

---

# Configuring Personal Data for a Quantified-Self Archive

**Ciaran B. Trace**

School of Information  
The University of Texas at Austin  
Austin, TX 78701, USA  
cbtrace@austin.utexas.edu

**Yan Zhang**

School of Information  
The University of Texas at Austin  
Austin, TX 78701, USA  
yanz@ischool.utexas.edu

## ABSTRACT

An audience exists for personal information, including quantified-self data, beyond an individual's social network and social communities. In the era of big data, the research and policy arenas are two areas where up-to-date assemblages of personal information have market value. In this ongoing study, we examine the long-term value of small data [2], acknowledging that there is also a societal need and an audience for rich, personalized collections of digital self-tracking records. Using qualitative research methods, we interviewed 18 people to explore the nature of self-tracking data that exists as a byproduct of daily life, and their sense of why and how their data could be archived for posterity. In the process, the intellectual and design challenges of a digital quantified-self archive are explored.

## KEYWORDS

Personal informatics; quantified-self; self-tracking; personal data; small data; archive; digital archive

## INTRODUCTION

The drive to capture and archive aspects of the human experience has deep roots in Western culture [16, 1]. In our ongoing research we seek to examine the ways in which self-tracking data has

---

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).

*CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland, UK.*

© 2019 Copyright is held by the author/owner(s).

ACM ISBN 978-1-4503-5971-9/19/05.

DOI: <https://doi.org/10.1145/3290607.3313017>

### Research in Progress [Trace & Zhang]

**Participants:** Eighteen adults (10 females and 8 males) between the ages of 18 and 42 participated in the study. Fourteen participants identified as white, (3 of whom identified as Hispanic or Latino), three as Asian, and one as African American.

All participants had tracked some aspect of their health and fitness as it was a basic criterion for participating in the study. In addition, 13 had tracked some aspects of their personal finances, 11 had tracked body and medical symptoms, and 8 had tracked some aspects of their environment.

Fourteen participants had tracked other aspects of their lives including habits and goals, travel/location, schedule, useful information, public transportation usage, computer usage, dreams, sociability, and utility usage.

**Methods:** Participants who signed up for the study were directed to complete an online questionnaire. The survey gathered demographic data and participants' basic characteristics vis-à-vis their self-tracking behavior and their experience in using self-tracking apps. In-person interviews were then conducted to ascertain participants' beliefs regarding the long-term use and value of quantified-self data. Interviews ranged from 44 to 153 minutes in length (Mean = 72 minutes).

**Analysis:** Survey data was analyzed using descriptive statistics. All interviews were recorded and transcribed in full. The resulting 590 pages of qualitative data were coded using the software program, MAXQDA.

perceived meaning and value in and after the life of the creator, including how such data could form part of the larger historical record, curated in an institutional archive. It is the field of archival science that has developed a framework for the management of collections of personal records, at scale. In the United States this framework rests on 100 years of professionalization, and an associated network of repositories serving an increasingly diverse constituency (scholarly to avocational).

Archival collections document human activity in the form of literacy practices, including such “‘socially evolved and pattern activities’ as letter writing, keeping records and inventories, keeping a diary, writing memos, posting announcements, and so on” [11]. With the adoption of personal computing in the 1980’s, the archive has expanded to capture a world of increasing digitality. Such is the case with Salman Rushdie’s archive housed at Emory University, where evidence of his scholarly and personal activities is inscribed in journals, manuscripts, and correspondence, and in digital files on four Apple computers and a hard drive. With the rise of the smartphone in the 2000’s, archivists are turning their attention to curating and making accessible evidence of human activity in the post-PC era. As one such form of mobile literacy, self-tracking data is poised to become a new part of the historical and societal record. Stakeholder engagement is key to the process of archiving the mobile life – including those interested in understanding self-tracking in action (researchers, app developers, and self-trackers) and those interested in its curation (memory workers). This paper begins such efforts with a study of self-trackers and their perceptions of the long-term value of self-tracking data. In the process, the intellectual and design challenges of a digital quantified-self archive are explored.

### METHODS

Eighteen people who had self-tracked for six months or more were recruited for the study in fall 2017 (see sidebar 1). Participants initially completed a survey which gathered basic demographic data and characteristics vis-à-vis their self-tracking behavior. In-person semi-structured interviews were then conducted to ascertain the beliefs of the participants regarding the long-term use and value of personal quantified-self data. The interviews were aligned around two sets of questions. The first set invited each participant to consider their own self-tracking practices across their lifetime. In the second set of questions (the results of which are outlined in this paper), we invited participants to speculate on the ‘what if’ - what if 50 years from now their self-tracking data was donated to an archival repository. Participants were asked to consider this scenario from their own perspective and from the perspective of a future user (the intended audience for their self-tracking data).

### FINDINGS

#### Emotive Aspects of the QS Archive

Prior research suggests that the act of passing on historical information outside of a family context can be “less controversial” given that cultural heritage institutions provide for more temporal distance between the creator and the subsequent user of the digital information [15]. Our participants were conflicted about the idea that their self-tracking data would one day be preserved in an archive and made available to the public. Positive emotions centered on the hope that the data

### Emotive and Value Driven Aspects of the QS Archive

“I don't want to just die and be like, okay, that was [participant's name]. Bye. Like, I'd like people to be able to look up to me and ... I'd like to be someone else's goal.” [P8]

“I'm totally fine with it. I realize that's a bit contradictory because I'm saying, “I don't wanna share this with people, it's so private.” I guess to the extent that it can help people understand life in 2017, I think that's pretty cool.” [P13]

“I think the furthest away that we'd go from my own life, that would have, you know, obviously, no impact on, on me and I would have no hesitation providing any of that.” [P 15]

“It's just such a small aspect of my life to where I don't think ... it's a big aspect to me... So, I don't know if it could actually influence how people know my life story. But they could know a part of a time period in my life.” [P12]

“I think the biggest thing they would notice is that, how I've tracked more and more things over time. And my attitudes towards kind of just efficiency and... I think that's just a reflection of kind of the 21st century times where we're just trying to do a lot in a very little amount of time... What would be missing is I think still a lot of the kind of every day kind of... bits and pieces between a lot of this...So this kind of just tells on a very typical day...but I don't think it gives a lot of the in between kind of details” [P9].

would be of benefit to others, on the satisfaction of having their progress permanently memorialized, on the sense of awe at being remembered as part of history, and on the curiosity in looking back on personal activities and in imagining what other people would do with the data.

The potential for data to embarrass them or their family was the main hesitancy. Reticence was also on show, with some participants expressing a desire for anonymity and to keep their data private. This was particularly the case with participants who participated in analogue and/or more qualitative forms of self-tracking that mimicked the more intimate genres of a diary or journal. Yet, present in some of the participants' answers was the notion that time would overcome such qualms. The longer the data survived, the fewer concerns participants had for how the data would be received and used, with concerns also diminishing about the impact the data could have on their own legacy and that of loved ones. In this case death was seen as the great leveler, generally putting all such cares to rest.

### Articulating the Value of the QS Archive

Participants viewed some types and forms of self-tracking data as having insufficient value to warrant preservation in the archive. Data that fell within this category was limited in temporal scope and documented isolated activities deemed insufficiently interesting or all-encompassing. Despite such misgivings, participants were able to articulate conditions in which their self-tracking data could have value to others, including what it could reveal about the self, and society.

Continuous historical self-tracking data tied to the individual was seen as providing concrete information about personal lifestyle patterns and achievements. Participants thought that associating that data with additional baseline information could also shed light on broader societal issues such as spatial and temporal patterns of commuting, fluctuations in the cost of living, changes to the built environment, etc. From a more speculative standpoint, participants felt that their historical self-tracking data could be useful in making inferences about an individual's personality (e.g. ‘organized’ or ‘obsessed with numbers’), and about culture and society at a particular time and place (e.g. the pervasive reliance on technology, and the greater individualism and image consciousness of the millennial generation). Participants were also cognizant of issues around the paucity, accuracy and/or truthfulness of certain self-tracking data, particularly data entered manually. In these instances, participants speculated about how these obstacles could be overcome including through the use of embedded sensors to augment self-tracking data with automatically captured biophysical data.

### Overall Design Considerations for the QS Archive

Participants had an emergent understanding of the quantified-self archive, speculating on how the data should be presented to the public as a personal portrait of the self and, more fundamentally, how self-tracking technologies could be redesigned to better support meaning making in the future. Some of the participants' discussion on how to settle the data in place revolved around the extent to which self-tracking data needed to be curated to be intelligible for future users. For example, did people think their data needed to be ‘tamed’ [9], ‘transformed’ [8], or ‘pre-processed’ [10] prior to

### Design Considerations for a QS Archive

“Like I mentioned a little bit earlier [social media] gives kind of meaning to all of the in between points of what the self-tracking kind of data... pertains to... And then I think the other thing would just be... my calendar app as well... Just to kind of see as the day kind of progresses...what that self-tracking data is reflective of... Just about everything that I do for the most part, or major kind of important meetings, whatever it is, that's on my calendar.” [P9]

“Sometimes I'll go on a walk and I'll take pictures of the walks I go on... The day we had, you know, 22,000 steps in London, I got tons of photos from that day, so that might make it more meaningful.” [P2]

“I wanna tell you what ... what my brain was like at each year. Like ... 2008, was like me in panic mode, 'cause my mom was like, "I'm not paying for college." And I'm like, "Oh crap. I need to get together a plan.”” [P12]

“As you increase the amount of objective data you increase the importance of the accompanying subjective interpretive data. So, I think the best thing you could do to understand... is just to like interview friends about my personal habits.” [P10]

“I guess if they were going to draw any like historical conclusions, I'd hope it wouldn't just be mine. I'd hope that they would kind of put it alongside other data points. I'd hope that they would have it like, oh this is how you fit into this.” [P7]

or as part of being handed over to an archive. Participants felt that self-tracking data should be ‘augmented’ to create an archive richer in context and overt or possible meanings. Amplifying the narrative potential of a quantified-self archive meant preserving self-tracking data in parallel with the data from a person’s documentary ecosystem. This data stood in for what P9 called the “in-between points” of life, along with data that functioned as what we term ‘emotive in-fill.’

A point of contention was whether an attempt should be made to ‘locate’ historical self-tracking data in the mind of the individual, using rich description to announce its creator’s intentions more plainly. This is analogous to the argument [3] that qualitative or textual data must be derived from, and combined with, quantified-self data to capture a richer context of lived experience. Some participants felt that personal written reflections outlining a person’s history, state of mind, feelings, and motivations would be revelatory for any potential user of the archive. Others viewed this information as suspect in that it was non-concomitant to how the activity of self-tracking was originally carried out. Overall, the idea of archiving self-tracking data alongside browser histories, social media posts, calendars, financial records, photographs, and interviews was to provide a space in which objective, subjective, emotive, and impressionistic data could be used in concert to shed light on an individual lived experience.

Questions regarding the form that the self-tracking data should take in the archive resurfaced an idea alluded to in discussions about the value of historical self-tracking data. Participants understood the quid pro quo of using self-tracking apps, an arrangement in which there are dueling claims to data. Participants understood data as simultaneously personal and collective in nature, a private resource for the individual and a private commercial asset for the app development company. In the case of five participants, the notion that the personal provenance of self-tracking data might take center stage in the public archive was set aside in favor of an imagined archive in which the individual’s data was presented in combination with other app users. In seeking to co-mingle provenance and amalgamate their data as part of a public archive, participants privileged the value of pooled data in helping to see trends and make comparisons, and simply as a way of indicating that they were engaged in “something bigger” (P18) than themselves.

### Presenting Data from the QS Archive

Whether individualized or aggregated in form, self-tracking data was thought to have the most meaning when rendered to the imagined public in visual modes (pictures, graphs, and 3D renderings, etc.) and temporally situated. Raw/unprocessed data, representing daily or weekly increments was seen as too granular (and even boring) by some of the participants. These participants were more confident in the utility of providing access to processed or summative data that shed light on a person’s activities over monthly or yearly intervals. Such data was seen as capable of being read in a narrative manner, providing obvious points of reflection and comparison. Other participants saw the research value in providing public access to fine-grained daily or weekly detail, data that allowed the user to make their own extrapolations and determinations.

### Presenting Data from the QS Archive

“I don't think it should be daily 'cause there's a lot of days where I actually don't put anything in. I think it would be more data that points to achievements of specific goals or milestones because it's, it's long process that takes times based on like your weight from one point to another.” [P12]

“I think summative data... So, it's funny as an individual to say I'm not so important, but I think that it's important that we sort of put ourselves in the real place that, you know, one day I'll be dead, and my legacy may be moderate and it's okay. I'm not Bruno Latour... Uh, now I also want to say that it's very important to have the full raw data, which counters what I just said, for those that are interested in the specifics.” [P15]

“I feel like that would be silly to be, like, here's one person's ... Look, look how lazy he is and how few times he walks, or that he walked a whole bunch. I just don't think that that's interesting or that someone would care. I think that saying this big trend of a whole user base of Weight Watchers, that would be more interesting.” [P18]

” I think there are other things, such as like, barometric pressure or ...pupil dilation, or how other friends subjectively felt ... what kind of mood they felt I was in, that are things that are like, profoundly inaccessible to me, and if technology can construct workarounds and like access those, I think those could be huge well springs of information.” [P10]

In a nod to the materiality and ubiquity of smart phones today, participants mentioned the value in housing historical self-tracking data within and accessible through the original hardware and software, thus providing insights into how such technologies were amalgamated into daily living. Participants also speculated about how current self-tracking apps could be redesigned to make the data more meaningful both now and in the future. At issue here is the extent to which current self-tracking tools are designed to provide support for activities that involve long-term use and reflection using historical data [12, 5, 7]. In seeking to redesign what others have called an “interface to the record” [6], the participants placed an emphasis on redesigning self-tracking apps to boost their “expressive potential” [14] at the personal and collective level.

Echoing current trends, participants wished for an app that would allow their entire documentary ecosystem to be integrated into one digital space. Consideration was also given to redesigning apps to collect additional layers of contextual metadata, including data that could be added automatically (e.g. weather and location data), and data that the individual could add him/herself (including the ability to ‘tag’ notable data in some way). Design considerations were also aligned with a desire from some participants to foreground and heighten points of comparison (including with other app users), with a request that such insights be supported via the app interface. A request for additional ways to associate images with self-tracking data was also of concern, with a photo seen as a corroborating data point, capable of revealing more personal and emotive aspects of the activity tracked.

### SUMMARY

Gary Wolf's talk at the National Archives of the Netherlands in 2017 began a dialogue between the QS and archival community. Connections between archival science and HCI are also in a promising yet nascent phase [1, 4, 13]. This paper outlines our ongoing effort to bring these communities together to articulate what it means to archive QS data for the future. As part of this work we recognize that the focus of system designers and ‘design considerations’ can expand beyond immediate and individual interaction and use, to consider long-term and societal interaction with and use of systems and their associated data. In the first stage of the research, outlined here, self-trackers were interviewed to understand the perspective of data creators, those who would donate their personal digital traces to an archival repository. In the process an understanding of the personal, emotional, practical, and societal value of the data was unearthed, along with the meanings ascribed to data at various levels of granularity and aggregation. To conclude, design considerations for the QS archive were presented, allowing for their further investigation by researchers, developers, and the archival community alike.

### ACKNOWLEDGMENTS

This work was supported by the Governor Bill Daniel Fellowship, School of Information, The University of Texas at Austin.

## REFERENCES

- [1] Amelia Abreu and Amelia Acker. 2012. Traces of Heritage: an archival approach to digital collections. *Heritage Matters Workshop (CHI '12)*.
- [2] Amelia Abreu and Amelia Acker. 2013. Context and collection: a research agenda for small data. In *Proceedings of the iConference 2013*, 549–554.
- [3] Theresa Dirndorfer Anderson and Roberto Martinez-Maldonado. 2016. Building a qualified self around lifecycles of experience and thinking. In *Proceedings of the CHI 2016 Workshop on Personal Information Management (PIM 2016)*.
- [4] Jenny Bunn. 2015. Exploring the potential for cross disciplinary working with archives and records management. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*, 2169–2174.
- [5] Chris Elsdén. 2014. Situated remembering with digital technology. In *Proceedings of the 2014 companion publication on Designing interactive systems (DIS Companion '14)*, 145–149.
- [6] Chris Elsdén, Abigail C. Durrant, and David S. Kirk. 2016. 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)*, 2819–2831.
- [7] Chris Elsdén, David S. Kirk, and Abigail C. Durrant. 2016. A quantified past: toward design for remembering with personal informatics. *Human-Computer Interaction* 31, 6: 518–557.
- [8] Daniel A. Epstein, Alan Borning, and James Fogarty. 2013. Fine-grained sharing of sensed physical activity: a value sensitive approach. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing (UbiComp '13)*, 489–498.
- [9] Daniel A. Epstein, Felicia Cordeiro, Elizabeth Bales, James Fogarty, and Sean A. Munson. 2014. Taming data complexity in lifelogs: exploring visual cuts of personal informatics data. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14)*, 667–676.
- [10] Daniel A. Epstein, Bradley H. Jacobson, Elizabeth Bales, David W. McDonald, and Sean A. Munson. 2015. From "nobody cares" to "way to go!": a design framework for social sharing in personal informatics. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*, 1622–1636.
- [11] Colin Lankshear and Michele Knobel. 2006. *New Literacies: Everyday Practices & Classroom Learning*. McGraw-Hill.
- [12] Ian Li, Anind K. Dey, and Jodi Forlizzi. 2011. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th international conference on Ubiquitous computing (UbiComp '11)*, 405–414.
- [13] Elisabet M. Nilsson and Sofie Marie Ottsen Hansen. 2018. The co-archiving toolbox: designing conditions for diversity in public archives. In *Proceedings of the Design Research Society (DRS) 2018 International Conference: Catalyst, Volume 2*, 717–728.
- [14] William Odom, Richard Harper, Abigail Sellen, David Kirk, and Richard Banks. 2010. Passing on & putting to rest: understanding bereavement in the context of interactive technologies. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*, 1831–1840.
- [15] Lisa Thomas and Pam Briggs. 2014. An older adult perspective on digital legacy. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational (NordiCHI '14)*, 237–246.
- [16] Nora Young. 2012. *The Virtual Self: How Our Digital Lives Are Altering the World Around Us*. McClelland & Stewart.