Personas Changing Over Time: Analyzing Variations of Data-Driven Personas During a Two-Year Period

Soon-gyo Jung
Hamad Bin Khalifa University
Doha, Qatar
sjung@hbku.edu.qa

Joni Salminen
Hamad Bin Khalifa University
Doha, Qatar
jsalminen@hbku.edu.qa

Bernard J. Jansen
Hamad Bin Khalifa University
Doha, Qatar
jjansen@acm.org

ABSTRACT

One of the critiques of personas is that the underlying data that they are based on may stale, requiring further rounds of data collection. However, we could find no empirical evidence for this criticism. In this research, we collect monthly demographic data over a two-year period for a large online content publisher and generate fifteen personas each month following an identical algorithmic approach. We then compare the sets of personas month-over-month, year-over-year, and over the whole two-year period. Findings show that there is an average 18.7% change in personas monthly, a 23.3% change yearly, and a 47% change over the entire period. Findings support the critique that personas do change over time and also highlight that changes in the underlying data can occur within a relatively short period. The implication is that organizations using personas should employ ongoing data collection to detect possible persona changes.

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

ACM ISBN 978-1-4503-5971-9/19/05. DOI: https://doi.org/10.1145/3290607. 3312955

KEYWORDS

Personas; Web analytics; online data representations

Table 1: Key Constructs and Definitions for Data-Driven Personas.

Persona
An imaginary person created from data that represents a user segment for content, system, or product

Persona Description
The depiction of a persona, normally in 1-2 pages that synthesize information about the persona

Demographics
Statistical data relating to a population or particular groups within it

User Segment
Groups of individuals that are similar in specific ways, such as age, gender, interests or behaviors

Persona Creation
Creation of personas, typically via the use of surveys or focus groups

Data-driven Persona
Creation of personas via the employment of actual user data

1 INTRODUCTION

Personas portray user segments presented as imaginary people, usually within the form of a persona profile containing attributes of the segment that the fictionalized person represents. Personas are integrated into development cycles in many domains [4, 6, 7] for a variety of projects types. Personas are traditionally developed using qualitative methods and data collection approaches, such as surveys and focus groups, normally resulting in a handful of personas shown to the end users.

Personas have been criticized for a variety of reasons, including staling [3], in that the underlying data from which they were created is no longer valid, resulting in expiring persona profiles. This requires further rounds of data collection to update the personas, which is, using the typical manual persona creation methods, time consuming and expensive. In fact, without periodically new data, persona users are uncertain whether the personas are representative of their *current* target users. This limitation is especially acute for those that distribute products online, with potential audiences being large, diverse, and frequently changing behaviors.

However, despite this routinely stated critique, we could locate no specific research investigating if personas actually do become outdated. Similarly, there is a lack of research determining how often additional data collection is needed for persona creation or presenting a methodology to verify if and when personas change.

As such, there are several unanswered questions. *Do personas actually change over time? If so, how often do they change? What is the pace of change? Is the change gradual or rapid? How many personas change? How does one identify when personas change?* These are some of the questions that motivate our research. Naturally, some of the answers to these questions may vary organization to organization; however, we could locate no empirical research showing that personas change in *any* context or environment.

In this research, we collect user data monthly during a two-year period for a large international YouTube content-producing organization with hundreds of thousands of social media followers. We then generate 15 personas from each monthly dataset following an identical algorithmic methodology for data-driven persona creation [2]. We then compare the changes in personas by month, by year, and over the data collection period. Our findings show that personas change at all levels of comparison and that they can change quite rapidly. See Table 1 for key constructs and definitions applicable to this research. We present a brief literature review, research objectives, methodology, and results. We end with implications and promising future research directions.

2 Review of Literature

Stated benefits aside [8], there are still concerns about the value of personas. One of the most noted concerns being that multiple or periodic rounds of data collection may be needed to keep the personas updated [12]. This criticism, widely present in the persona literature [3], is based on the assumption that there are times of instability and change in user populations [5]. This criticism is based on the assumption that there are periods of instability and change in user populations, i.e.,

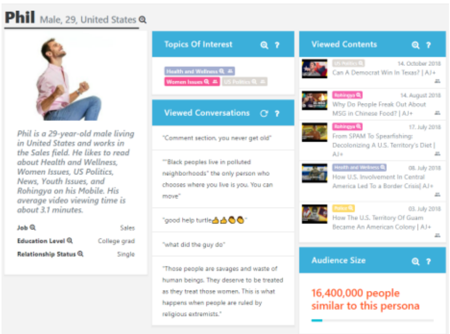

Figure 1: Example persona from the automatic persona generation (APG) system. The persona is generated from YouTube user data.

the data from which the personas are created. However, we could find no prior research that personas do change over time or that one needs to keep personas updated [16, 17], and we could find no prior research documenting an investigation of changes in personas.

This is an important research gap to address, as creating personas is not a cheap, easy, or quick process, given it has historically involved ethnography studies or focus groups [8]. When automated methods are employed [1, 2], there is still effort and time involved in the creation process. Therefore, whether or not additional rounds of data collection are needed has practical implications, notably for those organizations with large user populations, for example, online content publishers whose audiences' preferences can rapidly fluctuate.

3 Research Objectives

Our research goals are twofold. First, we investigate whether or not personas change over time and, if so, how much. Secondly, as specific changes are naturally data dependent, we validate a methodological approach for determining changes in personas over time. In pursuit of these goals, we address three research questions, which are:

- *Do personas change monthly?*
- *Do personas change yearly?*
- *Is there a change in personas during the data collection period?*

4 Methodology

For this research, we investigate these questions using data from a major producer of YouTube content. With the increased availability of online user and customer data, there is the opportunity to use data-driven personas derived directly from a system's users or a company's customer analytics data [15]. In fact, personas developed from this approach can be created automatically [9]. An example of a complete persona from the system is shown in Figure 1.

This organization had more than 580,000 subscribers and thousands of pieces of online content at the time of the study. As such, the organization is representative of those entities that distribute products via major online platforms, such as other content producers, app developers, SaaS providers, etc. The YouTube Analytics API provides statistics for each piece of content and various user profile data, (e.g., gender, age, country location) at an aggregate level. Individual user data is not provided to safeguard users' privacy. Via the YouTube Analytics API, we collect the detailed record of product interactions by country, gender, and age group for each piece of content. We collect data for all videos published each month from October 2016 through September 2018.

In terms of producing the personas, the applied approach for generating data-driven personas is discussed in prior work [9, 10]; therefore, we only briefly present it here. The approach relies on non-negative matrix factorization (NMF) to take the aggregated user data and identify unique behavioral patterns [13], associating these unique sets of behavioral patterns with demographic attributes, and then using other algorithmic approaches to generate a complete persona profile.

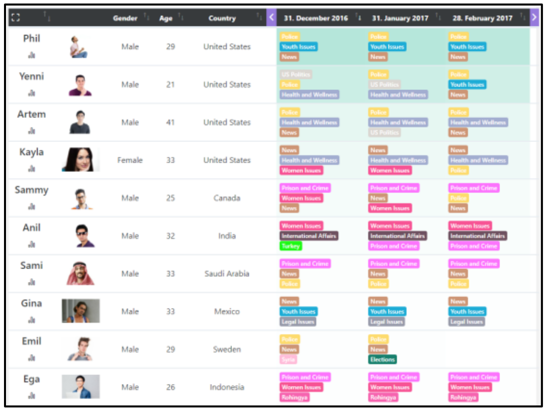

Figure 2: A portion of the persona listings with 10 of the 33 personas showing and their appearance in three of the 24 months. Note: the colors boxes are results of topical classification and are not related to the study.

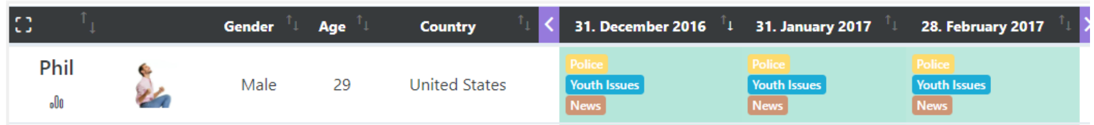

Figure 3: Example of one persona from the persona listings with the three of the 24 data collection months showing. The colors boxes are results of topical classification (not related to the study).

This approach has several advantages, including that is responsive to interactions with both existing and new content, which is important as our datasets are cascading (i.e., existing content will get new user interactions and new content is added that has no prior user interactions). As such, the result of the process is the set of most distinct personas in terms of both behavioral and demographic user attributes.

We apply the identical methodological approach to each monthly dataset, generating fifteen personas each iteration. Although 15 personas are 3-5 times the standard number of organizational personas [14], we deem the higher number of personas reasonable for organizations with varied online audiences. The result of monthly data collection and repeated analysis is a series of monthly sets of organizational personas over the period.

Once we have the complete series of 24 data sets, we then list the 15 personas for each month. After this, we compare the listing of personas for each month to the original list of 15 from October 2016. See Figure 2 (i.e., a listing of multiple personas). We use a system to automate this process and display the listings of personas, as shown in Figure 3 (i.e., one persona listing for readability).

We compare the list of personas month-over-month (MoM) (i.e., overlap expressed with respect to the previous month), year-over-year (YoY) (i.e., overlap expressed with respect to the previous year), and the overlap from the first to the last data collection during the period (C_{FL}) (i.e., overlap expressed with respect to the start of the period). Specifically, we define our metrics as the following:

$$MoM = (P_{PM} \sqcap P_{CM}) / P_{PM}$$

$$YoY = (P_{PY} \sqcap P_{CY}) / P_{PY}$$

$$C_{FL} = (P_F \sqcap P_L) / P_F$$

Where P_{PM} is the set of personas in the prior month, P_{CM} is the set of personas in the current month, P_{PY} is the set of personas at the start of the prior year, P_{CY} is the set of personas at the end of the current year, P_F is the set of personas at the first data collection, and. P_L is the set of personas in the last data collection. Once we have the overlap for each metric, it is trivial to calculate the change (i.e., the change for each period is just one minus the appropriate metric). Returning to our research questions, we now examine the change in personas over time.

Table 2: MoM change showing absolute values and percentages for average, standard deviation, maximum, minimum, and median.

Metric	*Monthly Change*	*% of Change*
Average	2.8	18.3%
Std. Dev.	1.8	12.2%
Max	7.0	46.7%
Min	0.0	0.0%
Median	2.0	13.3%

Table 3: Aggregated MoM changes showing the number of personas changing and the months that this number of personas change occurred, along with the percentage of change from the list of 15 and the percentage of occurrence overall.

No. of Personas	*Months of Occurrence*	*% Change*	*% Occurrence*
0	1	0.0%	4.2%
1	6	6.7%	25.0%
2	6	13.3%	25.0%
3	4	20.0%	16.7%
4	2	26.7%	8.3%
5	3	33.3%	12.5%
6	1	40.0%	4.2%
7	1	46.7%	4.2%
	24		100.0%

Table 4: YoY change showing absolute values and percentages.

Oct-16	*Oct-17*	*% Change*	*Sept-18*	*% Change*
NA	4	26.7%	3	20.0%

Table 5: Period change showing absolute value and percentage.

Beginning (Oct-16)	*End (Sept-18)*	*% Change*
	7	47%

5 RESULTS

5.1 Month-over-Month Change in Personas

We first computed the MoM change, with findings presented in Table 2. As shown in Table 2, the average MoM change was 2.8 personas per month (standard deviation of 1.8). The minimum MoM change was 0, and the maximum MoM change was 7. The median was 2. On average, the MoM change was 18.3%.

In terms of the distribution of MoM changes, the analysis results are shown in Table 3. As shown, there was one month with no MoM change in the persona sets. Most of the MoM changes were one or two personas from the set of 15, which is in line with the average. There were two months with a high percentage of MoM change (6 and 7 of the 15 personas).

These spikes in the MoM change in personas (e.g., the change of 6 and 7 personas) was of interest to us, so we investigated the underlying data for these two months. The months of high change are aligned with a period when the organization was experiencing a dramatic increase in the popularity of content, and its user base was intensely growing. During these months, the organization's content attracted a substantially larger viewer base, reaching millions of views for some of their videos. We conjecture that the population substantially shifted in composition, reflecting a higher than normal rate of change in the personas (more than double the normal rate of change). Since this point, the MoM fluctuations has returned to the normal rate of change.

5.2 Year-over-year Change in Persona

We then computed the YoY change, with findings presented in Table 4. As shown in Table 4, there was a YoY change of 4 personas (26.7%) during year one, and then a YoY change of 3 personas (20.0%) for year two. So, the YoY change is slightly higher (average of 3.5 personas) relative to the MoM change (average of 2.8 personas), indicating that some of these personas are dropping out of the top fifteen and then returning. We investigate this further when we examine the change in personas over the entire period.

5.3 Change in Persona During the Entire Period

We then computed the change in personas over the entire period, with findings presented in Table 5. As shown in Table 5, there were 7 personas that changed from the beginning to the end of the data collection period, with the numbers tracking with the YoY changes.

6 DISCUSSION AND IMPLICATIONS

In this research, we show that personas do change over time, at least for this organization that we examined, and present a simple and practical technique for determining changes in personas. The main implication is that organizations using personas should routinely engage in continuous data collection in order to detect potential changes in their user base. This data collection can be done via traditional means or using automated methods, as done here. The advantage of using online data sources for persona creation is the effortless comparison of changes over time.

Future Research

Future research could entail using the presented comparison approach to determine the optimal number of personas over a period of time, which could be the number where the MoM, or whatever period, of change, is minimized.

This would also address another open gap in personas research, namely, what is the "right" number of personas.

Moreover, our analysis could be enhanced by investigating, in depth, what the changes to the personas were, and why these changes are important for the organization's development decisions.

7 CONCLUSION

We confirm that personas can change over time, empirically validating a criticism of personas that their updating requires continual data collection. In the data set that we present here, there were substantial changes in personas, even within a relatively short period. This indicates that organizations incorporating personas, in whatever manner the data is collected, should engage in continual data collection to ensure personas are responsive to changes in the user population. Future research can look at creating personas from user segments in other domains [11, 13].

REFERENCES

[1] J. An, H. Kwak, J. Salminen, S. G. Jung, and B. J. Jansen, "Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data," *Social Network Analysis and Mining,* vol. 8, p. 54, 2018.

[2] J. An, H. Kwak, J. Salminen, S. G. Jung, and B. J. Jansen, "Imaginary People Representing Real Numbers: Generating Personas from Online Social Media Data," *ACM Transactions on the Web,* vol. 12, p. Article 27, 2018.

[3] C. N. Chapman and R. P. Milham, "The Personas' New Clothes: Methodological and Practical Arguments against a Popular Method," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2006, pp. 634–636.

[4] P. Dharwada, J. S. Greenstein, A. K. Gramopadhye, and S. J. Davis, "A Case Study on Use of Personas in Design and Development of an Audit Management System," in *Human Factors and Ergonomics Society Annual Meeting Proceedings*, Baltimore, Maryland, 2007, pp. 469-473.

[5] A. Drutsa, G. Gusev, and P. Serdyukov, "Periodicity in user engagement with a search engine and its application to online controlled experiments," *ACM Transactions on the Web,* vol. 11, pp. 1–35, 2017.

[6] K. A. Ericsson and H. A. Simon, "Verbal reports as data," *Psychological Review,* vol. 87, pp. 215-251, 1980.

[7] E. Friess, "Personas and Decision Making in the Design Process: An Ethnographic Case Study," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Austin, Texas, USA, 2012, pp. 1209-1218.

[8] K. Goodwin and A. Cooper, *Designing for the Digital Age: How to Create Human-Centered Products and Services.* Indianapolis, IN: Wiley, 2009.

[9] S. Jung, J. An, H. Kwak, M. Ahmad, L. Nielsen, and B. J. Jansen, "Persona Generation from Aggregated Social Media Data," in *ACM Conference on Human Factors in Computing Systems 2017 (CHI2017)*, Denver, CO, 2017, pp. 1748-1755.

[10] H. Kwak, J. An, and B. J. Jansen, "Automatic Generation of Personas Using YouTube Social Media Data," in *Hawaii International Conference on System Sciences (HICSS-50)*, Waikoloa, Hawaii, 2017, pp. 833-842.

[11] C. Li, J. Bai, Z. Wenjun, and Y. Xihao, "Community detection using hierarchical clustering based on edge-weighted similarity in cloud environment," *Information Processing & Management,* vol. 56, pp. 91-109, 2019.

[12] S. Mulder and Z. Yaar, *A Practical Guide to Creating and Using Personas for the Web*: New Riders, 2007.

[13] Pablo Sánchez and A. Bellogín, "Building user profiles based on sequences for content and collaborative filtering," *Information Processing & Management,* vol. 56, pp. 192-211, 2019.

[14] J. Pruitt and J. Grudin, "Personas: Practice and Theory," in *Proceedings of the 2003 Conference on Designing for User Experiences*, San Francisco, California, 2003, pp. 1-15.

[15] P. D. Vecchio, G. Mele, V. Ndou, and G. Secundo, "Creating value from Social Big Data: Implications for Smart Tourism Destinations," *Information Processing & Management,* vol. 54, pp. 847-860, 2018.

[16] G. Viana and J. M. Robert, "The practitioners' points of view on the creation and use of personas for user interface design," in *International Conference on Human-Computer Interaction* 2016, pp. 233-244.

[17] X. Zhang, H.-F. Brown, and A. Shankar, "Data-driven Personas: Constructing Archetypal Users with Clickstreams and User Telemetry," presented at the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, California, USA, 2016.