Data Dashboard: Exploring Centralization and Customization in Personal Data Curation

Francesco Vitale[1], **Janet Chen**[1], **William Odom**[2], **Joanna McGrenere**[1]
[1]The University of British Columbia, Vancouver, British Columbia, Canada
[2]Simon Fraser University, Surrey, British Columbia, Canada
fvitale@cs.ubc.ca, janet.chen@alumni.ubc.ca, wodom@sfu.ca, joanna@cs.ubc.ca

ABSTRACT
The advent of cloud platforms and mobile devices has complicated personal data management and decisions about what data to keep or discard. To explore how to help people *curate* their data, we designed Data Dashboard, a prototype system that provides: 1) a centralized overview of data from across platforms and devices, 2) customizable filters for sorting through many types of data. We evaluated the prototype with 18 participants. Building on top of previous literature, we use the concept of *data boundaries* (the idea of invisible but important separations across data) to explain participants' reactions to the prototype. We show that centralizing data in a single management system *blurs* boundaries and requires safety guarantees. Customization options, instead, *uphold* boundaries and reinforce user control. We discuss how to use these results for integrating data boundaries in the design of new tools, rethinking the language of personal data, and envisioning a post-cloud future.

Author Keywords
Personal data management; personalization.

CCS Concepts
•**Human-centered computing** → **Empirical studies in HCI;** *Interface design prototyping;*

INTRODUCTION
The past decade has changed the way we think about personal data. Ten years ago cloud storage platforms were in their infancy. Dropbox launched in 2008, iCloud in 2011, Google Drive in 2012. Before the advent of these platforms and before the popularity of mobile devices, people largely thought of their data as limited to files and folders stored on computers. Now, personal data is a buzzword at the center of political debates and regulatory efforts. For everyday users, managing data across platforms and devices is increasingly challenging [93, 94]. The growing amount of personal data that people create and store, with varied types of data, sometimes created automatically by technology, has shifted the HCI community's attention towards questions about data ownership, privacy, and curation [56]. In this paper, we look at a broad range of data types (from documents and photos to apps and automatic logs of usage, like browser histories) and investigate how to support people in *curating* this disparate set of items. We define personal data (or information) curation as the process of storing, organizing, and re-accessing digital data over time [102], focusing on storing and organizing decisions. Recent work shows that people need better tools to curate data, especially on cloud platforms [60, 94]. Two key challenges complicate people's practices and possible design efforts.

The first challenge for personal data curation is the growing number of devices and cloud platforms that people use [93]. The distributed, often fragmented nature of personal data undermines awareness of ownership, making curation difficult: it is hard to curate data if you do not know what you own [72]. One approach to address this issue is ***centralization***, that we define as providing an aggregated overview of data from different places in a single central tool [72, 74]. Our work asks the following questions: Can centralization help people decide what personal data to keep or discard? If so, how should we approach the design of centralized tools? With cloud platforms like Google Drive and Dropbox encouraging users to centralize and sync all data to the cloud, investigating this question can help us understand the consequences of this approach.

The second challenge for building curation tools is the subjective nature of personal data management and curation [11, 75, 94]. It is difficult to create a solution that can satisfy different types of users who have different management styles and curation approaches [60, 95]. A common design approach to deal with individual differences is to turn to *personalization* or ***customization***–the idea of tailoring a system to specific users' needs and characteristics. From this comes our second set of research questions: Given the personal and subjective nature of data curation, can customization help? How desirable is a personalized approach to data curation? And, what are the aspects that make it more or less desirable?

To answer these questions, we designed Data Dashboard, a centralized and customizable system for curating personal data. We evaluated an interactive prototype of Data Dashboard with 18 participants who had different approaches to data curation, asking them to go through five potential use scenarios.

DIS '20, July 6–10, 2020, Eindhoven, Netherlands.

ACM ISBN 978-1-4503-6974-9/20/07 ...$15.00.
http://dx.doi.org/10.1145/3357236.3395457

Drawing on previous work around digital data, we use the concept of ***data boundaries*** (the idea of conceptual lines that prescribe where to store personal data and how) to understand participants' reactions to centralization and customization. We show that centralization *blurs* boundaries and introduces a dilemma around privacy and security, requiring explicit safety guarantees. Customization, on the other hand, is easier to accept because it *upholds* boundaries. We discuss what these results mean for future data curation tools.

Our work makes three contributions: 1) we provide additional empirical evidence for the role of data boundaries in personal data curation, and use it to understand participants' reactions to design choices–specifically, we show reactions to centralization and customization; 2) we offer an approach to address key challenges in data curation by designing a unified tool with personalized functions; 3) we outline design and research directions for future data curation tools focused on integrating data boundaries into design, rethinking the language of personal data, and envisioning a post-cloud future.

BACKGROUND AND RELATED WORK

Related work falls into three main areas: (i) personal data curation and management; (ii) data boundaries; (iii) customization. Additional related work appears in the prototype description.

Personal data curation and management

Referring to work by Whittaker [102], we define *personal data curation* as the process of keeping, organizing, and reaccessing personal data over time. In our work, we focus on keeping and organizing decisions: these stages of curation are less supported by current tools compared to data retrieval. Like past studies [93–95], we define *personal data* as the data that people store, create, or interact with on their devices and cloud platforms, taking a holistic approach and letting participants consider a wide array of data types. This definition goes beyond traditional notions of personal information as limited to documents or files and folders [27]. Data Dashboard includes several examples of data types that show the evolving and disparate nature of personal data (e.g., apps, logs), with some types (e.g., browser history, tracking data) often unwillingly stored automatically. To limit our scope, we do not include specific life-logging applications, but our ideas can apply to this domain, covered in detail by past work [36, 38, 53].

Studies on Personal Information Management (PIM) and curation show that people have different tendencies when deciding what data to keep or discard, with some keeping most of their data [55, 89, 94] and others trying to reduce the number of items they store [55, 94]. Similarly, there are individual differences in organization, with some people being more organized than others [46, 51, 52]. In general, people find challenging deciding what to keep or discard and organizing their data [15, 64, 65, 89, 94, 102]. Several PIM studies propose and evaluate tools for curating, managing and retrieving documents [13, 25, 32, 34, 39, 57], photos [14, 70, 83, 106], contacts [12], or emails [9]. A related strand of work looks at how automating can improve task management [47] or cloud file management [16]. However, attitudes towards automating data curation show individual variation, with some people opposing the idea [95]. Our work further explores attitudes towards technology support for data curation.

The majority of past design work in PIM takes a quantitative approach for evaluating prototypes, using lab experiments or usage log analysis in field deployments. These studies largely look at time on task and similar metrics for measuring success. A few studies, instead, use a more qualitative approach, focused on teasing out participants' attitudes. In our work, we use a qualitative approach in line with research through design [104] and reframe the Data Dashboard prototype as a tool to understand broader aspects of data curation, instead of focusing only on its usability aspects. This approach follows past work on digital data that uses systems to prompt discussion with participants [43, 44, 71, 79, 95].

Data boundaries

Past work points to the idea of *data boundaries* as a key lens to understand data curation. At a high level, we define data boundaries as conceptual lines that demarcate where to store personal data and how. Boundaries appear in studies on data management [75, 93–95, 101], collaboration [96, 98], communication [19, 20, 69], privacy [3, 6, 77, 78], and personal possessions (both physical [8, 22, 31, 88] and digital [61, 62, 73, 100]). We can argue that data management and curation are essentially about establishing and negotiating boundaries. This process has roots in cognitive models of how the human mind works: to make sense of a continuous world, we build categories and "draw mental boundaries around them" [75].

Research on digital data often looks at the contrast between physical and digital possessions, exploring the boundary between the two domains [42, 43, 55, 59, 73, 74, 76]. The notion of boundaries runs throughout studies on physical possessions, helping us understand how people experience them. For example, by using boundaries to mark the unique character of their spaces [31], give meaning to clutter [88], and establish what belongs in their home [22]. These are all practices that help in building an identity [8].

Boundaries also exist in the digital world. Sometimes they are explicit, as is the case when people use folders to create structure in their data [101], or when they build collections with criteria for what goes in and what does not [100]. More often data boundaries are implicit. They are enforced by tools and applications [9, 62, 93, 96], work and personal life [19, 20], group and family relationships [61, 69, 93, 97], activities [98], and context [95]. A key boundary to understand participants' reactions to centralization is the one around privacy. Previous work argues that privacy management is at its core about negotiating boundaries between the public and private sphere [3, 77, 78]. Privacy boundaries are "permeable," "murky," contextual [3], subjective and dynamic, evolving over time [6, 77, 78].

Our work provides additional evidence for how boundaries drive personal data curation and shows how to use the concept to understand reactions to centralization and customization. We did not have in mind the concept of boundaries when we designed Data Dashboard or the study, but we identified it in the analysis and then traced it back to past work.

Personalization and customization

Personalization involves adapting systems to users' individual preferences and characteristics [5, 17, 47, 66, 67, 87]. Past work highlights different levels of user involvement [17, 66, 67, 87]. In purely "adaptive" system-controlled personalization, the system does not directly involve users in its adaptions [87]. System-controlled personalization is largely implicit. Customization, by contrast, is the term used for user-controlled personalization: here, the user is responsible for changing the system through explicit actions. A middle ground approach involves mixed-initiative systems, where personalization is system-initiated but "approved" by the user [18, 49, 54]. Regardless of level of user involvement, personalized changes in the system can take place in the user interface (layout, color, fonts–these changes focus on aesthetics [103]), content (e.g., showing or hiding specific content to different users, as it happens on social networks [66,87]), or functionality (e.g., macros in spreadsheets, browser extensions, changing system settings–these changes focus on how the system works [47, 103]). Changes in functionality are often labeled "advanced customization or personalization" [47]. In this study, we focus on content and functionality changes.

Past work shows that personalization can help users complete tasks [18, 40, 41], but several factors can prevent users from customizing software: the time and knowledge required by customization [63]; levels of exposure, awareness, or social influence [5, 40]; individual differences [67]. In general, users seem to prefer mixed-initiative interfaces over more automatic adaptions [18, 67]. In the field of PIM, several studies look at automating information organization, with past work providing an overview of past efforts [49, 95]. However, only a few studies look explicitly at personalization in data management [47] and curation [70, 83, 95], suggesting this is a promising but underexplored approach, Our work investigates how to design customization mechanisms for data curation.

THE DATA DASHBOARD PROTOTYPE

We designed Data Dashboard to address two key challenges of personal data curation outlined in the introduction: the fragmentation of data and the subjective nature of curation.

Overview of the prototype

Data Dashboard is a centralized system that shows an aggregated overview of data stored on different devices and cloud platforms (e.g., Dropbox, Google Drive, iCloud) and a set of customizable filters for curating data. There are four sections: Activity, Explore Your Data, Quick Actions, Settings.[1]

Activity (Figure 1) provides an overview of recent data (Google Drive, Dropbox, and macOS provide similar views of recent files). For example, for a given day, it shows documents that users have created or edited, photos and screenshots they have taken, apps they have installed, and so on (these are just some examples from many possible types of data). Users can filter their activity by time (today, this week, this month, all time, custom period). Activity also has a section aggregating data shared with other people on collaborative services (e.g., Dropbox, Slack), grouped by the person it was shared with.

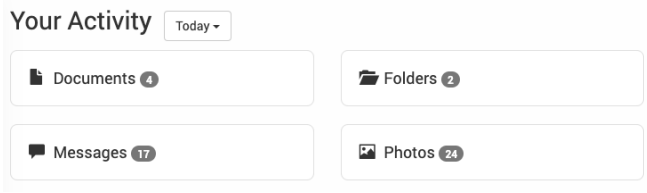

Figure 1. Activity shows an overview of recent data.

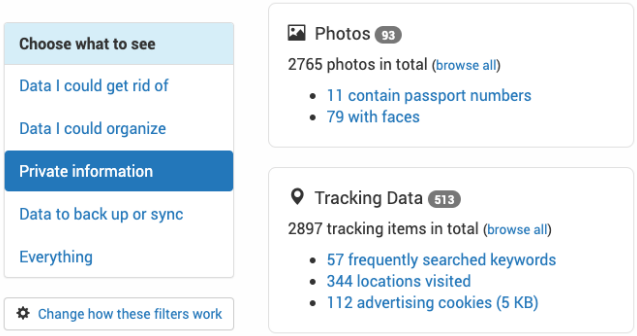

Figure 2. Explore Your Data shows an overview of all data from different devices and cloud platforms, that users can use filter.

Explore Your Data (Figure 2) shows an overview of all data users have on different devices and platforms, grouped by type (e.g., photos, messages, documents, cache & logs)[2]. Users can sort or hide the different types. They can also filter what the system shows, choosing on the left: "Data I could get rid of," "Data I could organize," "Private information," "Data to back up or sync." Users can customize these filters in Settings.

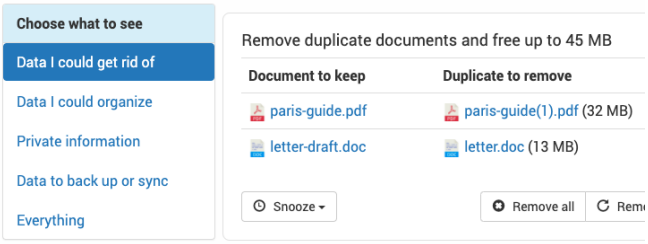

Figure 3. Quick Actions has recommendations for curating.

[1]The prototype is online at https://datadashboard.github.io. It works on all desktop browsers, but has some bugs in Safari. It is not optimized for mobile devices. The prototype is only an illustration of possible functions: it does not connect to any device or cloud platform. The supplementary materials to the paper include a video walkthrough of the prototype and additional figures of the sections.

[2]The complete list of data types we considered is visible in Explore Your Data after clicking the "Sort or hide the types of data you see" button. It includes: photos, screenshots, emails, text documents, presentations, spreadsheets, videos, audio, folders, contacts, applications, messages, cache and logs, bookmarks, ebooks, tracking data, browser history, passwords, notes and reminders, games. This is a comprehensive but not exhaustive list of types, meant to be a starting point in our exploration. Future work could extend our approach to a wider range of types (e.g., unintentionally digital traces collected by technology, like time spent on a device, clicks, queries.)

Quick Actions (Figure 3) provides a list of recommended actions for different data items. For example, the system suggests duplicates to remove, documents to rename, or message conversations to archive. Like in Explore Your Data, users can filter the recommendations using the sidebar filters (same options as in Explore Your Data). Users can also apply automatically suggested actions to similar items in the future.

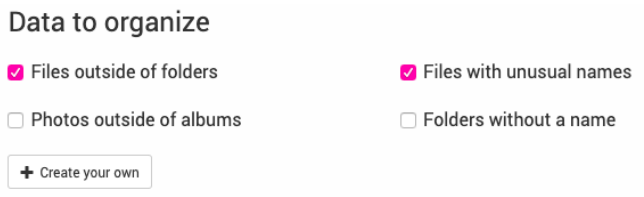

Figure 4. The Settings page allows user to customize data filters.

In **Settings** (Figure 4), users can customize how the filters for Explore Your Data and Quick Actions work. They can choose to include or exclude some default combinations of data types and criteria or create new, custom combinations. They can also add or remove connected devices and cloud accounts.

Rationale and design

Data Dashboard takes inspiration from previous empirical and design work on personal data [62, 72, 94, 95], while also combining and extending key aspects of existing commercial and research products. The visual design of the system takes direct inspiration from Google Dashboard and similar privacy dashboards. We wanted Data Dashboard to feel similar to existing systems so that participants could better imagine when to use it. Still, we had to build our own system to explore how centralization and customization could work together.

The idea of a centralized tool

One key inspiration for Data Dashboard is work by Odom et al. [72,74] on cloud platforms and the need for more awareness of digital possessions. Odom et al. suggest creating a "visual inventory" of digital possessions, "a place where 'my stuff' can be found, even if, in technical terms, it exists on many different servers, or many applications." A place to quickly go back where data is originally stored, preserving its context. Lindley et al. [62] explored a similar premise and found that a centralized web archive might not be the ideal solution for users. Data Dashboard, however, is not a central archive in itself, as it is largely an overview that provides links to data stored on different devices and platforms. Instead, we see it as a tool to curate a "meaningful archive" [65] within the systems that people already use to store and manage data. Some past PIM projects also explore the idea of centralizing data [25, 34, 57]. But they focus on retrieving data, rather than on helping users curate and decide what to keep or discard. Cloud platforms like Google Drive and Dropbox, instead, are moving towards centralization by encouraging users to "sync" all the data from their computers to the cloud as a backup [33, 84]. There are also commercial products that offer to unify separate cloud accounts, like odrive and MultCloud, but Data Dashboard has specific novel functions for curation (e.g., the data filters and their customization).

Providing a dashboard

A key function of Data Dashboard is to provide an overview of personal data by showing numbers for different categories (e.g., 49 blurry photos). Two systems inspired this approach: 1) Cardinal [30], a research tool that scans a user's computer and provides counts for the total number of files, breaking down numbers for some popular file types (e.g., photos). We wanted to use the same approach in a user-oriented system. 2) Google Dashboard, a privacy-oriented page that provides an overview of data created and stored in Google products [80]. (Yahoo also provides a similar interface for managing privacy settings, and work on GDPR compliance also proposes the idea of a dashboard [81].) Data Dashboard provides a similar function but extends it to entire personal data "ecosystems" [93], bringing together data from more than a single device or platform. Similarly, past work on data curation [95] and "cleaning" tools like Files, Clean My Mac, and CCleaner inspired the automatic recommendations in Quick Actions. However, we wanted Data Dashboard to be more comprehensive than similar tools. Most of these tools focus on freeing up space by finding items to discard based on their size. Instead, we wanted to help users go beyond freeing up space, and provide them with a broader set of actions.

Personalization mechanisms

Data Dashboard provides both content customization and functional customization. In terms of content, users can customize what data types they see and their order in Explore Your Data. In terms of functions, users can personalize how the sidebar filters in Explore Your Data and Quick Actions work, deciding what the system will show. This approach is in line with work on advanced personalization for task management [47]. The four options in the sidebar filters ("Data I could get rid of," "Data I could organize," "Private information," "Data to back up or sync") reflect different behavioral patterns and goals that people might refer to when curating data, based on past related work [89,94,95]. In addition, the three main sections (Activity, Explore Your Data, Quick Actions) target different types of users and data curation practices. Activity and Explore Your Data target users who prefer to explore data and decide on their own what to do by inspecting items individually. By contrast, Quick Actions targets users who welcome automation and expect an intelligent system to do things for them. Previous work details these contrasting attitudes to data curation [95].

Implementation

We built the prototype using AngularJS and Bootstrap 3. We designed Data Dashboard as a system that would be quick to turn into a horizontal prototype, that is, a prototype where only top-level functions are implemented to communicate the scope of the whole system [10]. All data in the prototype is "fake." This is a limitation of our approach, as previous work shows the value of using real participants' data [44]. However, we took this approach because the prototype would have required a much longer development time if we had to rely on devices' and platforms' APIs, and at this stage we wanted to explore its utility before committing to high development costs. Not using real data also protected participants' privacy and gave all participants the same experience when going through the

scenarios. The recommendations in Quick Actions, the data types in Explore Your Data, and number of items for each type are meant to illustrate the scope of the prototype and largely reflect common distributions of personal collections [28, 29].

EVALUATION METHODOLOGY

We evaluated Data Dashboard in a user study with 18 participants. In the evaluation, we collected participants' opinions about the Data Dashboard interface, the key ideas behind it, and possible scenarios of use. This specific approach to *Research through Design* [104] aims to use design artifacts and devices to frame and open prospective conversations with participants, like previous studies show [43, 44, 71, 79, 95]. This approach also explains the relatively minimal, under-designed nature of the prototype: we did not design a costly, high-fidelity prototype because our goal was a design exploration focused on eliciting participants' reactions. Before recruiting participants we ran two pilot sessions with members of our lab. We also gathered feedback from other lab members throughout the iterative development of the prototype, going from paper sketches to the interactive version we implemented.

Participants

We recruited participants from a university recruiting list and Craigslist in Vancouver, Canada. We used a screening survey (available in the supplementary materials) where we asked participants' age, occupation, main approach to data curation, and what devices, cloud platforms, and data management tools they used (all from a list of popular options). We received 169 responses to the survey. We contacted 38 respondents, 25 agreed to participate, and we ran the study with 20 (5 canceled or did not show up). After the first two sessions we made some changes in the prototype and the protocol, so we excluded the first two participants from the analysis. Thus, the study had 18 participants (12 women, 6 men, aged 18-64, median age: 33). Participants' occupations included accountant, background actor, business contractor, childcare provider, facilitator, occupational therapist, sales associate, social worker, student, postdoc, research manager, retired. Most participants self-reported average technical skills and no experience in computer science or programming. We recruited for a varied set of participants who used different tools, cloud platforms, and devices. Four participants did not use any curation tools. Of these four, one also did not use any cloud platforms. Participants also varied in their approach to curation, based on general organization practices, how much data they tended to keep or discard, and how they felt about their approach.

Procedure and data collection

The study sessions took place over one month, with each lasting 41-97 minutes (average: 64). The session had three parts (each lasting 10-30 minutes): an introductory interview, an exploration and scenario-based interaction with the prototype, a debriefing interview with two, short card-sorting activities. Participants interacted with the prototype using Chrome on a MacBook Air 13" laptop. We audio recorded the whole session and screen-recorded the prototype interaction. Whenever possible, two members of the research team conducted the session. One member asked questions, the other took verbatim notes. When only one member of the team could run the interview, we later transcribed the recording. After each interview, we compiled debrief notes with preliminary insights. Participants received $20 in compensation.

Introductory interview

In the first part of the session, we had an introductory interview focused on data curation practices. First we asked participants to remember and tell us about the last time they decluttered some of their data, by reviewing, organizing, discarding, archiving, or moving several items at once. Then we asked participants to show us examples of how they organized their data on their devices, how they decided what to keep or discard, and how they used specific tools –if any– to manage or curate their data (e.g., settings panels to clean up data, Google Dashboard, and so on, depending on what they mentioned using in the screening survey and in their interview answers).

Use scenarios

In the second part of the session, we introduced participants to the prototype and let them familiarize with it for a few minutes, prompting them to think aloud. After this initial exploration, we asked participants to go through five possible use scenarios with the prototype, once again asking them to think out loud.

Scenario (SC) 1 (space running out): "The space on your computer is running out. You want to find some data to discard. You are not sure where to start looking, but you know that you do not care too much about old documents."

SC2 (regular curation): "It is a rainy day. You have set aside some time for doing a regular cleanup of your devices. You usually do this every few months. You want to review your data and make sure everything is organized in your preferred way."

SC3 (exploring recent data): "You have 5 to 10 minutes in between meetings and errands. You decide to take a look at your recent data to get a sense of anything that needs taking care of."

SC4 (protecting data privacy): "You have heard about a data leak from a popular cloud storage platform that exposed personal information to hackers. You want to review what data you have stored on different cloud platforms that might pose a privacy risk in the future."

SC5 (safeguarding data across devices and platforms): "You are in the process of buying a new computer. You want to make sure that you are not going to lose any of the data you care about. You want to ensure that everything is stored in more than one place."

The scenarios are based on previous empirical work on user practices and individual differences in personal data curation [93–95]. We wanted to explore how different functions in the prototype can support different scenarios and user attitudes. The scenarios were also key for helping participants focus on concrete implications of use rather than low-level details of the prototype (like colors, fonts, buttons). Some of the scenarios mention specific devices to feel concrete, but we encouraged participants to see them as a starting point to discuss additional devices and broader situations or practices.

Debriefing interview

The third and final part of the session was a debriefing interview about the prototype. Here, we asked participants about their impressions of the system, clarifications about what they did during the scenarios, and then walked through all the four sections of the prototype one by one to gather more specific feedback. Finally, we had a short card-sorting activity where we asked participants to rank the five scenarios by how relevant they were to their own experience. We prompted participants to explain their ranking, elaborate on the match between different data management methods or tools (prototype included) and scenarios, and consider in what other situations they could imagine using the prototype. Participants used small paper printouts of the scenarios. We also asked participants to rank the usefulness of the prototype against any tools they mentioned using in the screening survey or during the interview. Once again, participants used paper printouts with the names of the different tools we had prepared for them.

Data analysis

We used thematic analysis to identify recurring themes and patterns from the sessions [23]. The analysis process took place over three weeks. Two members of the team conducted the bulk of the analysis and later discussed the themes with the other team members. We started with a round of open-coding, where the two team members coded data in parallel, seeing each other's codes. Then, we grouped codes into categories, and started thinking about themes and patterns across categories. Codes and categories were both inductive and deductive (based on insights from previous studies, and specific aspects or sections of the prototype). We discussed several iterations of possible themes, choosing specific areas of the analysis to focus on. After identifying the key lens of the analysis (centred around data boundaries and their effect on centralization and customization), we went back to the transcripts and re-coded them using only the three themes to check for consistency and make sure that our interpretation fully captured participants' experience.

RESULTS

In this section, we first provide some context for the analysis, based on participants' general reactions and usage of the prototype. Then, we delve into the key themes of the analysis.

Overall reactions and interactions in the scenarios

Most participants (12/18) had positive reactions to Data Dashboard, saying that it was "smart," "intuitive", "user-friendly," and would save their time. Several participants also preferred the system when comparing it to other tools they had used in the past. Some participants went so far as asking if they could have the system installed on their devices after the interview. But not all participants liked Data Dashboard. Participants who had mixed reactions (3) thought that some aspects of the system were unclear or unnecessary. Participants who had negative reactions (3), instead, said they did not need or want a tool like Data Dashboard. Some opposed the idea of a system deciding how to curate data. Others did not see data management as something worth their time. These negative reactions are in line with previous work and support the idea that different users have different needs when it comes to personal data curation and technology support [95]. Many participants also reflected on the potential privacy risks of centralizing personal data; we expand on this theme later.

Explore Your Data (EYD) and Quick Actions (QA) were the most commonly used sections of the prototype during the scenarios. Participants thought that EYD worked best for occasional, more focused scenarios and they used it more frequently in SC1 (getting rid of data to free up space) and SC2 (regular scheduled cleanup). Instead, QA would work better for short management episodes (SC3, taking a few minutes to look at recent data). Most participants also used Activity at some point, but several participants found it underwhelming and too similar to EYD. Several participants did not notice the Shared Data section or found it confusing. Some thought it would help them when working on collaborative projects. All participants except one discovered the sidebar filters in EYD and QA on their own, and most used them at some point during the scenarios. Most participants found the filters comprehensive and the idea of automatically clustering data helpful. Participants did not have clear requirements for how the system should generate suggestions or cluster data, but they expected "*a machine learning algorithm that gets better with time.*" (P18) Most participants also discovered the Settings page and its link to the sidebar filters but several found it initially confusing. In general, participants thought that it was a good idea to customize the filters because it gave them more control; we expand on this theme later.

Scenario 4 (protecting privacy) and 5 (safeguarding data across devices) presented some challenges for participants. When going through the privacy scenario, several participants scanned the system looking for a way to see all data stored in a specific platform (e.g., iCloud, Dropbox, Google Drive) or device (e.g., "my drive"). Often they could not find what they were looking for. But many participants also said that they prefer not to store private data on the cloud in the first place and that this scenario did not apply to them. In the last scenario, instead, many participants did not use the system and talked about their actual process of backing up files either manually or through completely automatic solutions (e.g., Apple's Time Machine, Google Photos). Some participants also had trouble with the language used by the system, saying they had no idea what "syncing" data meant: "*I know what data to back up means but I don't know what sync means. I don't know if they're related, maybe.*" (P16) Overall, these reactions highlight how participants saw privacy protection and backing up as processes that either take place outside of specific tools or require little input from them. Participants most commonly ranked these last two scenarios as the least relevant.

These results show how our scenario-based evaluation prompted participants to explore Data Dashboard and think about its design. They also suggest that our approach of combining centralization and customization has potential. To unpack its value and potential risks we now turn to the more interpretive part of the analysis, where we delve into key aspects of curation, centralization, and customization.

Data boundaries drive curation

When we analyzed participants' reactions to the prototype, one idea became key: data boundaries drive curation. We have briefly discussed what we mean by data boundaries in the related work section. Boundaries are an abstract concept that can explain how people enact curation of their personal data ecosystems [93]. People create implicit or explicit boundaries that separate different categories of information. These boundaries help people build their identity, mark areas of their life, and feel in control. For some people boundaries will be more malleable, for others more strict [75]. Here, we provide additional evidence for how people create, protect, and think of data boundaries. Then, we use the concept of boundaries to explain reactions to centralization and customization.

Creating boundaries

All participants implicitly talked about creating boundaries when deciding what data to store and where to store it. They mentioned rules for what goes where and why: "*I have my data compartmentalized:* [the] *tablet is just for reading.* [The] *laptop is for everything school related.*" (P9). A key boundary was between private and non-private information, with participants choosing where to store private information based on perceived privacy risks of devices and cloud platforms: "*I don't put any of my private stuff on the computer at all.*" (P6) Boundaries also helped participants distinguish and prioritize data based on importance. For example, P10 discussed moving data from one cloud platform to another to separate expendable and essential items: "*I started moving everything* [from Box] *to Dropbox, because I think it's just more reliable. So everything that I can't afford to lose, I'll stick it in Dropbox.*"

Breaking boundaries

The centralized approach of Data Dashboard prompted participants to reflect on how they often experienced breaking points in boundaries. Sometimes this was intentional. Reflecting on the suggestions and filters for "private information" in the prototype, one participant explained that sometimes it is necessary to break a boundary around privacy because of convenience: "[Having passport information on Google Drive is] *not ideal. But because of frequent travel, I am somewhere and I am filling in a form and I need that information,* [so] *either through Google Drive or email I was able to locate it. So, it's more about ease of access to the information that saves my time than anything else. Ideally, I don't like to have important documents even on email, but I haven't learned if there is a secure system to store them online.*" (P17)

In other cases, boundaries were broken unintentionally. For example, after looking at the different types of data in the prototype, one participant explained the frustration of having WhatsApp photos from other people go automatically into their own. In this case, the boundary between "my stuff" and "other people's stuff" was broken without permission: "*I tried to figure out on this device as well, some sort of filter where you can manage whether your images from WhatsApp, a group message thread, are automatically going into your photos or not. It drives me crazy that WhatsApp's photos go automatically into your photos. If there's more of a filtering system to help you organize that, that would be great.*" (P12)

Boundaries influence trust

The need to create and protect boundaries also influenced participants' trust in the systems and platforms to manage data. Several participants asked whether Data Dashboard would be associated with a specific brand because they tended to trust specific brands with their data: "*Google has created credibility over years so I know I have trust in that system.* [I use] *Mac because of* [its] *ease of use but I don't always store all my important info over there. They have a very tricky way. If you're locked out, only they can unlock it. I don't like that dependency on a third party.*" (P15) Others, instead, referred to the boundary between physical hardware and cloud platforms to explain their practices, wondering whether Data Dashboard would be largely local or cloud-based: "*You can't trust anything that you don't have control on the hardware. Especially with smartphones, because most of it goes through the cloud, you can't really trust anything.*" (P4)

Boundaries lead to fragmentation

A consequence of creating boundaries is data fragmentation [57,93]. This is one of the key challenges for data curation in an age of multiple devices and cloud platforms. When considering data boundaries, participants explained how fragmentation could be beneficial: "*I am isolating my data more and more rather than sharing it. The people who invented Facebook and all these platforms, they did it with good intentions. The problem is these platforms are now abused* [...] *You can't really trust their intentions so you have to protect yourself by isolating your data.*" (P4) But it could also be costly, leading to confusion and frustration: "*I've started working for different organizations, you have data from different places and sometimes it gets confusing. I have some things from my previous experiences and now everything is getting mixed up.*" (P3) From a design standpoint, we can ask whether fragmentation is a problem to be solved and how to solve it if that is the case. Next, we look at how centralization and customization intersect with data boundaries.

Centralization blurs data boundaries

Centralization is convenient

Participants saw something positive in centralizing data, with the different sections of Data Dashboard providing avenues to reconcile data from different devices and platforms: "*I really like how it's combining different sources of where these things could be stored.*" (P12) They thought that centralizing data provided some clear benefits, like saving time: "*This looks delightful. The most important criteria for a service like this for me is the time benefit I get from it* [...] *I don't mind giving them access to all my data.*" (P3) Or, making it easy to see everything in one place, as was the case in Explore Your Data: "*You see everything: your contacts, your bookmarks. That's really handy. I like that. It just feels very comprehensive.*" (P10) "*It's all different types of* [accounts]. *I like the way it's set up. It's quite clear, it looks good. I can see everything at a glance.*" (P16) They also hoped that the cards for different types of data they saw in Data Dashboard could break the barriers between different devices: "*If it were to show me text messages, it would be good for me because right now I have no way to see them from the computer.*" (P11)

Convenience goes hand in hand with risks

However, data centralization introduced a dilemma. As much as it seemed convenient, it created additional risks: "*If I have everything together,* [the] *ease of use is high but* [the potential risk for a] *security breach is higher as well.*" (P15) Especially when looking at Activity and Explore Your Data, most participants expressed concerns that touched on how centralization blurs data boundaries: "*Some people like to collaborate in one place, one app, access* [all] *email accounts and whatever, but that's not my preferred way.*" (P8) Based on these concerns, they wanted Data Dashboard to respect the boundaries that they created by choosing specific storing places for different types of data (a result consistent with past work [62]): "*I already upload most of my files to Google Drive, Dropbox. I don't feel that bothered. But if this accesses even the files that I specify not to be* [accessed for a specific function], *that would really bother me.*" (P5)

In particular, when considering privacy data boundaries, several participants felt uneasy about Data Dashboard: "*You already have questions about the security of iCloud, Google Drive, etc. and that's reduced when you go to third-party tools. Especially when you see* [that] *the algorithm can see private information.*" (P18) Some participants imagined the negative consequences of having all data centralized in one system and wondered whether this is the right approach. For example, after seeing a recommendation to review cloud documents containing passport or credit numbers in Quick Actions, one participant wondered about the consequences for privacy: "*But there is a link to the file and it says what's in the file, so what else do I need as a thief? You just gave everything on a platter. It doesn't make any sense to me.*" (P4) Some participants also had issues with terms related to privacy in EYD, saying they were unsure of what "tracking data" meant: "*Tracking data.... what does it include?.... Not totally sure.*" (P12) We used these terms because they are common in similar tools, but these reactions show the importance of language for building trust in the system and building awareness of what private data users might have stored on their devices.

Centralization requires guarantees

Because of the potential to blur and distort boundaries, participants expressed a need for explicit guarantees around centralization: "*I just want it to be secure* [so that] *no one else can get into the system. I don't know how... I have no idea how you do it. How do you protect something like that?*" (P16) They wanted to make sure that the system would respect data boundaries and mitigate potential security risks: "*It's the central point that commands all my accounts so it has to be highly secure.*" (P7) Participants rationalized the potential adoption of Data Dashboard by referring to the terms and conditions that the system would have and would need to strictly apply (although there was no such thing in the prototype): "*If I am using the system, the terms and conditions that are mandatory to be agreed upon, should not include that your data has any potential of being shared by any third party for any commercial use.*" (P17) If the terms are clear and the system feels reliable, then centralization becomes acceptable: "*It depends on the privacy statement really. If the agreement seems good enough and it seems a reliable service, I don't mind using it and it going through my things.*" (P3) If it is possible to reinforce a boundary between local data and data in the cloud, then the worries disappear: "*If this is an online interface, I'd have problems with it. But if it was offline with no interaction other than backing up, which I control, then I'm completely fine.*" (P5) Once again, these reactions highlighted the need for clear explanations around data curation, with participants often blaming themselves for being "not techy."

Even with potential issues mitigated, some participants wondered about the feasibility of centralization. They reflected on the underlying conflict between their expectations and business practices that impose borders around data: "*My question is, can you actually make it? And I'll be the first guinea pig. The big giants, they have big muscle and they compete.* [...] *I don't know if it's possible because of the conflict of interest of these business things.* [...] *I am using Google Drive because I have a Google account, iCloud* [because I have] *an Apple account. Does this* [system] *need to be linked to a certain company or email account or something? If this is like my online version of my external hard drive, yeah, I would like to have everything consolidated under the one big, safe roof.*" (P19) This conflict required a broader guarantee, one that puts user needs above business needs: "*If there is such a magnificent creation down the road, it could be under the one roof, but also secure and safe. That's kind of my dream.*" (P19)

Customization upholds boundaries

Customization can make boundaries visible

Most participants had positive reactions to the customization options in Data Dashboard: "*I like choices, so I'm all for it.*" (P17) Some were confused by the link between the sidebar filters and the Settings page, but in general they thought that customizing was a good idea. For example, participants liked the option of sorting and hiding data types in Explore Your Data because it helped them see the things they wanted to prioritize: "*I'm the sort of person who would limit what I see here. I would keep photos, bookmarks* [...] *It's also good that I can reorganize. Minimize what I see and prioritize what I see first.*" (P9) Similarly, the option of sorting through data using the personalized filters excited participants because it made it easier to look at their data: "*This is great to look at what's old, what's inactive, unreachable, unused, these are really good.*" (P12) Together, these options made personal boundaries more visible and tangible, and helped participants navigate different data curation scenarios: "*Sometimes I need to see all of the information, sometimes if I am running on short time, I wanna do a narrow search, instead of always seeing everything.*" (P17)

Customization allows to manipulate boundaries

A consequence of perceiving boundaries as more visible was that participants also felt they had more options for manipulating them. Several participants thought that the default options in the system were "novel" and "comprehensive" enough to meet their needs. Choosing what to include or not made them feel more in charge of the system: "*I have some kind of authority to make a selection of what I want. This is my priority, I want to check duplicate documents. These options are good because then I also have a sort of selection power.*" (P15) But

an additional positive aspect was the option of creating your own filters through custom combinations in Settings. While this function was confusing for some participants, they generally saw it as a useful way of setting specific boundaries: "[It is] *cool* [that] *you can make your own filters. I like that 'cause I can figure out what's important to me, my own criteria, I like the customization aspect.*" (P18)

Customization reinforces control

One possible disadvantage of customization is the time required for it, as we mentioned in the related work section. When discussing the customization options in Data Dashboard, and the Settings page in particular, participants reflected on the time required for setting the filters as they preferred. Several participants imagined it would not take a lot of time because the system provided default options to choose from: "*I don't think it would take that much time.*" (P12) Others imagined it would take some planning on their part: "*I think it might take me probably a couple of hours to figure this out. I would plan it out before I actually use the system.*" (P10) But in the end they thought that investing time in customizing was worth it: "*I don't mind investing the time if this was going to figure out the system, I would get out a piece of paper and think of recent data to get rid of.*" (P10). Taking the time to customize would lead to the system respecting their own priorities and boundaries: "*I think filters are always useful. I don't mind* [spending time customizing] *because everyone has something different they're looking for so it's good to have these filter so they have what's most important to them.*" (P3) Participants perceived this trade-off between time and control as necessary to counterbalance any potential risks coming from other aspects of the system. Feeling in control made them feel safer: "*When it comes to data, I am very cautious. So if I have the settings set correctly, hopefully it acts accordingly.*" (P19)

Overall, these positive reactions show that customization upholds data boundaries by making them more visible, allowing users to manipulate them, and reinforcing their control and involvement in data curation.

DISCUSSION

Our results around centralization and customization show the importance of data boundaries for curation. While the centralized nature of Data Dashboard had several advantages for curation, it also had the potential for blurring boundaries. Customization options, on the other hand, tended to uphold participants' boundaries. Past work on centralized personal archives [62] similarly found that storing personal data in a central place undermines the "different facets of the self," ignoring differences between unremarkable and valuable content. Our study suggests that customization can offset the negative aspects of centralization and better help users demarcate different types of content. Thus, combining the two approaches is a promising design direction that can balance conflicting user needs. Below, we reflect on some of the evaluation results and outline how to move forward in designing curation tools.

Integrating data boundaries into design

An implication from our work is that filtering and sorting data into "chunks" based on types or other automatic categorizations can be helpful. The examples in Data Dashboard suggest that working on algorithms that can filter and recognize different types of information is a promising direction. While there are previous technical efforts along these lines [7,24,86], more work is necessary to address the functional design of these mechanisms. One possibility is to envision more granular mechanisms that integrate data boundaries in their design and combine the key positive aspects of both centralization and customization. For example, we can imagine users being able to directly review and manipulate boundaries at the level of individual items: they could be looking at a document containing their passport number, and choose how private, important, or relevant they consider that type of information to be. Then, they could set the boundary to be valid only for a certain amount of time and apply it to similar items or just the single item, set it for one platform or multiple, all from a central control point. An explicit boundary management mechanism could allow users to better manage breaking points, defining whether a boundary can be broken and in which scenarios (e.g., "keep passport data off the cloud, unless I am travelling").

Another potential avenue for integrating boundaries into design is to think of them as objects that users can share and exchange. One goal of our study was to see if customization can help in designing a system that can accommodate individual user needs. Our results show that this is possible, although customization can be labor-intensive. However, we know that while each person is different, there might be clusters of users who take a similar approach to data curation. We can envision a way for users to adopt and then adapt the boundaries of other people. A system like Data Dashboard could use a sharing infrastructure to provide personalized defaults based on user models. This could reduce the time required for setting personal filters and make the system more appealing to those who are not willing to invest time in customization [48]. Of course, similar functions should be designed in a privacy-preserving way. Past work on digital collections [100] and smart journaling [35] can also help define ways for automating and enforcing the customization of data boundaries, envisioning them as mechanisms to group related content without the need for an overall complete or encompassing archive.

Rethinking the language of personal data

Another implication of our study is about rethinking the language of personal data curation, a process that has relevance beyond practical actions like deleting files, moving documents into folders, or uninstalling mobile applications because of storage space. These are concrete actions that need support. But the broader consequences of curation are about control and power relationships with the companies that collect, store, and access data as part of a "surveillance" apparatus [105]. The language used to explain where personal data is stored (from documents to location history) and what happens to it is essential for understanding and controlling its use. In our study, we saw that participants sometimes struggled with some of the terms used in the system, like "syncing data" or "tracking data" and pointed to the importance of language and terms for building trust in the system. These recurring reactions from a varied sample of "not techy" participants show how it is necessary to work on improving the language of everyday user

interfaces, especially in privacy and security-related scenarios. How can users curate and control their personal data if they are not aware of the vast range of personal data that exist and and what common actions in everyday tools do? Past work shows how language inconsistencies are common in popular operating systems for actions as simple as deleting data [45]. We argue that simplifying and unifying the language around personal data is necessary for supporting everyday technology users. As Zuboff says in describing strategies for fighting surveillance capitalism, naming is the first step in "confronting and taming the unprecedented." [68]

To help navigate a looming data-driven society [37], privacy and data regulations are paramount [1, 2], but design practice has a role to play too. There is an opportunity for future design initiatives to start addressing the underlying gap in personal data literacy. This would involve rethinking common technical terms for describing personal data, and creating initiatives, both within tools but also outside them, to build a common vocabulary around personal data. We imagine this as a community effort where designers and researchers explore how user language and the grammar of possible actions within data management are intertwined, an idea brought forward by previous work [50]. One option could be to promote consistency across operating systems, applications, and tools, creating a grounded, standard vocabulary that users are familiar with. This could involve a systematic study of users' terms that could be then integrated into "personal data standards," just like we have conventions for common design elements. Another option would be to let users define and teach their own language to the tools through mechanisms that leverage a link between individual items and categories of data. For example, systems could give users an example item, ask them to categorize it using their own words, and then apply this user-generated vocabulary in the interface.

Envisioning a post-cloud future

Finally, another important thread in our work is the contrast between physical devices and cloud platforms. Participants' reactions around centralization and privacy boundaries emphasized the tension between local and online data storage. Current setups force users to go through a third party for accessing data across devices. There is no practical alternative. This dynamic has consequences for personal data boundaries, with tools creating boundaries of their own and imposing specific structures on data (e.g., having to store files in a Dropbox folder if you want to access them on more than one device, rather than in a local and more personal structure). But what if we imagined an alternative that emphasizes users' perception of local devices as central, rather than giving priority to the cloud? This would require interoperability across data formats and storage devices to easily synchronize data without the need for a cloud-first structure. The tools and user interfaces would be similar to existing ones, but the functional paradigm would be different.

We imagine that future data infrastructures could rely on local, "social" clouds, maybe with a physical device acting as the central node connecting data links across devices. For example, members of a family, neighborhood, or social group could set up a "local cloud" that allows personal data to move freely among known and trusted members' devices. Items could be synchronized to different devices on an individual basis rather than by uploading them on a distant server. The local cloud would take advantage of the multiple devices for maximizing storage and preventing data loss through redundancy. A similar approach could make it possible to make data available across devices, while still respecting data boundaries and local structures. These ideas are speculative, but not entirely new [91,92]. Before the advent of commercial cloud computing, investigations on peer-to-peer and decentralized technologies for data access and storage were popular [4, 26, 58, 82]. Today, the increasingly critical discourse around technology companies, cloud platforms, and privacy might signal a shift in perceptions. Within this evolving landscape, there is renewed interest around decentralized computing [99], cooperative storage [90], and cross-media information spaces [85]. All together, current efforts and past explorations point to a promising domain that future design work can help explore.

LIMITATIONS AND FUTURE WORK

Our sample is meant to be generative and not statistically representative, so it has some limitations around age and gender balance. We did not screen for gender identity, but related studies have similar samples [21, 22] and notice no apparent differences in attitudes based on gender. We screened for age, but we did not get interest from many respondents over 50. Future studies can complement our work with a more representative sample to see if and how our results transfer to a broader population. Future work should also explore quantitative measures of user satisfaction for a design approach similar to ours, how to extend our efforts to other stages of data curation (i.e., retrieval), and what might be the psychological or logistical effects of spending time curating data (e.g., enjoyment, satisfaction, time saved, more efficient retrieval).

CONCLUSION

Storing, managing, and curating personal data is a challenging process made more complex by the mix of devices and cloud platforms that people regularly use. In this paper, we explore how centralization and customization can support people's behaviors. We found that centralization blurs personal data boundaries, while customization upholds them. Using this specific analysis lens helped us outline key challenges for designing data curation tools. As our relationship with data evolves and likely becomes more complicated, we see taking a step back and reflecting on key design decisions as essential for shaping future products. Our work shows that there is great potential for exploring how to design tools that integrate data boundaries as a core part of their functionality and provide new mechanisms for managing them. We hope our work will provide a starting point for innovative tools that can define future paradigms for data management and curation.

ACKNOWLEDGEMENTS

The prototype includes icons from Flat Icon. We thank the anonymous reviewers, along with Ivan Beschastnikh, Yelim Kim, Ashish Chopra, Taslim Arefin Khan, and Paul Bucci for their comments. This work was supported by the grant NSERC RGPIN-2017-04549 "Highly personalized user interfaces."

REFERENCES

[1] 2018. California Consumer Privacy Act (CCPA). (Oct. 2018). https://oag.ca.gov/privacy/ccpa

[2] 2018. Guide to the General Data Protection Regulation. (May 2018). https://www.gov.uk/government/publications/guide-to-the-general-data-protection-regulation

[3] Alessandro Acquisti, Laura Brandimarte, and George Loewenstein. 2015. Privacy and human behavior in the age of information. *Science* 347, 6221 (2015), 509–514.

[4] Atul Adya, William J. Bolosky, Miguel Castro, Gerald Cermak, Ronnie Chaiken, John R. Douceur, Jon Howell, Jacob R. Lorch, Marvin Theimer, and Roger P. Wattenhofer. 2003. Farsite: Federated, Available, and Reliable Storage for an Incompletely Trusted Environment. *SIGOPS Oper. Syst. Rev.* 36, SI (Dec. 2003), 1–14. DOI: http://dx.doi.org/10.1145/844128.844130

[5] Nikola Banovic, Fanny Chevalier, Tovi Grossman, and George Fitzmaurice. 2012. Triggering Triggers and Burying Barriers to Customizing Software. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2717–2726. DOI: http://dx.doi.org/10.1145/2207676.2208666

[6] Susanne Barth and Menno DT De Jong. 2017. The privacy paradox–Investigating discrepancies between expressed privacy concerns and actual online behavior–A systematic literature review. *Telematics and Informatics* 34, 7 (2017), 1038–1058.

[7] Ron Bekkerman. 2004. Automatic categorization of email into folders: Benchmark experiments on Enron and SRI corpora. (2004).

[8] Russell W Belk. 1988. Possessions and the extended self. *Journal of consumer research* 15, 2 (1988), 139–168.

[9] 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

[10] David Benyon. 2014. *Designing interactive systems: A comprehensive guide to HCI, UX and interaction design*. Pearson Edinburgh.

[11] 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.

[12] 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

[13] 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

[14] Ofer Bergman, Simon Tucker, and Somaya Dahamshy. 2018. The effect of demoting near-duplicate pictures. *Proceedings of the Association for Information Science and Technology* 55, 1 (2018), 755–756.

[15] Ofer Bergman and Steve Whittaker. 2016. *The Science of Managing Our Digital Stuff*. MIT Press.

[16] Ofer Bergman, Steve Whittaker, and Yaron Frishman. 2019. Let's get personal: the little nudge that improves document retrieval in the Cloud. *Journal of Documentation* 75, 2 (2019), 379–396.

[17] Jan Blom. 2000. Personalization: A Taxonomy. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems (CHI EA '00)*. Association for Computing Machinery, New York, NY, USA, 313–314. DOI: http://dx.doi.org/10.1145/633292.633483

[18] 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

[19] Robert Capra, Julia Khanova, and Sarah Ramdeen. 2013. Work and personal e-mail use by university employees: PIM practices across domain boundaries. *Journal of the American Society for Information Science and Technology* 64, 5 (2013), 1029–1044.

[20] Marta E. Cecchinato, Anna L. Cox, and Jon Bird. 2015. Working 9-5? Professional Differences in Email and Boundary Management Practices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3989–3998. DOI: http://dx.doi.org/10.1145/2702123.2702537

[21] EunJeong Cheon and Norman Makoto Su. 2018a. 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

[22] EunJeong Cheon and Norman Makoto Su. 2018b. "Staged for Living": Negotiating Objects and Their Values over a Porous Boundary. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article Article 36 (Nov. 2018), 24 pages. DOI: http://dx.doi.org/10.1145/3274305

[23] Victoria Clarke and Virginia Braun. 2014. Thematic analysis. In *Encyclopedia of critical psychology*. Springer, 1947–1952.

[24] Gabor Cselle, Keno Albrecht, and Rogert Wattenhofer. 2007. BuzzTrack: topic detection and tracking in email. In *Proceedings of the 12th international conference on Intelligent user interfaces*. ACM, 190–197.

[25] Edward Cutrell, Daniel Robbins, Susan Dumais, and Raman Sarin. 2006. Fast, Flexible Filtering with Phlat. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 261–270. DOI: http://dx.doi.org/10.1145/1124772.1124812

[26] Frank Dabek, M. Frans Kaashoek, David Karger, Robert Morris, and Ion Stoica. 2001. Wide-Area Cooperative Storage with CFS. *SIGOPS Oper. Syst. Rev.* 35, 5 (Oct. 2001), 202–215. DOI: http://dx.doi.org/10.1145/502059.502054

[27] Jesse David Dinneen and Charles-Antoine Julien. 2019a. The ubiquitous digital file: A review of file management research. *Journal of the Association for Information Science and Technology* (2019).

[28] Jesse David Dinneen and Charles-Antoine Julien. 2019b. What's in people's digital file collections? *Proceedings of the Association for Information Science and Technology* 56, 1 (2019), 68–77.

[29] Jesse David Dinneen, Charles-Antoine Julien, and Ilja Frissen. 2019. The Scale and Structure of Personal File Collections. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, Article 327, 12 pages. DOI:http://dx.doi.org/10.1145/3290605.3300557

[30] Jesse David Dinneen, Fabian Odoni, Ilja Frissen, and Charles-Antoine Julien. 2016. Cardinal: Novel software for studying file management behavior. In *Proceedings of the 79th ASIS&T Annual Meeting: Creating Knowledge, Enhancing Lives through Information & Technology*. American Society for Information Science, 62.

[31] Delphine Dion, Ouidade Sabri, and Valérie Guillard. 2014. Home sweet messy home: Managing symbolic pollution. *Journal of Consumer Research* 41, 3 (2014), 565–589.

[32] 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

[33] Dropbox. n.d. How to use selective sync. (n.d.). https://help.dropbox.com/installs-integrations/sync-uploads/selective-sync-overview

[34] Susan Dumais, Edward Cutrell, J. J. Cadiz, Gavin Jancke, Raman Sarin, and Daniel C. Robbins. 2016. Stuff I've Seen: A System for Personal Information Retrieval and Re-Use. *SIGIR Forum* 49, 2 (Jan. 2016), 28–35. DOI:http://dx.doi.org/10.1145/2888422.2888425

[35] Chris Elsden, Abigail C. Durrant, and David S. Kirk. 2016a. It's Just My History Isn't It? Understanding Smart Journaling Practices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2819–2831. DOI: http://dx.doi.org/10.1145/2858036.2858103

[36] Chris Elsden, David S Kirk, and Abigail C Durrant. 2016b. A quantified past: Toward design for remembering with personal informatics. *Human–Computer Interaction* 31, 6 (2016), 518–557.

[37] Chris Elsden, Mark Selby, Abigail Durrant, and David Kirk. 2016c. Fitter, happier, more productive: what to ask of a data-driven life. *Interactions* 23, 5 (2016), 45–45.

[38] 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

[39] 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

[40] Leah Findlater and Joanna McGrenere. 2010. Beyond performance: Feature awareness in personalized interfaces. *International Journal of Human-Computer Studies* 68, 3 (2010), 121–137.

[41] Saul Greenberg and Ian H Witten. 1985. Adaptive personalized interfaces—A question of viability. *Behaviour & Information Technology* 4, 1 (1985), 31–45.

[42] Jane Gruning. 2018. Displaying Invisible Objects: Why People Rarely Re-read E-books. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 139, 12 pages. DOI: http://dx.doi.org/10.1145/3173574.3173713

[43] 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

[44] 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

[45] Andreas Gutmann and Mark Warner. 2019. Fight to Be Forgotten: Exploring the Efficacy of Data Erasure in Popular Operating Systems. In *Annual Privacy Forum*. Springer, 45–58.

[46] Jacek Gwizdka and Mark Chignell. 2007. Individual Differences. In *Personal Information Management*, William Jones and Jaime Teevan (Eds.). University of Washington Press, Seattle and London, Chapter 12, 206–220.

[47] Mona Haraty and Joanna McGrenere. 2016. Designing for Advanced Personalization in Personal Task Management. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS '16)*. Association for Computing Machinery, New York, NY, USA, 239–250. DOI: http://dx.doi.org/10.1145/2901790.2901805

[48] Mona Haraty, Joanna McGrenere, and Andrea Bunt. 2017a. Online Customization Sharing Ecosystems: Components, Roles, and Motivations. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 2359–2371. DOI: http://dx.doi.org/10.1145/2998181.2998289

[49] Mona Haraty, Zhongyuan Wang, Helen Wang, Shamsi Iqbal, and Jaime Teevan. 2017b. Design and in-situ evaluation of a mixed-initiative approach to information organization. *Journal of the Association for Information Science and Technology* 68, 9 (2017), 2211–2224.

[50] 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

[51] Sarah Henderson. 2009. Personal Document Management Strategies. In *Proceedings of the 10th International Conference NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction (CHINZ '09)*. ACM, New York, NY, USA, 69–76. DOI:http://dx.doi.org/10.1145/1577782.1577795

[52] 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.

[53] Victoria Hollis, Artie Konrad, Aaron Springer, Matthew Antoun, Christopher Antoun, Rob Martin, and Steve Whittaker. 2017. What does all this data mean for my future mood? Actionable Analytics and Targeted Reflection for Emotional Well-Being. *Human–Computer Interaction* 32, 5-6 (2017), 208–267.

[54] 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

[55] 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

[56] 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

[57] William Jones, Predrag Klasnja, Andrea Civan, and Michael L. Adcock. 2008. The Personal Project Planner: Planning to Organize Personal Information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 681–684. DOI: http://dx.doi.org/10.1145/1357054.1357162

[58] Mahesh Kallahalla, Erik Riedel, Ram Swaminathan, Qian Wang, and Kevin Fu. 2003. Plutus: Scalable Secure File Sharing on Untrusted Storage. In *Proceedings of the 2nd USENIX Conference on File and Storage Technologies (FAST '03)*. USENIX Association, USA, 29–42.

[59] 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

[60] 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

[61] David S. Kirk, Shahram Izadi, Abigail Sellen, Stuart Taylor, Richard Banks, and Otmar Hilliges. 2010. Opening up the Family Archive. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work (CSCW '10)*. Association for Computing Machinery, New York, NY, USA, 261–270. DOI:http://dx.doi.org/10.1145/1718918.1718968

[62] 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

[63] Wendy E. Mackay. 1991. Triggers and Barriers to Customizing Software. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '91)*. Association for Computing Machinery, New York, NY, USA, 153–160. DOI: http://dx.doi.org/10.1145/108844.108867

[64] Catherine C Marshall. 2008a. Rethinking personal digital archiving, Part 1: Four challenges from the field. *D-Lib Magazine* 14, 3/4 (2008), 2.

[65] Catherine C Marshall. 2008b. Rethinking personal digital archiving, part 2: implications for services, applications, and institutions. *D-Lib Magazine* 14, 3 (2008), 3.

[66] Nolwenn Maudet. 2019. Dead Angles of Personalization: Integrating Curation Algorithms in the Fabric of Design. In *Proceedings of the 2019 on Designing Interactive Systems Conference (DIS '19)*. Association for Computing Machinery, New York, NY, USA, 1439–1448. DOI: http://dx.doi.org/10.1145/3322276.3322322

[67] Joanna McGrenere, Ronald M. Baecker, and Kellogg S. Booth. 2002. An Evaluation of a Multiple Interface Design Solution for Bloated Software. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '02)*. Association for Computing Machinery, New York, NY, USA, 164–170. DOI:http://dx.doi.org/10.1145/503376.503406

[68] John Naughton. 2019. 'The goal is to automate us': welcome to the age of surveillance capitalism. *The Guardian* (Jan. 2019). https://www.theguardian.com/technology/2019/jan/20/shoshana-zuboff-age-of-surveillance-capitalism-google-facebook

[69] Midas Nouwens, Carla F. Griggio, and Wendy E. Mackay. 2017. "WhatsApp is for Family; Messenger is for Friends": Communication Places in App Ecosystems. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 727–735. DOI: http://dx.doi.org/10.1145/3025453.3025484

[70] 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

[71] William Odom, Richard Banks, David Kirk, Richard Harper, Siân Lindley, and Abigail Sellen. 2012a. 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

[72] William Odom, Abi Sellen, Richard Harper, and Eno Thereska. 2012b. 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

[73] 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

[74] 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

[75] Kyong Eun Oh. 2017. Types of personal information categorization: Rigid, fuzzy, and flexible. *Journal of the Association for Information Science and Technology* 68, 6 (2017), 1491–1504.

[76] Daniel Orth, Clementine Thurgood, and Elise Van Den Hoven. 2019. Designing Meaningful Products in the Digital Age: How Users Value Their Technological Possessions. *ACM Trans. Comput.-Hum. Interact.* 26, 5, Article Article 34 (Aug. 2019), 28 pages. DOI: http://dx.doi.org/10.1145/3341980

[77] Leysia Palen and Paul Dourish. 2003. Unpacking "Privacy" for a Networked World. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. Association for Computing Machinery, New York, NY, USA, 129–136. DOI:http://dx.doi.org/10.1145/642611.642635

[78] Sandra Petronio. 2002. *Boundaries of privacy: Dialectics of disclosure*. Suny Press.

[79] James Pierce and Eric Paulos. 2014. Counterfunctional Things: Exploring Possibilities in Designing Digital Limitations. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14)*. Association for Computing Machinery, New York, NY, USA, 375–384. DOI: http://dx.doi.org/10.1145/2598510.2598522

[80] Eugenia Politou, Efthimios Alepis, and Constantinos Patsakis. 2018. Forgetting personal data and revoking consent under the GDPR: Challenges and proposed solutions. *Journal of Cybersecurity* 4, 1 (2018), tyy001.

[81] Philip Raschke, Axel Küpper, Olha Drozd, and Sabrina Kirrane. 2017. Designing a GDPR-Compliant and Usable Privacy Dashboard. In *IFIP International Summer School on Privacy and Identity Management*. Springer, 221–236.

[82] Antony Rowstron and Peter Druschel. 2001. Pastry: Scalable, Decentralized Object Location, and Routing for Large-Scale Peer-to-Peer Systems. In *Middleware 2001*, Rachid Guerraoui (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 329–350.

[83] Woo-Jong Ryu, Jung-Hyun Lee, Kang-Min Kim, and SangKeun Lee. 2017. MeCurate: Personalized Curation Service Using a Tiny Text Intelligence. 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, CHE, 269–272. DOI: http://dx.doi.org/10.1145/3041021.3054723

[84] Aakash Sahney and Dave Loxton. 2017. Introducing Backup and Sync for Google Photos and Google Drive. (July 2017). https://blog.google/products/photos/introducing-backup-and-sync-google-photos-and-google-drive/

[85] Beat Signer. 2019. Towards Cross-Media Information Spaces and Architectures. In *2019 13th International Conference on Research Challenges in Information Science (RCIS)*. IEEE, 1–7.

[86] Frank Smadja. 2003. Automatic categorization of documents based on textual content. (Sept. 16 2003). US Patent 6,621,930.

[87] S Shyam Sundar and Sampada S Marathe. 2010. Personalization versus customization: The importance of agency, privacy, and power usage. *Human Communication Research* 36, 3 (2010), 298–322.

[88] Laurel Swan, Alex S Taylor, and Richard Harper. 2008. Making place for clutter and other ideas of home. *ACM Transactions on Computer-Human Interaction (TOCHI)* 15, 2 (2008), 9.

[89] 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.

[90] Yee-Yang Teing, Ali Dehghantanha, Kim-Kwang Raymond Choo, Tooska Dargahi, and Mauro Conti. 2017. Forensic investigation of cooperative storage cloud service: Symform as a case study. *Journal of forensic sciences* 62, 3 (2017), 641–654.

[91] Max Van Kleek and Kieron OHara. 2014. The future of social is personal: The potential of the personal data store. In *Social Collective Intelligence*. Springer, 125–158.

[92] Max Van Kleek, Daniel A. Smith, Dave Murray Rust, Amy Guy, Kieron O'Hara, Laura Dragan, and Nigel R. Shadbolt. 2015. Social Personal Data Stores: The Nuclei of Decentralised Social Machines. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion)*. Association for Computing Machinery, New York, NY, USA, 1155–1160. DOI: http://dx.doi.org/10.1145/2740908.2743975

[93] 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

[94] 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

[95] Francesco Vitale, William Odom, and Joanna McGrenere. 2019. Keeping and Discarding Personal Data: Exploring a Design Space. In *Proceedings of the 2019 on Designing Interactive Systems Conference (DIS '19)*. ACM, New York, NY, USA, 1463–1477. DOI:http://dx.doi.org/10.1145/3322276.3322300

[96] Amy Voida, Judith S. Olson, and Gary M. Olson. 2013. Turbulence in the Clouds: Challenges of Cloud-based Information Work. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 2273–2282. DOI:http://dx.doi.org/10.1145/2470654.2481313

[97] Stephen Voida, W. Keith Edwards, Mark W. Newman, Rebecca E. Grinter, and Nicolas Ducheneaut. 2006. Share and Share Alike: Exploring the User Interface Affordances of File Sharing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. Association for Computing Machinery, New York, NY, USA, 221–230. DOI: http://dx.doi.org/10.1145/1124772.1124806

[98] Stephen Voida 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

[99] David Vorick and Luke Champine. 2014. Sia: Simple decentralized storage. *Nebulous Inc* (2014).

[100] 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

[101] 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

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

[103] Clemens Zeidler, Christof Lutteroth, and Gerald Weber. 2013. An Evaluation of Advanced User Interface Customization. In *Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration (OzCHI '13)*. Association for Computing Machinery, New York, NY, USA, 295–304. DOI: http://dx.doi.org/10.1145/2541016.2541037

[104] 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

[105] Shoshana Zuboff. 2015. Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* 30, 1 (2015), 75–89.

[106] Xenia Zürn, Mendel Broekhuijsen, Doménique van Gennip, Saskia Bakker, Annemarie Zijlema, and Elise van den Hoven. 2019. Stimulating Photo Curation on Smartphones. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval (CHIIR '19)*. Association for Computing Machinery, New York, NY, USA, 255–259. DOI: http://dx.doi.org/10.1145/3295750.3298947