

From the Small to the Large: Learnings from the deployment of Serendipitous Family Stories / StoryPlace.me

Frank Bentley

Motorola Mobility Research
Libertyville, IL USA
f.bentley@motorola.com

Santosh Basapur

Motorola Mobility Research
Libertyville, IL USA
sbasapur@motorola.com

William Hamilton

Texas A&M University
College Station, TX USA
bill@ecologylab.net

ABSTRACT

Research in the large is quite different from the more traditional small-scale field studies conducted in Ubicomp over the past decade. Large scale studies are not just small scale ones with a larger n , they allow researchers to answer fundamentally different questions about use and adoption and enable the studying of systems in real and messy social networks and situations. However, conducting large-scale research often requires large-scale resources. We present the evolution of the Serendipitous Family Stories system into StoryPlace.me and discuss how large-scale research required 5x more time and effort to prepare a system for a field trial.

Author Keywords field studies, large-scale research, deployments, scale, security.

ACM Classification Keywords H5.2 User Interfaces: Evaluations and Methodology.

General Terms Human Factors

INTRODUCTION

In recent years, researchers in the Ubicomp and CHI communities have been moving towards larger-scale public deployments of systems as ways to better understand system use “in the wild” outside of the constraints and influences of small-scale studies [e.g. 5-10]. The growth of mobile application stores and the adoption of popular smartphone platforms such as Android and iPhone have made it relatively easy to reach a large audience with new services and applications.

However, performing large-scale research is not as easy as releasing a typical research app into a market. Significant additional work on design, development, security, and scalability is often required at significant expense. Planning a study also requires a great deal of thought ahead of time in terms of what data needs to be logged and what questions can be answered. While in a small-scale study, semi-structured interviewing of participants allows for easy

modification of research protocols to match the data that is being received, changing what is logged in a large-scale study requires thousands of users updating your application.

We will discuss these issues in the context of the Serendipitous Family Stories / StoryPlace.me system, developed in Motorola Mobility Research in 2010-2011. We will begin with a description of the system and its development.

SERENDIPITOUS FAMILY STORIES

The Serendipitous Family Stories (SFS) system [2] was first built in the summer of 2010 to support a small-scale research experiment that fall. The idea for the concept came from a generative research project focused on communication between generations across distance [3]. The system consisted of a web interface that older adults could use to create video stories for their younger relatives about their lives growing up in Chicago. These stories were placed on a Google Map based on the locations where the original events took place. The creators’ younger relatives used an Android application over a four-week period to receive the stories. Story content remained hidden until the recipient happened upon the place where the story occurred at which point their phone would vibrate and they could “unlock” the story and play the corresponding video in the context of where it had originally occurred.

As this study would only involve 20 users and we would be meeting all of them in person to set up their devices, rapid prototyping methods were used to quickly create a system that would be “good enough” for the study [4] without taking too long to implement. Many features were left out of the system, for example the ability to manage contacts in the system or receive notifications when content was discovered or shared. Contacts were set up by researchers at the start of the study, and participants only had their one contact of their older/younger relative. In total the design, implementation, and testing of the system took less than 12 weeks.

We had many research questions that we wanted to answer with the SFS system. Some of these questions were best answered by a small-scale, mostly-qualitative study with a group of local participants, but others required a large-scale approach.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Large '11, September 18, 2011, Beijing, China.

Copyright 2011 ACM 978-1-4503-0924-0/11/09...\$10.00.

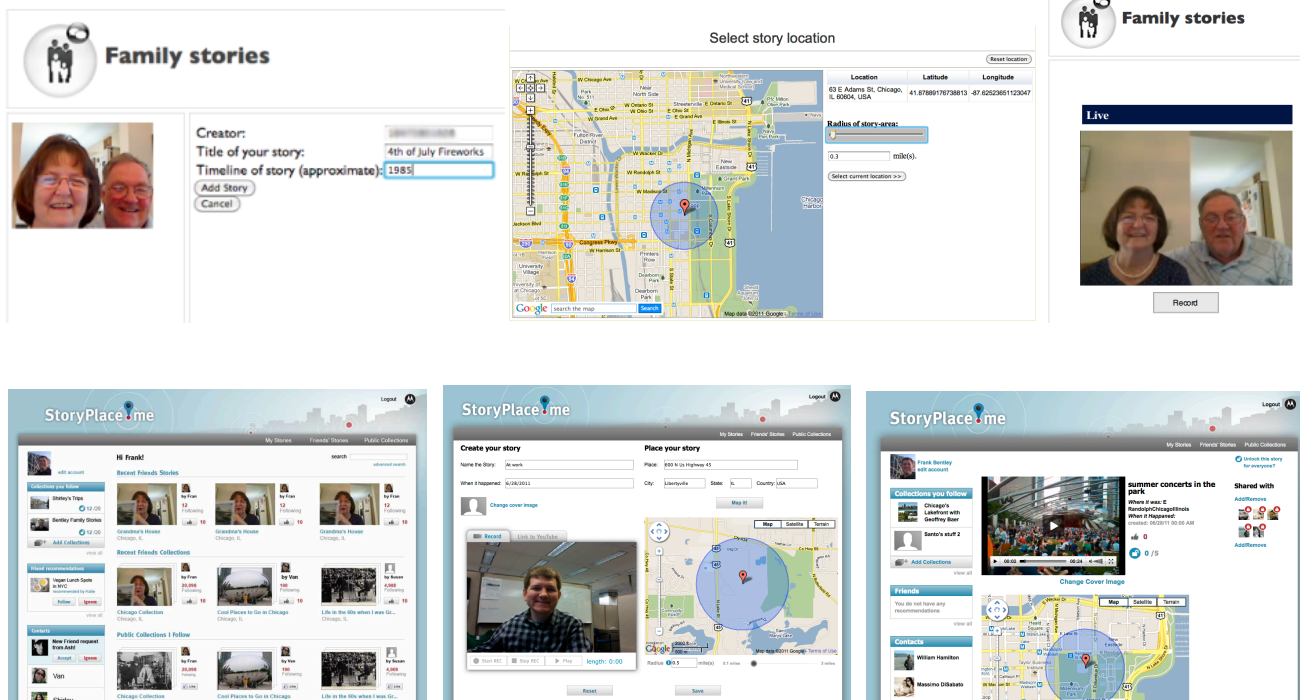


Figure 1: The web interface of Serendipitous Family Stories vs. the interface for StoryPlace.me. The public deployment needed to be much more polished and support a wider feature set to be accepted by a large audience.

Questions around how receiving content changed a relationship between an older adult and their adult child required in-depth interviews and voicemail logs throughout the study for us to fully understand their use. Likewise, understanding the experience of uncovering a story and any tensions that arise over the circumstances of finding it required deeper interaction with local participants.

WHY LARGE?

After our field trial, we realized that there were many other use cases for the concept of location-based video communication. Because our initial study was so focused, we were unable to see these uses in practice. In order to better understand these other emergent uses and how the system might be used by other populations, we decided to build a large-scale public beta. This system involved a much higher fidelity implementation that could meet the needs and expectations of many thousands of users worldwide without any handholding or setup from the research team. This system involved over one person-year of work on design, implementation, and testing.

One of our larger research questions was about the types of content that would be created in a system like this. As the reviewers of our Ubicomp paper on the first system pointed out, a small-scale study like our initial work cannot fully answer these types of questions. While we have gained some good initial ideas of the types of content that users create when sharing location-based asynchronous video, we

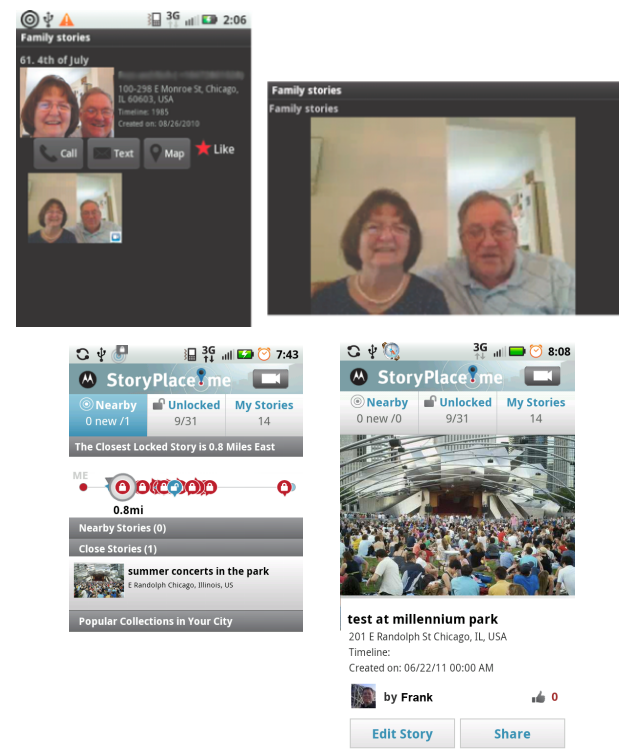


Figure 2: The mobile interfaces of the two systems.

did not have much certainty until we build a larger deployment and users were free to create content over many weeks and months.

Our other research questions centered on how location-based content is shared in propagated in social networks. Since it was quite difficult to recruit sets of parents/grandparents and adult children for the first study, we knew it would be nearly impossible to recruit larger social networks. Deploying a system publicly allows users to recruit their own friends in a natural way. In the large deployment we will be able to see how content spreads between friends and also learn more about the types of content that are created. Are most stories created as personal messages for one particular person? Or are they created for larger audiences? Can the location of a story be a predictor of how it is shared or how many people discover it? What types of content do people view when they are traveling versus in their home city? We believe that these questions can only be adequately answered with a larger deployment.

However, not all large-scale deployments are able to collect this sort of data and attract social groups to use. Morrison et al found it quite hard to get social groups using the World Cup Predictor app together in their large deployment [9]. If this is the key research question to be answered, it might be a bit of a gamble to spend the time and money to launch an app in the large, only to discover the lack of use of a key feature that is being studied.

THE COSTS OF LARGE

Since there are some research questions that can only be answered by large-scale deployment, a research team must carefully weigh the need and ability to answer these questions vs. the much larger costs and time required.

As discussed by [10], applications released in mobile App Stores and Marketplaces need to appear more polished than the typical research prototype. This often means hiring a graphic/interaction designer or at least using the time of one. This more polished design also likely complicates the implementation of the system adding extra weeks or months to the development cycle. As shown in Figures 1 and 2, our StoryPlace.me website and mobile app are professionally designed and significantly more complex than the original SFS system. The website makes frequent use of AJAX and Javascript and the mobile app contains complex animated widgets and much greater functionality than the first system.

Functionally, applications need to be robust as well. In a small-scale trial where researchers meet with the participants, it's usually acceptable to have an application that might crash at times or not have the most perfect interaction. Often, research questions in the small center on how a system is used and participants in small-scale trials are usually understanding of the fact that the system is still in development. However, apps in a public Marketplace

come with an understanding that they are complete and polished requiring a great deal more testing and bug fixing before launch.

Development for a large-scale deployment also needs to be more robust. In a small-scale trial that is not known to the public, security and scalability are not as big of a concern. The initial SFS prototype used the device phone number as the identifier for the user and did not provide any encryption of requests to the server. A public-facing system needs to be much more robust and we adopted an OAuth-style interface for making calls to the server as well as increased the security on the web interface for StoryPlace.me.

Releasing a system publicly also involves a good deal of additional fact checking and code review. Depending on the license, open-source components that are extended may have to be re-released to the community. Proper searches must be conducted to ensure that any names in use in the system are not trademarked in every country where the app will be released. Local privacy and data reporting laws must be taken into consideration in presenting terms of use and opt-in/opt-out policies. Etc. In small-scale research, these are not concerns since software is not publicly distributed and research overviews and consent forms cover the privacy and use of data. Most research teams are well equipped to create these forms, but may not have the legal expertise required to release a system globally.

We also had to consider scale. While the original system could likely support many thousands of users, releasing a system publicly can create large spikes in use, especially if the service is mentioned on a major blog or retweeted by a user/company with a large number of followers. For us, this meant moving to a more scalable cloud-computing model with a load balancing system and the ability to quickly increase capacity. This added additional time to the development as well as cost in running the server instances compared to the small server in a closet for SFS.

This scale also results in a large amount of data being returned from the public trial participants. This amount of data can be overwhelming for researchers used to small-scale mostly qualitative research. Tools like Morrison and Chalmers SGVis [8] or complex queries to databases or analysis clusters become necessary and further complicate the analysis. Finding team members with the appropriate quantitative skills can also be challenging.

Overall, preparing for the large deployment was close to 5x the cost in terms of design/development time.

COMBINING LARGE AND SMALL

Large-scale research can bring about many benefits and new types of understanding, especially around adoption of social systems or systems where content needs to be created by a large number of people. These are research questions that are not easily answered by smaller scale field trials.

Researchers may also observe emergent behavior in large-scale deployments that leads to further research questions.

When deciding to go small or large, the resources required for either path are often a deciding factor. However, often the best answer is to do both. By building an initial prototype, we were able to learn quite a bit about use in one very particular context. We learned about the process of finding stories in the real world, and the types of content that participants wanted to create. These findings gave us the knowledge we needed in order to build a larger system that could be successful in a public trial. We also learned small technical details that would allow for a more successful public deployment such as additional parameters that we should be logging in order to better understand the flow of application use. The public trial then provided us with additional data about adoption and spread of media within social networks that could help us to further improve the system and better understand mobile content creation and the use of asynchronous mobile video communications.

Just performing a usability analysis or other discount methods on the early prototype would not have unveiled these deeper uses. Without the small-scale study in-between the concept and the larger trial, the resulting system would not have been as interesting and would likely have only focused on personal content instead of the larger opportunity we uncovered in push location-based media.

We believe that small should go before large, but also that additional small-scale studies are useful once a large deployment is released. Studies like Ames and Naaman's ZoneTag [1] work can help researchers understand exactly why they are seeing the results that they are in the quantitative data from a large deployment. But these studies do not make the earlier, rapid prototype and small-scale field evaluation irrelevant. These early studies are a critical part of the iterative design of a successful large-scale deployment built on findings from previous evaluation.

DISCUSSION

Small-scale and large-scale research are complementary in terms of the types of questions that are best answered with each style. Small-scale research however is much easier to conduct in terms of implementation effort and study execution. As researchers, we need to carefully consider the answers we can obtain from large-scale deployments against the effort required to properly deploy one. While some researchers have been successful in deploying relatively simple applications publicly (e.g. [4,9]), deploying more complex systems to many thousands of users is quite a different level of effort.

As we begin to receive data from our deployment, we will learn about key differences in use between small and large-scale deployments and hope to have preliminary findings to present by the time of the workshop.

REFERENCES

1. Morgan Ames and Mor Naaman. 2007. Why we tag: motivations for annotation in mobile and online media. In *Proceedings of CHI 2007*.
2. Frank Bentley, Santosh Basapur, and Sujoy Kumar Chowdhury. 2011. Promoting Intergenerational Communication Through Location-Based Asynchronous Communication. In *Proceedings of Ubicomp 2011*.
3. Frank Bentley, Gunnar Harboe, and Pallavi Kaushik. 2010. Involving Seniors in Ethnographic-style Work: Initial findings from a study on long-distance communication. *CHI 2010 Workshop on Senior-Friendly Technologies: Interaction Design for the Elderly*.
4. Frank Bentley and Crysta Metcalf. Rapid Prototyping and Field Evaluation of Mobile Experiences. *CHI 2009 Workshop on Mobile User Experience Research: Challenges, Methods & Tools*. April, 2009.
5. Sebastian Büttner, Henriette Cramer, Mattias Rost, Nicolas Belloni, and Lars Erik Erik Holmquist. 2010. $\phi 2$: exploring physical check-ins for location-based services. In *Proceedings of the 12th ACM international conference adjunct papers on Ubiquitous computing (Ubicomp '10)*. ACM, New York, NY, USA, 395-396.
6. Niels Henze, Martin Pielot, Benjamin Poppinga, Torben Schinke, and Susanne Boll. My App is an Experiment: Experience from User Studies in Mobile App Stores. 2011. *International Journal of Mobile Human Computer Interaction (IJMHCI) – special issue on Research in the Large*.
7. Emiliano Miluzzo, Nicholas D. Lane, Shane B. Eisenman, and Andrew T. Campbell. 2007. CenceMe: injecting sensing presence into social networking applications. In *Proceedings of the 2nd European conference on Smart sensing and context (EuroSSC'07)*, Gerd Kortuem, Joe Finney, Rodger Lea, and Vasughi Sundramoorthy (Eds.). Springer-Verlag, Berlin, Heidelberg, 1-28.
8. Alistair Morrison and Matthew Chalmers. SGVis: Analysis of Data From Mass Participation Ubicomp Trials. 2011. *International Journal of Mobile Human Computer Interaction (IJMHCI) – special issue on Research in the Large*.
9. Alistair Morrison, Stuart Reeves, Donald McMillan, Matthew Chalmers. 2010. Experiences of Mass Participation in Ubicomp Research. *Ubicomp 2010 workshop on Research in the Large*.
10. Robert Schleicher, Alireza Sahami Shirazi, Michael Rohs, Sven Kratz, Albrecht Schmidt. WorldCupinion: Experiences with an Android App for Real-Time Opinion Sharing during Soccer World Cup Games. 2011. *International Journal of Mobile Human Computer Interaction (IJMHCI) – special issue on Research in the Large*.