Revelado: Exploring the Preservation of our Digital Data

Rebecca Gulotta
Carnegie Mellon
Human-Computer Interaction Institute
Pittsburgh, PA 15232 USA
rgulotta@andrew.cmu.edu

Haakon Faste
Carnegie Mellon
Human-Computer Interaction Institute
Pittsburgh, PA 15232 USA
hfaste@cs.cmu.edu

Jodi Forlizzi
Carnegie Mellon
Human-Computer Interaction Institute
Pittsburgh, PA 15232 USA
forlizzi@cs.cmu.edu

CHI 2012, May 5–10, 2012, Austin, TX, USA.
ACM xxx-x-xxxx-xxxx-x/xx/xx.

Abstract
Everyday, Internet users generate large amounts of digital data about their lives, habits, and experiences. Increasingly, this data is being preserved online – aided by the prevalence of social networking sites and third party systems that track and collect information about our activities in both the digital and physical world. Despite the overwhelming amount of data that is available, it is unclear how it might be transferred or understood by future generations. In particular, we are interested in the preservation and transmission of secret or discarded aspects of users' online identities. In this short paper, we describe the creation of a tool called Revelado, which allows users to store this type of information online such that it might be accessed by future generations. In doing so, we explore issues related to the ownership of information, digital identity, and the inheritance of digital information.

Keywords
Technology heirlooms, digital inheritance, identity

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Experimentation, design

The Spectrum of Digital Information

In a 2008 study, Zhao, Grasmuck, and Martin described that users whose online profiles are tied to anchored, or offline, relationships are more likely to develop digital identities that represent an idealized version of themselves [6]. This should come as no surprise; social networks offer a plethora of opportunities for us to curate the most presentable or self-actualized aspects of our identities. Even with content contributed by other users, we are able to exert some degree of control over these third-party additions. Who amongst us has not un-tagged an unflattering photo or furtively omitted some popular volumes from their favorite books list? In this way, the version of ourselves that we present to others is what we want it to be, if not how we want others to perceive it.

However, these public identities only represent part of what we share online. There is also data that we unintentionally archive, or that is being archived without our knowledge on our behalf. For example, our browsing history, receipts from online purchases, and records of long since completed conversations are often stored without our having made an active decision to do so. This passively collected data paints an equally compelling portrait as our curated identities do – it is a telling account of our day-to-day activities. It is also a depiction of less-public aspects of our digital lives. In contrast to status updates or tweets that are broadcast to legions of readers, this data is distinctly compelling because it is not intended to be shared with a large audience.

In addition to the data that we curate, and the data that is captured about us, there is a third category: the data we hide or discard. Certainly, this data is no less valuable - what we withhold or throw away in life may have great significance in death (or perhaps in the years in between). Anonymous accounts allow us to separate our anchored digital identities from less idealized or polished representations of ourselves [5]. As such, they are valuable vehicles for self-expression and may reveal more than we allow ourselves to show through our public personas. However, the ulterior nature of these identities makes it less likely that the information they contain will be passed down or shared with future generations. In this way, a user's digital legacy may be an incomplete picture of their experiences.

Of course, not all of the information we hide or abandon online is associated with secret aspects of our identities. There are also occasions in which users simply stop using or producing content for a particular web service. Estimates vary, but it is clear that the vast majority of blogs are eventually abandoned [3]. Though these blogs are no longer active, the information posted to them, and other accounts that have been abandoned by users, may still be of interest to future generations.

Revelado

Revelado, a web application we are developing, is centered on this third type of data. Revelado is designed to encourage visitors to share hidden or discarded aspects of their lives, digital or otherwise, such that this information might be transferred to and explored by future generations. The goal for this application is that it will allow people to save, and

someday share, data that they might not have otherwise passed down for any number of reasons.

Visitors to the site are free to anonymously upload information that will remain sealed until a selected date or event in the future. Users are also asked, but not required, to provide contextual data about what they are storing. For each piece of information provided, Revelado generates a key that will allow the recipients of that key to access the information once it is no longer sealed.

Revelado is currently in development and we are exploring the ways in which these keys will be transferred to their recipients. One major component of this work is to determine the digital and physical forms this key could take, and to develop an understanding of how they might fit into existing practices related to death, dying, and the ownership of goods. Traditionally, physical objects play a central role in these practices, but are subject to being lost, broken, or forgotten over time. Digital objects are similarly susceptible to being forgotten, and are plagued by issues regarding future compatibility. Just as a physical object can be broken, it is impossible to ensure the availability of digital data given the rapid cycles in which new technologies are introduced and others lose their foothold in the digital realm.

Outside of practical concerns regarding accessibility, there are also considerations related to the contextual significance of the objects or keys generated by Revelado. In a 2008 study, Petrelli, Whittaker, and Brockmeier explored how objects become mementos imbued with memory and meaning [2]. They found that the habits surrounding the selection and use of mementos varied greatly from person to person. Generally speaking, people chose objects that represented important events and relationships, and that allowed them to remember personal experiences. The significance of these objects and the ways in which they acquired that significance illustrates the difficulty of creating personally relevant and valued mementos that represent digital data. In a related paper, Odom, Zimmerman, and Forlizzi explored the curation and value of virtual possessions among teenagers [1]. In doing so, they provide guidelines for the development of personally meaningful representations of digital data and raise questions about how other user populations view virtual possessions, such as people approaching the end of their lives.

In addition to exploring the manifestation and form of the keys generated by Revelado, its development brings to light issues related to the ownership of digital data. The laws and policies governing digital ownership and access to information are ambiguous [4]. As a result, even in our current context, users may not have clear expectations about who owns the data they have produced and who can access that data. Furthermore, it is unclear how changing policies might affect future recipients of digital data. These issues are of great importance as they will impact the viability of creating systems with multi-generational life spans.

Future Work

Clearly there are a number of open questions regarding the design of systems related to both death and the curation of digital identities. Through the development of Revelado, we hope to explore these issues and to learn about how users view the incorporation of this

hidden or discarded data as a part of their digital legacy.

Acknowledgements

We would like to thank the members of the Creative Systems Lab for their contributions to this work. This work was funded by NSF grant #0750271.

Citations

[1] Odom, W., Zimmerman, J., & Forlizzi, J. (2011). Teenagers and their virtual possessions: Design opportunities and issues. *CHI 2011.*

[2] Petrelli, D., Whittaker, S., & Brockmeier, J. (2008). AutoTopography: What can physical mementos tell us about digital memories? *CHI 2008.*

[3] Quenqua, D. (2009, June 5th). Blogs falling in an empty forest. *The New York Times.* Retrieved from: http://www.nytimes.com/2009/06/07/fashion/07blogs.html

[4] Semitsu, J.P. (2011). From Facebook to the mug shot: How the dearth of social networking privacy rights revolutionized online government surveillance.

[5] Turkle, S. (1997). Life on the Screen: Identity in the Age of the Internet. Simon & Schuster.

[6] Zhao, S., Grasmuck, S., & Martin (2008). Identity construction on Facebook: Digital empowerment in anchored relationships. *Computers in Human Behavior, 24,* 1816-1836.