Retrieved January 16, 2019 from <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>
- [15] EunJeong Cheon and Norman Makoto Su. 2018. The Value of Empty Space for Design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 49, 13 pages. DOI: <http://dx.doi.org/10.1145/3173574.3173623>
- [16] Victoria Clarke and Virginia Braun. 2014. Thematic analysis. In *Encyclopedia of critical psychology*. Springer, 1947–1952.
- [17] Scott Davidoff, Min Kyung Lee, Anind K Dey, and John Zimmerman. 2007. Rapidly exploring application design through speed dating. In *International Conference on Ubiquitous Computing*. Springer, 429–446.
- [18] Paul Dourish, W. Keith Edwards, Anthony LaMarca, and Michael Salisbury. 1999. Presto: An Experimental Architecture for Fluid Interactive Document Spaces. *ACM Trans. Comput.-Hum. Interact.* 6, 2 (June 1999), 133–161. DOI: <http://dx.doi.org/10.1145/319091.319099>
- [19] Daniel A. Epstein, An Ping, James Fogarty, and Sean A. Munson. 2015. A Lived Informatics Model of Personal Informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 731–742. DOI: <http://dx.doi.org/10.1145/2750858.2804250>
- [20] Scott Fertig, Eric Freeman, and David Gelernter. 1996. Lifestreams: An Alternative to the Desktop Metaphor. In *Conference Companion on Human Factors in Computing Systems (CHI '96)*. ACM, New York, NY, USA, 410–411. DOI: <http://dx.doi.org/10.1145/257089.257404>

- [21] Leah Findlater and Joanna McGrenere. 2004. A Comparison of Static, Adaptive, and Adaptable Menus. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. ACM, New York, NY, USA, 89–96. DOI: <http://dx.doi.org/10.1145/985692.985704>
- [22] Roxana Geambasu, Tadayoshi Kohno, Amit A Levy, and Henry M Levy. 2009. Vanish: Increasing Data Privacy with Self-Destructing Data.. In *USENIX Security Symposium*, Vol. 316.
- [23] Elisa Giaccardi, Elvin Karana, Holly Robbins, and Patrizia D'Olive. 2014. Growing Traces on Objects of Daily Use: A Product Design Perspective for HCI. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14)*. ACM, New York, NY, USA, 473–482. DOI: <http://dx.doi.org/10.1145/2598510.2602964>
- [24] Andreas Girgensohn, Sara A Bly, Frank Shipman, John S Boreczky, and Lynn Wilcox. 2001. Home Video Editing Made Easy-Balancing Automation and User Control.. In *INTERACT*, Vol. 1. 464–471.
- [25] Glenn Greenwald and Ewen MacAskill. 2013. NSA Prism program taps in to user data of Apple, Google and others | The Guardian. (7 June 2013). Retrieved January 16, 2019 from <https://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data>
- [26] 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. DOI: <http://dx.doi.org/10.1145/2470654.2466240>
- [27] Rebecca Gulotta, Alex Sciuto, Aisling Kelliher, and Jodi Forlizzi. 2015. Curatorial Agents: How Systems Shape Our Understanding of Personal and Familial Digital Information. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3453–3462. DOI: <http://dx.doi.org/10.1145/2702123.2702297>
- [28] Jacek Gwizdka. 2000. Timely Reminders: A Case Study of Temporal Guidance in PIM and Email Tools Usage. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems (CHI EA '00)*. ACM, New York, NY, USA, 163–164. DOI: <http://dx.doi.org/10.1145/633292.633383>
- [29] Jacek Gwizdka. 2001. Supporting Prospective Information in Email. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems (CHI EA '01)*. ACM, New York, NY, USA, 135–136. DOI: <http://dx.doi.org/10.1145/634067.634150>
- [30] Jacek Gwizdka. 2002. Reinventing the Inbox: Supporting the Management of Pending Tasks in Email. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems (CHI EA '02)*. ACM, New York, NY, USA, 550–551. DOI: <http://dx.doi.org/10.1145/506443.506476>
- [31] Richard Harper, Siân Lindley, Eno Thereska, Richard Banks, Philip Gosset, Gavin Smyth, William Odom, and Eryn Whitworth. 2013. What is a File?. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, New York, NY, USA, 1125–1136. DOI: <http://dx.doi.org/10.1145/2441776.2441903>
- [32] B. Hecht, L. Wilcox, J.P. Bigham, J. Schöning, E. Hoque, J. Ernst, Y. Bisk, L. De Russis, L. Yarosh, B. Anjum, D. Contractor, and C. Wu. 2018. It's Time to Do Something: Mitigating the Negative Impacts of Computing Through a Change to the Peer Review Process. (29 March 2018). Retrieved January 16, 2019 from <https://acm-fca.org/2018/03/29/negativeimpacts/>
- [33] Eric Horvitz. 1999. Principles of Mixed-initiative User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, 159–166. DOI: <http://dx.doi.org/10.1145/302979.303030>
- [34] Jina Huh, Martha Pollack, Hadi Katebi, Karem Sakallah, and Ned Kirsch. 2010. Incorporating User Control in Automated Interactive Scheduling Systems. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems (DIS '10)*. ACM, New York, NY, USA, 306–309. DOI: <http://dx.doi.org/10.1145/1858171.1858226>
- [35] Amy Hurst, Jennifer Mankoff, Anind K. Dey, and Scott E. Hudson. 2007. Dirty Desktops: Using a Patina of Magnetic Mouse Dust to Make Common Interactor Targets Easier to Select. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology (UIST '07)*. ACM, New York, NY, USA, 183–186. DOI: <http://dx.doi.org/10.1145/1294211.1294242>
- [36] Jasmine Jones and Mark S. Ackerman. 2016. Curating an Infinite Basement: Understanding How People Manage Collections of Sentimental Artifacts. In *Proceedings of the 19th International Conference on Supporting Group Work (GROUP '16)*. ACM, New York, NY, USA, 87–97. DOI: <http://dx.doi.org/10.1145/2957276.2957316>
- [37] William Jones. 2007. Personal information management. *Annual review of information science and technology* 41, 1 (2007), 453–504.
- [38] 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. DOI: <http://dx.doi.org/10.1145/2851581.2856481>

- [39] Daniel Kahneman and Amos Tversky. 2013. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*. World Scientific, 99–127.
- [40] Joseph 'Jofish' Kaye, Janet Vertesi, Shari Avery, Allan Dafoe, Shay David, Lisa Onaga, Ivan Rosero, and Trevor Pinch. 2006. To Have and to Hold: Exploring the Personal Archive. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 275–284. DOI: <http://dx.doi.org/10.1145/1124772.1124814>
- [41] Mohammad Taha Khan, Maria Hyun, Chris Kanich, and Blase Ur. 2018. Forgotten But Not Gone: Identifying the Need for Longitudinal Data Management in Cloud Storage. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 543, 12 pages. DOI: <http://dx.doi.org/10.1145/3173574.3174117>
- [42] Sarah Kim. 2013. *Personal digital archives: preservation of documents, preservation of self*. Ph.D. Dissertation.
- [43] Moon-Hwan Lee, Seijin Cha, and Tek-Jin Nam. 2015. Patina Engraver: Visualizing Activity Logs As Patina in Fashionable Trackers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1173–1182. DOI: <http://dx.doi.org/10.1145/2702123.2702213>
- [44] Moon-Hwan Lee, Oosung Son, and Tek-Jin Nam. 2016. Patina-inspired Personalization: Personalizing Products with Traces of Daily Use. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. ACM, New York, NY, USA, 251–263. DOI: <http://dx.doi.org/10.1145/2901790.2901812>
- [45] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A Stage-based Model of Personal Informatics Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 557–566. DOI: <http://dx.doi.org/10.1145/1753326.1753409>
- [46] Jun Li, Joo Hwee Lim, and Qi Tian. 2003. Automatic summarization for personal digital photos. In *Proc. of ICICS-PCM*, Vol. 3. Citeseer.
- [47] Siân E. Lindley, Catherine C. Marshall, Richard Banks, Abigail Sellen, and Tim Regan. 2013. Rethinking the Web As a Personal Archive. In *Proceedings of the 22Nd International Conference on World Wide Web (WWW '13)*. ACM, New York, NY, USA, 749–760. DOI: <http://dx.doi.org/10.1145/2488388.2488454>
- [48] Siân E. Lindley, Gavin Smyth, Robert Corish, Anastasia Loukianov, Michael Golembewski, Ewa A. Luger, and Abigail Sellen. 2018. Exploring New Metaphors for a Networked World Through the File Biography. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 118, 12 pages. DOI: <http://dx.doi.org/10.1145/3173574.3173692>
- [49] Caroline Lo, Justin Cheng, and Jure Leskovec. 2017. Understanding Online Collection Growth Over Time: A Case Study of Pinterest. In *Proceedings of the 26th International Conference on World Wide Web Companion (WWW '17 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 545–554. DOI: <http://dx.doi.org/10.1145/3041021.3054189>
- [50] Mary Madden and Lee Rainie. 2015. Americans' Attitudes About Privacy, Security and Surveillance | Pew Research Center. (20 May 2015). Retrieved January 16, 2019 from <http://www.pewinternet.org/2015/05/20/americans-attitudes-about-privacy-security-and-surveillance/>
- [51] Clara Mancini, Yvonne Rogers, Arosha K. Bandara, Tony Coe, Lukasz Jedrzejczyk, Adam N. Joinson, Blaine A. Price, Keerthi Thomas, and Bashar Nuseibeh. 2010. Contravision: Exploring Users' Reactions to Futuristic Technology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 153–162. DOI: <http://dx.doi.org/10.1145/1753326.1753350>
- [52] Catherine C Marshall. 2008. Rethinking personal digital archiving, Part 1: Four challenges from the field. *D-Lib Magazine* 14, 3/4 (2008), 2.
- [53] Charlotte Massey, Sean TenBrook, Chaconne Tatum, and Steve Whittaker. 2014. PIM and Personality: What Do Our Personal File Systems Say About Us?. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 3695–3704. DOI: <http://dx.doi.org/10.1145/2556288.2557023>
- [54] Justin Matejka, Tovi Grossman, and George Fitzmaurice. 2013. Patina: Dynamic Heatmaps for Visualizing Application Usage. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 3227–3236. DOI: <http://dx.doi.org/10.1145/2470654.2466442>
- [55] David Merritt, Jasmine Jones, Mark S. Ackerman, and Walter S. Lasecki. 2017. Kurator: Using The Crowd to Help Families With Personal Curation Tasks. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 1835–1849. DOI: <http://dx.doi.org/10.1145/2998181.2998358>
- [56] Richard Mortier, Jianxin Zhao, Jon Crowcroft, Liang Wang, Qi Li, Hamed Haddadi, Yousef Amar, Andy Crabtree, James Colley, Tom Lodge, Tosh Brown, Derek McAuley, and Chris Greenhalgh. 2016. Personal Data Management with the Databox: What's Inside the Box?. In *Proceedings of the 2016 ACM Workshop on Cloud-Assisted Networking (CAN '16)*. ACM, New York, NY, USA, 49–54. DOI: <http://dx.doi.org/10.1145/3010079.3010082>

- [57] Pere Obrador, Rodrigo de Oliveira, and Nuria Oliver. 2010. Supporting Personal Photo Storytelling for Social Albums. In *Proceedings of the 18th ACM International Conference on Multimedia (MM '10)*. ACM, New York, NY, USA, 561–570. DOI: <http://dx.doi.org/10.1145/1873951.1874025>
- [58] William Odom, Richard Banks, David Kirk, Richard Harper, Siân Lindley, and Abigail Sellen. 2012. Technology Heirlooms?: Considerations for Passing Down and Inheriting Digital Materials. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 337–346. DOI: <http://dx.doi.org/10.1145/2207676.2207723>
- [59] William Odom and Tijs Duel. 2018. On the Design of OLO Radio: Investigating Metadata As a Design Material. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 104, 9 pages. DOI: <http://dx.doi.org/10.1145/3173574.3173678>
- [60] William Odom, Abi Sellen, Richard Harper, and Eno Thereska. 2012. Lost in Translation: Understanding the Possession of Digital Things in the Cloud. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 781–790. DOI: <http://dx.doi.org/10.1145/2207676.2207789>
- [61] William Odom, John Zimmerman, Scott Davidoff, Jodi Forlizzi, Anind K. Dey, and Min Kyung Lee. 2012. A Fieldwork of the Future with User Enactments. In *Proceedings of the Designing Interactive Systems Conference (DIS '12)*. ACM, New York, NY, USA, 338–347. DOI: <http://dx.doi.org/10.1145/2317956.2318008>
- [62] William Odom, John Zimmerman, and Jodi Forlizzi. 2011. Teenagers and Their Virtual Possessions: Design Opportunities and Issues. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 1491–1500. DOI: <http://dx.doi.org/10.1145/1978942.1979161>
- [63] William Odom, John Zimmerman, and Jodi Forlizzi. 2014. Placelessness, Spacelessness, and Formlessness: Experiential Qualities of Virtual Possessions. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14)*. ACM, New York, NY, USA, 985–994. DOI: <http://dx.doi.org/10.1145/2598510.2598577>
- [64] William T. Odom, Abigail J. Sellen, Richard Banks, David S. Kirk, Tim Regan, Mark Selby, Jodi L. Forlizzi, and John Zimmerman. 2014. Designing for Slowness, Anticipation and Re-visitation: A Long Term Field Study of the Photobox. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1961–1970. DOI: <http://dx.doi.org/10.1145/2556288.2557178>
- [65] Blaine A Price, Karim Adam, and Bashar Nuseibeh. 2005. Keeping ubiquitous computing to yourself: A practical model for user control of privacy. *International Journal of Human-Computer Studies* 63, 1-2 (2005), 228–253.
- [66] Kopo Marvin Ramokapane, Awais Rashid, and Jose Such. 2017. “I feel stupid I can’t delete...”: a study of users’ cloud deletion practices and coping strategies. In *Proceedings of the Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. USENIX Association.
- [67] Corina Sas and Steve Whittaker. 2013. Design for Forgetting: Disposing of Digital Possessions After a Breakup. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1823–1832. DOI: <http://dx.doi.org/10.1145/2470654.2466241>
- [68] Corina Sas, Steve Whittaker, and John Zimmerman. 2016. Design for Rituals of Letting Go: An Embodiment Perspective on Disposal Practices Informed by Grief Therapy. *ACM Trans. Comput.-Hum. Interact.* 23, 4, Article 21 (Aug. 2016), 37 pages. DOI: <http://dx.doi.org/10.1145/2926714>
- [69] Darshana Sedera and Sachithra Lokuge. 2018. Is Digital Hoarding a Mental Disorder? Development of a Construct for Digital Hoarding for Future IS Research. In *Proceedings of the International Conference on Information Systems, ICIS 2018, San Francisco*.
- [70] Margo I Seltzer and Nicholas Murphy. 2009. Hierarchical File Systems Are Dead. In *HotOS*.
- [71] George Sweeten, Elizabeth Sillence, and Nick Neave. 2018. Digital hoarding behaviours: Underlying motivations and potential negative consequences. *Computers in Human Behavior* 85 (2018), 54–60.
- [72] 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. DOI: <http://dx.doi.org/10.1145/2818048.2820017>
- [73] 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. DOI: <http://dx.doi.org/10.1145/3173574.3174161>
- [74] Stephen Volda and Saul Greenberg. 2009. WikiFolders: Augmenting the Display of Folders to Better Convey the Meaning of Files. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1679–1682. DOI: <http://dx.doi.org/10.1145/1518701.1518959>

- [75] Stephen Volda and Elizabeth D. Mynatt. 2009. It Feels Better Than Filing: Everyday Work Experiences in an Activity-based Computing System. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 259–268. DOI: <http://dx.doi.org/10.1145/1518701.1518744>
- [76] Rebecca D. Watkins, Abigail Sellen, and Siân E. Lindley. 2015. Digital Collections and Digital Collecting Practices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3423–3432. DOI: <http://dx.doi.org/10.1145/2702123.2702380>
- [77] Markus Weiland and Raimund Dachzelt. 2008. Facet Folders: Flexible Filter Hierarchies with Faceted Metadata. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems (CHI EA '08)*. ACM, New York, NY, USA, 3735–3740. DOI: <http://dx.doi.org/10.1145/1358628.1358922>
- [78] Roger Whitham and Leon Cruickshank. 2017. The Function and Future of the Folder. *Interacting with Computers* 29, 5 (2017), 629–647. DOI: <http://dx.doi.org/10.1093/iwc/iww042>
- [79] Steve Whittaker. 2011. Personal information management: from information consumption to curation. *Annual review of information science and technology* 45, 1 (2011), 1–62.
- [80] Steve Whittaker, Ofer Bergman, and Paul Clough. 2010. Easy on that trigger dad: a study of long term family photo retrieval. *Personal and Ubiquitous Computing* 14, 1 (2010), 31–43.
- [81] Steve Whittaker, Vaiva Kalnikaitė, Daniela Petrelli, Abigail Sellen, Nicolas Villar, Ofer Bergman, Paul Clough, and Jens Brockmeier. 2012. Socio-Technical Lifelogging: Deriving Design Principles for a Future Proof Digital Past. *Human-Computer Interaction* 27, 1-2 (2012), 37–62. DOI: <http://dx.doi.org/10.1080/07370024.2012.656071>
- [82] Steve Whittaker and Candace Sidner. 1996. Email Overload: Exploring Personal Information Management of Email. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '96)*. ACM, New York, NY, USA, 276–283. DOI: <http://dx.doi.org/10.1145/238386.238530>
- [83] Richmond Y. Wong, Deirdre K. Mulligan, Ellen Van Wyk, James Pierce, and John Chuang. 2017. Eliciting Values Reflections by Engaging Privacy Futures Using Design Workbooks. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 111 (Dec. 2017), 26 pages. DOI: <http://dx.doi.org/10.1145/3134746>
- [84] Xuan Zhao and Siân E. Lindley. 2014. Curation Through Use: Understanding the Personal Value of Social Media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2431–2440. DOI: <http://dx.doi.org/10.1145/2556288.2557291>
- [85] John Zimmerman. 2005. Video Sketches: Exploring pervasive computing interaction designs. *IEEE pervasive computing* 4, 4 (2005), 91–94.
- [86] 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. DOI: <http://dx.doi.org/10.1145/1240624.1240704>