Data-Driven Persona Development

Jennifer (Jen) McGinn
Sun Microsystems, Inc.
35 Network Drive, Burlington, MA 01803
Jennifer.McGinn@sun.com

Nalini Kotamraju
Sun Microsystems, Inc.
17 Network Circle, Menlo Park CA 94025
Nalini.Kotamraju@sun.com

ABSTRACT
Much has been written on creating personas – both what they are good for, and how to create them. A common problem with personas is that they are not based on first-hand customer data, and if they are, the data set is not of a sample size that can be considered statistically significant. In this paper, we describe a new method for creating and validating personas, based on the statistical analysis of data, which is fast and cost effective.

Author Keywords
Personas, persona development, data gathering, data analysis, factor analysis, user research, survey, interview.

ACM Classification Keywords
H5.2. Information interfaces and presentation (e.g., HCI): User-centered Design, Theory and Methods.

INTRODUCTION
The term "personas" was first introduced by Cooper [5] to describe a new way of creating user profiles. A persona is a precise description of a hypothetical user and what s/he wishes to accomplish [3]. Personas "provide a conduit for conveying a broad range of qualitative and quantitative data, and focus attention on aspects of design and use that other methods do not" [15].

Some [2] make crisp distinctions between user roles, user segments, extreme characters, and personas. However, we take the broader perspective of Mulder and Yaar [12] and Pruitt and Adlin [14], that there are many types of personas, which are differentiated by the methods used to create them. We also agree with Murata [14, p. 228] that "personas are about norms, not about exceptions".

Much has been written on creating personas – both what they are good for [1, 5, 7, 12, 14, 15, 16], and how to create them [1, 3, 4, 7, 12, 14, 15, 16]. A common problem with creating personas is that they are not based on first-hand

customer data [15], and if they are, the data set is not of a sample size that can be considered statistically significant.

COMMON PERSONA DEVELOPMENT PROBLEMS
Four problems commonly occur in persona development efforts [3, 15]. Of those, only one problem has to do with the development of the personas themselves[1]: "The characters were not believable; either they were obviously designed by committee (not based on data) or the relationship to the data was not clear" [15].

The lack of connection between the data and the personas may have been a result of Pruitt and Grudin's [15] inability to conduct original research as part of their persona development process. They cite the lack of time and money as the reasons for using existing research, but their persona development initiatives took much longer than expected, and they did not meet their goal to "have every statement in our Personas generated from or related to user data or observation."

By contrast, our method was relatively fast and cheap, easy to scope, and all persona data was directly derived from the research that we performed.

METHOD
Our method effectively turns the traditional persona research, development [1, 6, 7, 9, 11, 12, 14, 16], and validation process [1, 6, 9, 14, 16] on its head, by changing the order of activities [8, 16]. When data gathering is done to create personas, it often starts with collecting existing data, performing field studies, drafting the personas, and then, if there is time and budget left, validating the persona groups using a survey.

We did not develop the personas for our own use, but were contracted to develop them by the training group inside our company. We *could* have created personas that had ages, certain computer proficiencies, and other demographic attributes, but they most likely would not have answered the specific questions of our client team.

What we did, instead, was to involve our client team early in the process, to define the persona attributes that were

---

[1] The other three problems are: the personas are not communicated well; how to use the personas is not well understood; and persona development efforts lack the support of senior management. [14]

most meaningful to the team, and to articulate the questions that the team needed to have answered, which we discuss in the next section.

We then created a survey, based on that stakeholder input, which was sent to people who were identified as representative of our target market. We received over 1300 responses to the survey, and from those responses, performed factor analysis to allow the persona groups to emerge. Next, we interviewed people from each group, and used that information to make adjustments to the groupings.

Stakeholder Involvement in the Survey Design

The internal client team that sponsored this work was heavily involved in the survey design process. We told the team that there were several attributes that were commonly specified in personas, such as name, age, education level, and years of experience. We could certainly gather that data, but we wanted the personas to be of the highest value possible to the team – to ensure that the personas that we delivered would be used by the client.

So we asked: what are the problems that you are trying to solve? What don't you know about your customers that you wish you did? What are the biggest challenges facing you in the coming fiscal year, and what data would help you face those challenges? We spent about three weeks iteratively designing the survey questions and having the questions reviewed by different parts of the organization. As a result, the persona data is salient to the client team because it answers questions that are specific to its marketing and development domains.

For example, our personas were developed for a training organization, so the attributes that they were interested in learning more about were related to training experiences and professional certification. These questions included how much training had customers taken over the last two years, the amount spent on that training, who made the decision to take training, what professional certifications customers held, whether and why they valued those certifications, and so on.

The Persona Survey

A third-party vendor whom we use regularly conducted the survey. We selected potential respondents from an internal database of people who were already working with our company, and they were invited to participate in the survey, by going to a specific URL.

The invitation and survey were translated from English into simplified Chinese, Japanese, and neutral Spanish (those languages were identified by the client organization). We already knew which country each invitee was from, so their country of origin determined in which language version of the invitation and survey they received.

In all, there were 18 multiple-choice survey questions that gathered both demographic and behavioral data. Because we were using an internal database of participants, there was some data that we didn't need to collect, because we already had it: name, company, country, address, phone number, and what product they were using.

That said, we *did* ask demographic questions, such as their highest level of education, the business domain of their company, the number of employees in their company and how long they'd been doing their current kind of work. In addition, we asked about the hardware and software platforms they currently used.

The domain-specific questions focused on their experiences and attitudes regarding technical training and professional certification. Such as, what inhibited them from taking training, and what motivated them to take training? We asked who was responsible for making decisions about how much training they received and which classes they took; how much training they'd taken over the last few years and how much they or their company had spent on it; which certifications they currently held, and why they held them. Lastly, we asked about the types of training they took, and what was important when selecting a course vendor.

Goodwin [7] tells us to use patterns of behavior to drive the persona development. So, as we discuss in the next section, rather than ask people from around the world to identify job titles that described what they did, we instead asked them to identify their job tasks and then to rank order those tasks.

We did not ask about age, marital status, hobbies, or for any other data that was not directly related to our business domain. Our final question allowed respondents to opt in to a 60- to 90-minute telephone interview, in return for $50.

In all, we received over 1300 survey responses from more than 90 countries. We then used exploratory factor analysis to group the responses into personas.

The Quantitative Analysis

Factor analysis, a data reduction technique, represents a large set of variables in terms of a smaller number of new variables [10]. These newly created variables underlie the original, larger number of variables. We chose exploratory factor analysis for two reasons. First, we wanted to use a technique that would generate groupings of people's work tasks, rather than use a technique that required a priori groups. Factor analysis allowed us to see how the responses formed groups of tasks; these factor-driven groups – and resulting personas – were an artifact of the data.
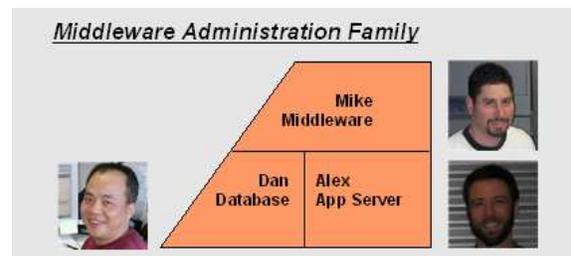

Figure 1: Persona family example

We created the initial groups by factor analyzing[2] responses to questions about which of twenty-five tasks people performed regularly as part of their jobs. Some of these tasks were: "I evaluate software"; "I perform database administration"; "I manage people"; and "I am a full-time student or educator". Our analysis yielded an initial set of factors. Each of the factors was a group that included only the task measures that loaded most highly on it, meaning that each of the twenty-five work tasks was associated with only one factor group.

We then placed people into groups based on their responses to another series of questions about which of tasks were the most time-consuming (respondents were only given the option to rank order the tasks that they had selected in the previous question; not all 25 tasks). For example, people who indicated that they spent the most time performing database administration were placed into the factor group that included the measure for performing database administration.

The second strength of factor analysis is that it allowed us to easily see the relationships between the factors or groups. We then factored the original factors to make visible the relationship between the original factors. While each of the original factors remained as persona groups, we were also able to talk about the factors of factors as "families" of personas. This analysis allowed us to identify relevant patterns of work task behavior in the survey responses, and these patterns became the basis of the personas.

The Qualitative Data and Analysis
We knew that the survey results would give us a lot of data on *what* people did. The interviews were designed to tell us *why*, by asking probing, open-ended questions that could not easily be analyzed by statistical methods.

Once we had our initial groups, we identified people who had agreed to be interviewed from each of the groups, and invited them to talk over the phone. The interviews took 30 to 90 minutes, each, and 26 interviewees answered a set of 11 open-ended questions.

All of the interviews were conducted with participants in the United States. We believed that this would not introduce any significant bias because they were already grouped, by tasks performed, with people around the world. Pragmatically, using participants in the U.S. was the only way that we could ensure compliance with international laws, and we could send the incentives in a single currency.

---

[2] We used Stata 9.0 for Macintosh OS 10 to analyze the responses using the principal components factor method, and the factor structure was analyzed using Varimax rotation. For a discussion of best practices of factor analysis, see Russell, D. W. 2002 "In Search of Underlying Dimensions: The Use (and Abuse) of Factor Analysis in Personality and Social Psychology Bulletin." *Personality and Social Psychology Bulletin* 28(12): 1629-1646.

Revising the Persona Groups after the Interviews
Initially, we did not think of the interviews as being a validation technique [15, 16] for the survey design, but that is exactly what they became. After conducting the interviews, we sometimes saw a strong alignment with the group they were in (proving only, perhaps, that statistics works), and in a few cases, a very weak alignment.

We interpreted the weak alignment to imply that the question set, which we had used to create our factors, was imperfect. We hadn't used demographics [4, 12, 14, 16] to define our personas, so our factors were not implicitly validated. Instead, our groups were defined by time on task, and when we went back to look at the data, some new patterns emerged. For example, the interviews showed that some of the people who responded that they performed database administration as their most time-consuming task were actually full-time database programmers, so they were re-categorized into a developer group.

As a result of the interview findings, we created one new subgroup in the first family, and we removed a group in the second family. In addition, some respondents were moved from a group in the second family to a group in the third family. What the interviews had uncovered were weaknesses in the questions that we asked. Some tasks were important, while others just obfuscated the results.

Four additional interviews were conducted to ensure that we had multiple interviewees in each persona category.

The Results
The survey and interview data were combined into posters: one for each of the eleven personas, and then one that described the four families.

As suggested by Pruitt and Adlin [14], we gave each persona a fictional alliterative name, such as "Manny the Manager", and used photos of employees in their work environments as the faces of our personas. Each poster described how the persona spent his or her day, her or his biggest challenges, goals, demographic data, and data that was specific to the market domain [6].

COMPARISON TO OTHER METHODS
Pruitt and Adlin [14] acknowledge that "Collecting, analyzing, and translating insights generated from data into personas takes time and energy." In all, we spent four people months to: solicit stakeholder input; develop, translate, and run the survey; conduct the interviews; analyze the data; and create the personas.

Compared to creating personas from existing data [9] or no data, our method did take longer, but the data was real, it was gathered specifically for the purpose of creating the personas, and our method allowed for assumptions to be challenged [11]. In addition, the results were easy to defend.

Our method was fast and cheap, particular compared to other recent industry experiences of spending several people years on ethnographic research and persona development (20 to 66 people months for Microsoft [13] and "many months" to "several person years" for Cisco [15]). In addition, we did not have to worry that a single interviewee would skew the results or worry that we talked to the wrong people – we had statistically significant sample sizes to base our personas on, and the interviews provided insight and direct quotes.

CONCLUSION
Our method of creating data-driven personas provides several benefits over traditional methods.

- The personas are based on real customer data, and are, therefore, more readily accepted by the client.
- The data has been gathered specifically for the purpose of creating personas, so what the practitioner needs to know has been collected.
- The internal team is involved all the way through the process.
- The assumptions and beliefs of the internal team (as well as those of the practitioner) can be tested, rather than accepted as fact.
- For a project of this size and scope (having the variety and complexity of the population that it does), the time to complete the data gathering and analysis is people years faster and therefore cheaper, than high-quality ethnographic techniques.
- The data set collected is of a statistically significant quantity.
- By using factor analysis, the number of resulting personas appears from the data organically, rather than imposing a number a priori.
- And lastly, the factor analysis allowed us to see relationships between the personas, as "families", rather than just presenting the personas in a flat organizational structure.

ACKNOWLEDGMENTS

REFERENCES
1. Adlin, Tamara, et al. Panel: Putting Personas to Work. *Proceedings of CHI 2006*, ACM Press (2006), 13 – 16.
2. Aquino, Plinio Thomaz, Jr, and Filgueiras, Lucia Viela Leite. User Modeling with Personas. *Proceedings of CLIHC'05*, ACM Press (2005), 277 – 281.
3. Blomquist, Asa and Arvola, Mattias. Personas in Action:Ethnography in an Interaction Design Team. Proceedings of NordCHI 2002, ACM Press (2002), 197 – 200.
4. Cooper, Alan, Reimann, Robert, and Cronin, David. *About Face 3: The Essentials of Interaction Design*, Wiley (2007).
5. Cooper, Alan. *The Inmates Are Running the Asylum*, Sams (1999).
6. Goodwin, Kim. *Getting from Research to Personas: Harnessing the Power of Data.* http://www.cooper.com/insights/journal_of_design/articles/getting_from_research_to_perso_1.html
7. Goodwin, Kim. *Perfecting Your Personas*. http://www.cooper.com/insights/journal_of_design/articles/perfecting_your_personas_1.html
8. Grudin, Jonathan and Pruitt, John. Personas, Participatory Design and Product Development: An Infrastructure for Engagement. *Proceedings of the Participatory Design Conference 2002*, 144-161.
9. Khalayi, Nina, et al. Persona Based Rapid Usability Kick-Off. *Proceedings of CHI 2007*, ACM Press (2007), 1771 – 1776.
10. Kim, Jae-On & Mueller, Charles W. *Introduction to Factor Analysis: What It Is and How To Do It*. Sage Publications (1978).
11. Markensten, Erik and Artman, Henrik. Procuring a Usable System Using Unemployed Personas. *Proceedings of NordCHI'04*, ACM Press (2004), 13–22.
12. Mulder, Steve with Yaar, Ziv. *The User Is Always Right: A Practical Guide to Creating and Using Personas for the Web*, New Riders (2007).
13. Neiters, James E., Ivaturi, Subbarao, and Ahmed, Iftikhar. Making Personas Memorable. *Proceedings of CHI 2007*, ACM Press (2007), 1817 – 1823.
14. Pruitt, John and Adlin, Tamara. *The Persona Lifecycle: Keeping People in Mind Throughout Product Design*, Morgan Kaufman (2006).
15. Pruitt, John and Grudin, Jonathan. Personas: Practice and Theory. *Proceedings of the 2003 Conference on Designing for User Experiences (DUX'03)*, ACM Press (2003).
16. Sinha, Rashmi. Persona Development for Information Rich Domains. *Proceedings of CHI 2003*, ACM Press (2003), 830 – 831.