Lost in Translation: Three challenges for the collection and use of data in personal informatics

Jed R. Brubaker
Department of Informatics
School of Information and Computer Sciences
University of California, Irvine
jed.brubaker@uci.edu

Sen H. Hirano
Department of Informatics
School of Information and Computer Sciences
University of California, Irvine
jed.brubaker@uci.edu

Gillian R. Hayes
Department of Informatics
School of Information and Computer Sciences
University of California, Irvine
jed.brubaker@uci.edu

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
ACM 978-1-4503-0268-5/11/05.

Abstract
As personal informatics tools grow in number, interpretation and interoperability of data will become increasingly important. In this paper, we share insights from our development of Estrellita, a medical tracking system for high-risk infants, and our ongoing research in the area of social media. We highlight three challenges relevant to the area of personal informatics: determining context, assumptions around "personal" use, and the social implications of representation.

Keywords
Personal informatics, data collection, visualization, medical informatics, social media, representation

ACM Classification Keywords
H.4 Information Systems Applications; H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous. K.6 Management of Computing and Information Systems

General Terms
Design, Human Factors

Introduction and Background

The domain of personal informatics (PI) is expanding. PI is increasingly networked and embedded, sourcing data from 3rd parties and appearing as features within larger systems. We have experienced benefits and challenges associated with this expansion during our research on and development of medical and social media applications. In this paper, we describe Estrellita, a medical system that includes PI features inside of a larger, multi-purposed, application [4]. We then outline challenges raised during development, informed by our theoretical work in the area of social media [2].

Estrellita supports parents in collecting data about their prematurely-born infants, sharing these data with interested stakeholders, and incorporating the collection of these data into parent practices and the clinical workflows. The system includes two interfaces: an Android-based mobile application (primarily used by parents) and a web-based interface (primarily used by clinicians).

The mobile application allows parents to collect pertinent information about their baby's health. Additionally, the system prompts parents to report on their mood in order to detect signs of depression. Collected information is synchronized to a centralized server allowing medical providers to review patient data and to ensure that both the caregiver and infant are healthy. Parents can also review reports on their own behavior and mood as well as metrics about their baby's progress (*e.g.,* weight gain) and share this information via social networks like Facebook. Finally, these data are reviewed with parents during consultations with medical professionals.

The development of Estrellita raised questions about how best to implement PI features given this broader ecology. Informed by ongoing research in social media focusing on representational practices within computing, we articulate three challenges from our work that are relevant to the larger community of PI researchers. They include ambiguities around ***context***, challenges to assumptions of ***personal*** use, and a consideration of ***representation***. We detail each below.

The Challenge of Context

PI is inhabited by a plethora of topic-specific but frequently disaggregated tools. These disparate efforts can result in data silos that may adequately capture the user in specific contexts but often lack extensive portability and interoperability. When datasets are made openly available (as is common with social media) they often lack sufficient information with which to qualify and interpret data. Straightforward services with standardized APIs, such as those provided by Twitter and Facebook, have been a boon to PI and visualization tools more generally. Yet, while it is easy to think of ways in which data from services like Twitter might be meaningfully analyzed and presented to end users, researchers and designers are left with the burden of interpreting the nature of these systems, the meaning embedded into the structure of their data, and the behavior of users. As Bowker writes, "*data comes in a dizzying set of categorical bins. You always have to know some context*" [1:116]**.**

Interpreting these systems is further complicated by the deployment of PI features within larger systems and the adoption of NoSQL and non-relational storage solutions. One requirement of Estrellita was that our

data be stored in a secure a centralized Personal Health Records (PHR) system through which data could be accessed by doctors and other client software. One might imagine interesting ways in which data from multiple 3rd-party clients might be combined to provide a more holistic view of the patient's life. However, in the absence of data standards, finding these data, let alone trying to understand their structure and meaning, is challenging.

Even simple visualizations or recommendation engines can obfuscate the nature of the data that underlie their behavior. For example, users may understand how to qualify the term "Friend" within the context of Facebook, but systems that source and present Facebook relationships may overlook or omit important limitations of the data (this can include systems designed by Facebook themselves; for example, see [2]). Thus, when researching PI, we must be cognizant of the means by which these systems were developed, the conditions under which they are used, and the resultant meaning of the data they capture and represent.

The Challenge of "Personal"

PI tools are not used in social or technological isolation. While typically designed around a single user, PI tools involve a host of other social and technological actors. For example, exercise applications like Nike Fit or the FitBit allow users to broadcast their activity via social networks, and one can imagine how the use of a calorie counter can involve a significant other when making dinner plans.

The existence of multiple actors was an important explicit consideration during the development of Estrellita. A diverse set of people care for premature infants, including parents, grandparents, friends, social workers, pediatricians, and specialists. Designing a comprehensive record-keeping system that would meet all of the often-conflicting needs of all of these individuals would be impossible in practice.

Estrellita attempts to balance these various needs, in part, through the inclusion of PI features and techniques. In addition to capturing important medical information, parents can collect information that they feel is relevant and share it with health providers, close relatives, and friends. Although we anticipate parents will appreciate these features, inside Estrellita they raise questions about data collection and use. Pediatricians, for example, are likely to prioritize certain data regardless of the desires of the parent. This scenario is further complicated by the fact that for many of our features, the "personal" is already once removed – the information collected by the parent is about their child.

Estrellita may be an extreme example, but in our research it served to remind us about the ways in which technology is always situated in a larger network of use that extends beyond a single user. When designing PI tools, designers should consider the needs of not just the user of the system, but those surrounding the user as well.

The Challenge of Representation

A central premise in PI is that once data are collected, information can be represented back to the user in ways that enable the process of reflection [3]. However, this requires that the behavior and practices that PI tools seek to represent first be made amenable

to computation. Often, PI approaches this requirement through quantitative indexing: calories are counted, dollars are added, and time is tracked. This process, however, necessarily reduces some types information and eliminates others all together. Here we consider the nature of data representation within PI.

In recent work we have stressed the importance of computation as a representational practice through which socio-technical practices emerge [2]. Technological systems dictate what behavior can be digitized, and often influence that very behavior through the process of capturing it. This presents both a challenge and an opportunity for PI tools. As those familiar with PI are aware, the adoption of a specific tool can go hand-in-hand with the adoption of the behavior it is meant to support. The adoption of a fitness-oriented PI tool, for example, may be the byproduct of a new gym routine (or visa versa). This is certainly the case with Estrellita. Most parents will have never tracked baby-related information to the extent that having a premature infant requires. Likewise, we hope to encourage certain behavior (*e.g.,* reading to the infants) through acts of tracking to improve long-term outcomes for the infants.

However, we must remember that the captured data is a biased representation of the reality on which it is based. After all, "*databases are not filled with the actual people, places, and activities they describe but rather with representations of them*" [2]. The translations that occur as information moves from real-world experience to the database is compounded as users modify their data capture practices, not only in relationship to changes in their world, but to the world presented back to them through the PI system as well.

Conclusion

We have outlined three challenges to the design and study of personal informatics. As tools to collect and reflect on personal data increase, their use becomes increasingly situated in larger social and technological networks. As such, issues regarding the context of data, delineating what is "personal," and the meanings of various representations produced by PI tools will continue to be important for both researchers and designers.

Acknowledgements

We would like to thank Karen Cheng, Leslie Liu, Monica Tentori, Dini Baker, Marni Nagel, and Sheba George for their work on the design of Estellita, the clinicians and hospital staff who collaborated with us on this project, and members of STAR Group and LUCI. This work was supported in part by the Robert Wood Johnson Foundation's Project HealthDesign, Google, and NSF CAREER grant #0846063.

References

[1] Bowker, G. (2005). Memory practices in the sciences. MIT Press Cambridge, MA.

[2] Brubaker, J. R. and Hayes, G. R. (2011, to appear). SELECT * FROM USER: Infrastructure and Socio-technical Representation. Proc. CSCW 2011. Hangzhou, China.

[3] Li, I., Dey, A., and Forlizzi, J. (2010). A stage-based model of personal informatics systems. Proc. CHI 2010. Atlanta, GA.

[4] Liu, L., Hirano, S., Tentori, M., Cheng, K., Park, S. and Hayes, G. R. (2011, to appear). Improving Communication and Social Support for Caregivers of High-Risk Infants through Mobile Technologies. Proc CSCW 2011. Hangzhou, China.