A Template for Data-Driven Personas: Analyzing 31 Quantitatively Oriented Persona Profiles

Joni Salminen[1,2], Kathleen Guan[3], Lene Nielsen[4], Soon-gyo Jung[1], and Bernard J. Jansen[1]

[1] Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar
[2] University of Turku, Turku, Finland
[3] Georgetown University, Washington, DC, United States
[4] IT University of Copenhagen, Copenhagen, Denmark
jsalminen@hbku.edu.qa

Abstract. Following the proliferation of personified big data and data science algorithms, data-driven user personas (DDPs) are becoming more common in persona design. However, the DDP templates are seemingly diverse and fragmented, prompting a need for a synthesis of the information included in these personas. Analyzing 31 templates for DDPs, we find that DDPs vary greatly by their information richness, as the most informative layout has more than 300% more information categories than the least informative layout. We also find that graphical complexity and information richness do not necessarily correlate. Furthermore, the chosen persona development method may carry over to the information presentation, with quantitative data typically presented as scores, metrics, or tables and qualitative data as text-rich narratives. We did not find one "general template" for DDPs and defining this is difficult due to the variety of the outputs of different methods as well as different information needs of the persona users.

Keywords: Personas; data-driven personas; information design; algorithms

1 Introduction

Quantitative data-driven user personas (DDPs) provide an alternative to qualitatively created personas (QCPs). DDPs can represent user populations in ways that are statistically valid, replicable via algorithms and verifiable by statistical metrics [1, 2]. The proliferation of DDPs is driven by the rise of "personified big data" [3] from social media and online analytics platforms that provides new opportunities to generate personas describing digital user populations.

Moreover, data science algorithms and machine learning libraries have made it possible to automate persona creation processes [4, 5] and to automatically update the personas when the underlying user data changes [6]. Based on these advantages, Human-Computer Interaction (HCI) scholars have proposed many types of profiles and layouts for DDPs [7–10], with varying complexity and informational content. The general goal is to increase quantitatively reliable information in personas.

Nonetheless, the multitude of layouts and templates for DDPs has resulted in two challenges: (1) *there is a lack of a general template for DDPs*, meaning that researchers and practitioners are uncertain of what information to include when using quantitative methods and online user data for the creation of DDPs. Moreover, (2) *it is not well-known what the boundaries are of DDPs relative to QCPs*. QCPs are based on social constructivism and human meaning-making [11] and the understanding that human persona creators infer from other humans (the users) when creating the persona. It has been postulated that persona creation is an immersive practice that in itself enhances understanding about the users. In turn, DDPs, might be limited in their ability to capture human nuances and understand meanings of social importance, as the persona creation takes place via probabilistic calculations that humans have little or no interaction with.

Thus, there is a need for research that critically examines the boundaries of the practice of DDPs and the information included in such personas. Figure 1 illustrates this concern with an example of a statistically valid but potentially non-useful persona. Previous research on DDPs fails to deliver a critical analysis such as this, focusing primarily on evaluating DDPs using technical accuracy metrics [1]. While Nielsen et al. [12] have analyzed the templates of user personas developed by Danish companies, such a review has not been conducted for DDPs personas specifically that, as we argue, require a dedicated analysis of their own.

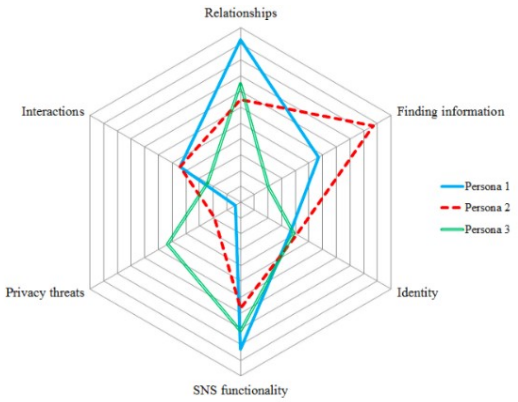


Fig. 1. DDP based on a quantitatively identified patterns of user behavior [13]

Understanding DDPs is important because in-depth information about user motivations and pain points may not be readily available when relying solely on quantitative methods. This is because machine learning methods rely on probabilistic learning rather than a true understanding of human nature, and thus have limited ability to detect human pain points, needs, and wants, as well as goals of individuals [14]. Algorithms are unable to capture tacit information or to understand why a person acts the way he or she does. This limitation might form a fundamental obstacle for the value and usefulness of DDPs, as personas traditionally rely on their ability to convey human-centric information. For HCI, it is a principle of primary importance that personas appear as realistic profiles of otherwise cold and unempathetic “target groups” (as descriptions that cannot evoke empathy), thus enhancing stakeholders’ focus on end user needs [15].

Personas typically contain demographic information, as well as user goals and motivations [12]. The principle of rounded personas [16] calls for the persona to contain all the necessary information for stakeholders using the personas.

Thus, it is important to identify and discuss the boundaries of the information design of DDPs for research and practice. For this, a review of layouts and information of DDPs is needed. To this end, this analysis specifically focuses on DDPs, specifically on their layouts and information designs. Using systematic review methods, we locate and retrieve 31 DDP templates from prior research. We analyze the information in these templates using an extended version of the categories by Nielsen et al. [12].

Relevant studies containing persona profiles were identified, and the content of the persona profiles was extracted to answer the following research questions (RQs):

- **RQ1:** What information do quantitative personas typically contain?
- **RQ2:** What patterns can be found in quantitative persona layouts?
- **RQ3:** How are purely quantitative personas different from qualitative or mixed quantitative-qualitative ones?

Our results indicate gaps in information design for DDPs, demonstrating the limitations of purely quantitative methods to generate rounded personas that serve stakeholders' information needs in a holistic way. To remedy these gaps, we outline potential avenues for the use of algorithms, both independently and in collaboration with humans, to generate more holistic, more rounded DDPs than the current state of the art provides. As such, we provide an important contribution of combining algorithms and machine learning techniques with online user data and human judgment in order to create user personas that involve the benefits of quantitative data but also contain the type of information needed to understand the humans being behind the profile.

2 Related Literature

Persona templates are characterized by influencing each other and very few have looked at research for inspiration. Anvari et. al. [17] have looked at cognitive psychology and learning for inspiration on what to include in the persona description. Nielsen [18] takes inspiration from filmscript writing.

Looking at the literature concerning what to include in the descriptions, there are some variations. Bornet and Brangier [19] describe in their study of the literature how three categories define the persona: (a) identity of the persona, (b) attitude towards the product or service, and (c) context of usage. Their study is built upon nine texts written between 2001 and 2009. Floyd et al. [20] report from 13 papers written between 1999 and 2006 and differentiates between seven kinds of personas that have different characteristics. Some types refer to authors of persona literature and advocacy, others to variations in use context. The types vary in how detailed they have, according to how much and what data they are built upon and the purpose of creation.

Nielsen et al. [12] analyzed 12 templates from 2006-2013. The study shows that the attitude towards the product and the context of use is often intertwined; thus the infor-

mation can be divided into two main areas: (a) personality that includes various information about demographics and personality traits (b) information related to the specific area to design for such as technology use, a-day-in-the-life, products goals and behavioral information. Apart from this, some researchers suggest adding business information such as market size and brand relationships [21–23]. Finally, a few researchers suggest indicating differences that can affect the persona, such as differences between international markets [21, 22, 24] and different behavior according to disabilities [25].

Looking at recommendations from personas based on design team's assumption, the literature recommends a limited amount of information such as name and demographics, behaviors and beliefs, needs and goals [26].

Common is that for both the qualitative data-based personas and the assumption-based personas is that the suggestions are not based on research across disciplines and large amounts of cases but are based on individual experiences and single case studies.

When it comes to the application areas, there is almost no area where personas has not been applied; digital services [27], learning [28], health care [29] are among the most common areas and target groups are both children [30], adults and users with special needs [31] using both mobile devices [32], and web services.

Previous research has shown that DDPs can take many forms and shapes. For example, Aoyama [7] used conjoint analysis to create DDPs for software embedded in digital consumer products. Holden et al. [9] developed "biopsychosocial" DDPs of elderly patients with heart failure using quantitative survey data. DDPs have also been applied in fashion [8], ecommerce [33], news [34], and many other domains. The diversity in persona information design, thus, appears to originate on one hand from *the specificity of the methods applied* – with the intuition that the outputs of different methods enable different information to be used for persona development – and, on the other hand, from *the varying information needs of persona users*, which inarguably affect the goals of the persona development endeavor. Thus, the consequence is that the field is embedded in the diversity of proposed design templates for DDPs. This diversity reflects the increasing relevance of DDPs for researchers and practitioners in user-centric industries.

Moreover, the design of DDPs has been explored both empirically, using experimental designs, and conceptually, by crafting research agendas that entail open questions for what is considered as "optimal" persona template. For example, Hill et al. [35] experimented with two persona designs: one that includes multiple pictures (consisting both of males and females) for a given persona and another one that has only one picture. Using a controlled laboratory study with eye-tracking measurement, they found that the use of multiple pictures may represent an appropriate technique to expand the persona users' understanding of the persona as a gender-free (or, "multi-gender") user segment rather than evoking gender stereotypes [35].

Similarly, Salminen et al. [34] experimented with persona profiles: one with lifestyle photos and the other with a single portrait picture. Contrary to Hill et al. [35], their findings indicated the use of multiple photos can distract and confuse the persona users, possibly because these are more used to the conventional template of the persona including only one photo [34]. Nonetheless, neither Hill et al. [35] or Salminen et al. [34] found that multiple photos would decrease the user engagement with the persona.

In another experimental study, Salminen et al. [36] presented 38 professionals with two alternate layouts: one that used numbers-oriented information presentation style and another one that used text-oriented style. They found that the numbers-oriented template was perceived significantly more usefulness by analysts but significantly less complete by both marketers and analysts [36]. The visual engagement with the persona profiles was found not to vary significantly between the templates [36].

Conceptually, persona information design in the context of DDPs has been raised as one of the prominent research areas [14, 37]. For example, Anvari et al. [38] discuss the use of personality traits in personas: it is unclear how well such traits that require subject-matter expertise and human analysis could be automatically added to DDPs.

3 Methodology

The persona layouts analyzed in this research were retrieved using systematic review methods. Two academic databases (Google Scholar and ACM Digital Library) were consulted for initial identification of articles. Identical literature searches were carried out for both databases in June 2019. The search phrases were devised with references to DDPs ("quantitative personas", "data-driven personas", "procedural personas") in addition to specific methodologies ("automatic persona generation", personas + cluster analysis | clustering | conjoint analysis | factor analysis | latent semantic analysis | matrix factorization | principal component analysis).

Snowball sampling was also applied [39] to identify additional DDP articles. In total, the searches yielded 138 unique articles, which were first assessed by reading the titles and abstracts and, subsequently, reviewing the full texts. The criteria for including an article in the final sample were:

- full research article (no short articles, books or theses)
- published in peer-reviewed journal or conference
- written in the English language
- empirical paper that develops personas using quantitative data

After a full text review, 49 (35.5%) articles remained. For the purposes of this researchers, we further excluded articles, which did not attach graphical representations of their final personas (i.e., persona layouts). At this stage, 30 final articles remained, and their persona layouts were extracted for further analysis. Data from each paper's persona layout(s) was recorded using a standardized data extraction form [40] with subcategories built on the previous work of Nielsen et al. [12] (see Table 1). In addition, the methodology conducted by each study (i.e., whether the paper used statistical and/or numerical techniques such as k-means cluster analysis, solely or in combination with qualitative methods such as ethnography) was also recorded.

Furthermore, the categories were analyzed within the contexts of the authors' methods and goals in their respective papers. This included three papers that contained illegible layouts (i.e., too small or blurry), but nonetheless offered sufficient details in their text regarding the individual components of the persona layouts.

The following section presents the findings. Appendix 1 shows the recorded data.

Table 1. Information extracted from each persona layout, with examples

Subcategory	Description of information content	Examples (verbatim whenever possible)
Name	Full name, first name, or epithet	Eric Transon [41] (p. 632), "Lazy Experts" [42]
Age	Age (or age range) ascribed to the persona	Age 23, "senior student" [7] (p. 6)
Gender	Gender ascribed to the persona	Male/ female
Personality and psychographics	Character traits and disposition of the persona	"Very satisfied with life, usually gets the social support she needs" [43] (p. 66)
Lifestyle	Living situation, leisure, work-life balance	"Lives in central California, frequently walks and gardens" [43] (p. 66)
Experience	The person's experience with the product	"Never interacted with a robot before (…)" [44] (p. 8)
Daily work context	The persona's role and duties in the workplace	"Daily use of e-mail, browsing the web" [7] (p. 6)
Product related behaviors	How the persona interacts with technology and/or tools in the workplace	"During the interaction she kept saying that AIBO was cute and she was enjoying it" [44] (p. 8)
Product goals	What the persona hopes to achieve	"Wants reliable access to all journal articles he needs" [41] (p. 632)
Scenarios	Specific events involving the persona in relation to the product	"I mainly use the library website to find citations or to check whether I can get articles I've found in Google Scholar for free" [41] (p. 632)
A day in the life	Daily context for persona in relation to the product	"She goes out at least once in every two weeks with fellow hikers ... frequently jogs in the field of Shenzhen University" [45] (p. 599)
Market size	Sample size of analyzed population that matches a particular persona	Percentage of time spent in a knowledge worker action section [46]
Color-coding to indicate segment	Color tagging for details of the persona	Yellow highlight
Use of facial picture	Photograph of real person included	N/A
Use of cartoon picture	Cartoon image to represent persona	Cartoon image depicting girl
Reference to sources	Source of data or explanation of metrics	Link to research references [47]
Disabilities	Handicaps of the persona (particularly for papers written in healthcare contexts)	Heart health metrics [9]
International considerations	Cultural heritage, ethnicity and/or citizenship	Non-aboriginal [48]
Explanations	Tooltip definitions	Link to research references [47]

4 Findings

4.1 Levels of Information Richness

The persona layouts varied in "richness," which we define as *containing multifaceted, well-rounded information regarding the persona*. We quantitatively calculated the richness of personas by tallying the total pieces of information (i.e., subcategories present) within each persona layout. The most complex persona layout contained information for 14 subcategories [47], while the least complex contained only 4 [8, 13]. The mean number of subcategories was 8.83, while the standard deviation was 2.57.

Based on the descriptive statistics, the persona layouts were divided into three levels of richness styles: "simple" (4 to 7 subcategories), "moderate" (8 to 10 subcategories), and "high" (11 to 14 subcategories) (see Table 2 for examples). We selected the number of subcategories for the levels after examining the entire dataset and identifying the natural 'breakpoints' in the number of subcategories. Half of the persona layouts (50%) fell under the "moderate" category, with the remainder falling relatively evenly between either "simple" (26.6%) or "high" (23.3%) richness.

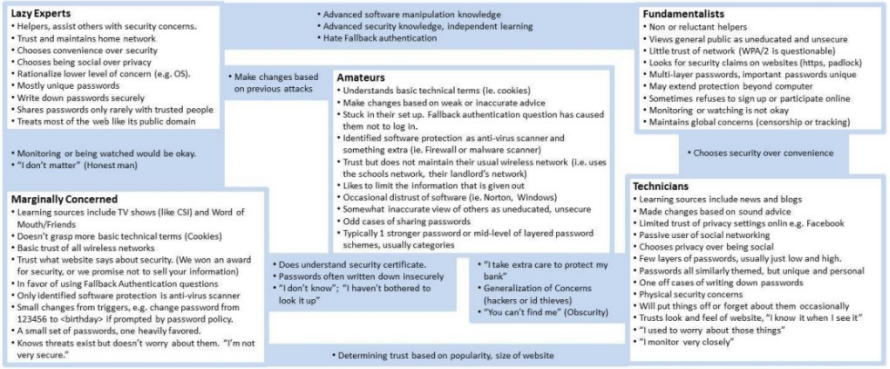
The graphical complexity and information richness of the personas do not necessarily correlate. For example, one persona layout [13], while an interesting graphical way of presenting personas, was questionable in its informativeness for end users; such extreme cases of abstraction were thus categorized under "simple" style despite their graphical complexity (see Table 2).

Persona layouts falling under the most "simple" information style, as exemplified by the layout from Dupree et al. [42] in Fig. 2, contained sparse information limited to bullet points detailing common behaviors. The persona is not identified with characteristics to make it human, like a name, or demographic and psychographic information; instead, it may only be labeled with a general epithet, such as "Lazy Experts" (close to what Floyd et al. [20] term as user archetypes).

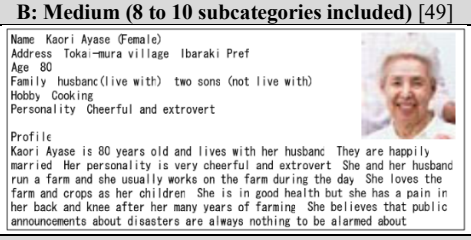
Most "simple" persona layouts could be regarded as "skeleton personas" [43] that can be further enriched with details once time, costs, or limited data are removed as barriers. Persona layouts falling under the "moderate" information style – the most common category – reflect what such an upgrade in resources can result in. As exemplified by Kanno et al. [49], personas in this category are enriched with human-like elements, such as a full name, age, gender, and details on leisurely activities and temperament. In many cases, a photo of a real person is enclosed. The persona layout also contains a short narrative (or in some cases, detailed bullet points) about the persona's daily life scenarios and design-related goals.

Table 2. Examples of each category of persona layouts with varying richness

A: Simple (4 to 7 subcategories included) [42]



B: Medium (8 to 10 subcategories included) [49]



C: High (11 to 14 subcategories included) [41]

Finally, persona layouts falling under the "high" information category are enriched with the most details (see "C" in Table 2). They extend beyond "moderate" information layouts through the inclusion of quotes, graphical representations, and categorization of the persona's information. In short, persona layouts in this category contain not only more comprehensive information on demographic and psychographic details, but also

categorize details in direct relation to the authors' objectives. For example, Templeman-Kluit and Pearce [41] categorize specific details under library usage and frustrations, which are in direct line with the authors' topics of inquiry. Graphical symbols illustrate what relevant devices or subscriptions the persona has (the authors' point of interest). This contrasts with the personas in the "moderate" category ("B" in Table 2), which usually contain only a short narrative with details that are not necessarily arranged into meaningful categories. As such, persona layouts in the "high" category go beyond mere personification and become mediums of analysis, as users can view these layouts to quickly discern between relevant information from various categories.

4.2 Mixed Methods vs. Purely Quantitative Methods

Most articles (56.7%, N=17) adopted solely quantitative methods, while 43% (N=13) of the papers adopted mixed methods (i.e., used both qualitative and quantitative methods). Among the articles that adopted quantitative methods only, 29.4% (N=5) fell under the "simple," 47.1% (N=8) under the "medium," and 23.5% (N=4) under the "high" information styles. Among the articles that adopted mixed methods, 23.1% (N=3) fell under the "simple," 53.8% (N=7) fell under the "moderate," and 23.1% (N=3) fell under the "high" information styles.

As such, quantitative articles fell relatively more often under the "simple" category (29.4% versus 23.1% for mixed method studies). Beyond this, no other major differences in richness could be observed between papers, either adopting quantitative methods solely or in combination with qualitative methods. These findings are similar to Nielsen et al.'s, who found in their analysis that companies with the lack of a formal quantitative data collection protocol nonetheless still resulted in final personas as lengthy and with descriptions just as thorough as those with data [12].

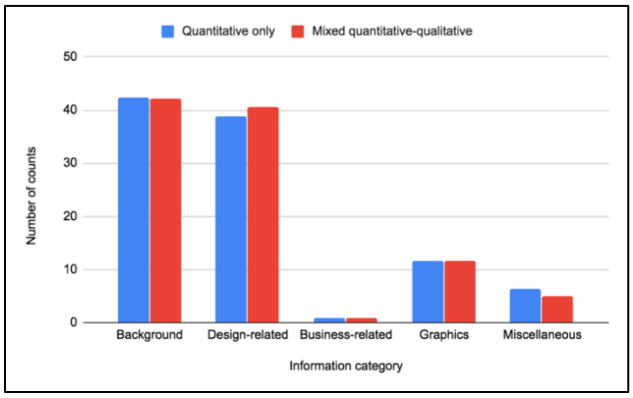


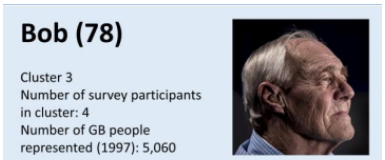
Fig. 2. Information category distribution by methodology

When analyzing the number of information pieces from personas from papers adopting mixed methods, 42.1% fell under the background, 40.5% under the design-related, 0.8% under business and marketing related, 11.6% under the graphics, and 5.0% under miscellaneous categories. A similar distribution was observed for personal layouts from

papers adopting only quantitative methods, 42.3% of details fell under the background, 38.7% under the design-related, 0.9% under business and marketing related, 11.7% fell under the graphics, and 6.3% under miscellaneous categories (see Fig. 2).

Table 3 displays examples of how solely quantitative versus mixed method approaches differ in presentation of persona layouts. Each of the example layouts (quantitative from Goodman-Deane et al. [50] and qualitative from Tu et al. [45]) detail mostly background and design-related information. Nonetheless, the solely quantitative approach [50] results in a chart-like presentation of the details, with "scores" directly representing the quantitative data from the survey. On the other hand, the mixed method approach [45] results in more narrative-like, contextual descriptions.

Table 3. Purely quantitative versus mixed method persona



Bob (78)

Cluster 3
Number of survey participants in cluster: 4
Number of GB people represented (1997): 5,060

Bob is retired and lives with his wife. He likes spending time with his grandchildren, watching TV and swimming. He uses a basic mobile phone to communicate with friends and family.
Bob has Spondylitis, angina and age-related hearing loss. He wears a hearing aid.

Capability	Score	Description
Walking	3	Can walk 200 yards but not 400
Fine-finger	3	Can pick up a pin with each hand but not tie a bow without difficulty
Strength	3	Can pick up a pint of milk in each hand but not a bag of potatoes
Hearing sound	3	Can hear a telephone ring but not follow TV at a volume others find acceptable
Hearing speech	2	Can hear someone talking in a loud voice but not use an ordinary telephone

Basic information
Name：Anna
Gender: female
Age: 30
Education background: university graduate
Occupation: mid level management
Characteristic: Extrovert . Anna loves movie、drama、coffee and music. She has a stable job and wonderful family. Anna is a fashionable. She enjoys name brand products. However, because of her limited income, she tries to buy products with best value. Anna is receptable for new products, but she only buys it until she tries them or some friends recommend them.
Income per month: 5000 RMB
Consumption per month: 3000 RMB
Outdoor Experience:
Anna started outdoor activities about a year ago. Now she a owns a complete set of branded outdoor gears. Although Anna is busy with job, she will go out at least once every two weeks at least.
Use's goal for listening to music in outdoor activities:
1. Outdoor activities such as hiking is tiring. Listening to music is a good motivator.
2. I love music. Sport and music are indispensable parts of my life.

Use's background
Two years ago, Anna, who has a stable family and a stable job, found herself in a sub health condition. She began to change her lifestyle and eating habits. She started to do some exercise to improve her physical health. About one year ago, Anna started to participate in outdoor activities. She now owns many branded outdoor gears. Anna enjoyes outdoor as she can embrace the nature. She goes out at least once in every two weeks with fellow hikers from an internet forum. She has made many good friends who love outdoor just like her. Sometimes Anna goes out alone. She likes jogging as much as climbing. She frequently jogs in the field of Shenzhen University.

Purely quantitative persona [50]	**Mixed method persona [45]**

Furthermore, the former only captures "work related issues," "daily life context," and "product related issues" in subcategories with the design-related category, while the latter captures these in addition to the "product goals," "scenarios," and "a day in the life" subcategories. The former also does not capture any personality and psychographic information within the background category, while the latter infuses many of these details into narrative form. Thus, we surmise that the type of information collected for persona development (i.e., quantitative vs. qualitative) may carry over to the actually design of the persona, with numerical details such as graphs, scores, metrics, and tables being more common with purely quantitative personas and text-focused, narrative-like descriptions more prevalent in mixed method personas.

However, most layouts (especially in the "high" information richness category) combine both information styles, with some numerical cues and some textual information. The degree of text vs. numbers in data-driven personas is an open research question,

with some previous research showing that the persona developers' choices can affect the persona perceptions of users [36].

5 Discussion

While Nielsen et al. [12] found that most persona layouts from Danish organizations had a "strict distinction" between personas and scenarios, our analysis found that personas were generally intertwined with descriptions of present scenarios. Many of the richest personas (i.e. falling under the "high" information category) had narratives infused into the persona layouts' descriptions to give them a more human-like quality. Interestingly, Nielsen et al. [12] reported that companies found this style "difficult to use (…) for the design of future solutions and as a result cancelled using the method [of intertwining]" (p. 6). This suggests the importance of considering the layout and categorization of information in a persona in conjunction with the researchers' and/or practitioners' needs. From our analysis, the persona layout from dos Santos et al. [44] is an example of how pertinent scenarios can be embedded into personas in a manner that remains relevant to stakeholders (see Fig. 3).

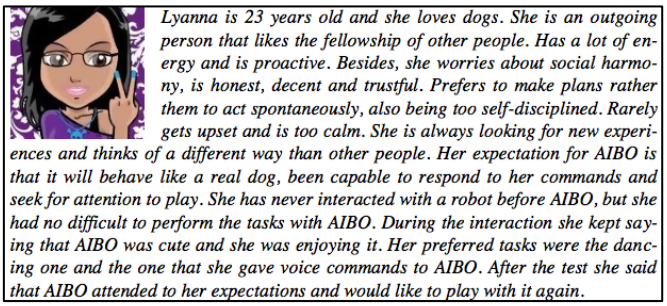


Fig. 3. Example of scenarios embedded into personas [44]

Furthermore, while Nielsen et al. [12] found a "noteworthy difference" in the "lack of information on income, urbanicity, and lifestyle" (p. 5) in their analysis, we did not find this to be the case. Moreover, Nielsen et al. also found that market segments were rarely captured in persona templates from Danish companies, which we also did not find to be the case. This may be attributed to a greater diversity in contexts observed in our set of persona layouts. As we only included studies that were data-driven and did not exclude by geographical region, such demographic details were intentionally included in many of the persona layouts. For example, studies developing personas for e-health devices found it pertinent to capture lifestyle and subpopulation distribution percentages [9], while studies conducted in market research and business contexts captured relevant income, lifestyle, and urbanicity data of potential customers [2, 45].

Nielsen et al.'s [12] finding that Danish persona descriptions lacked business and marketing related information was also reflected in our analysis of international, data-driven persona layouts. We also found it to be the case that "even though it is stated in several interviews that personas are used as a strategic tool and in marketing, the descriptions do not reflect this" (p. 6). Specifically, information pertaining to competitors,

business objectives, and brand relationship were not captured in any of the persona layouts we analyzed. Rather, relevant information to businesses was mostly indirectly captured in the design-related categories, in relation to the personas' product goals, work-related issues, and daily life or work context. This may suggest that the persona layouts were designed to envision common product usage scenarios among customers rather than to explicitly to illustrate and correspond to business objectives.

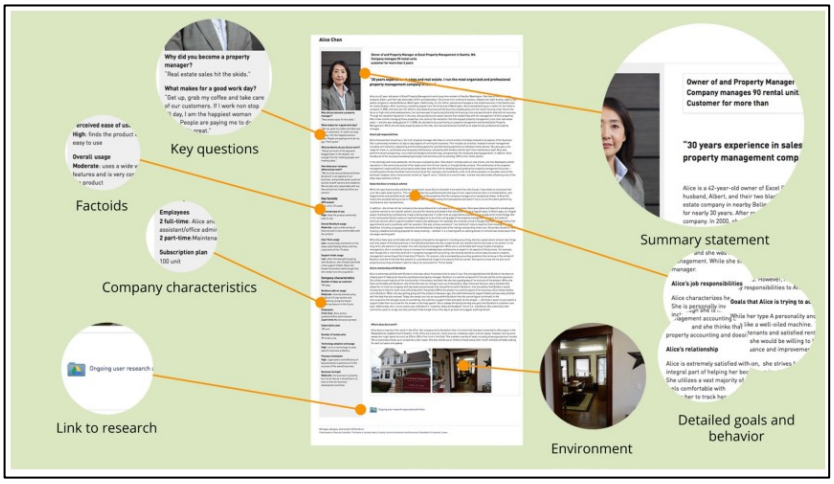


Fig. 4. Example of visual identification aids in personas [47]

Moreover, the persona layouts in our analysis also reflected Nielsen et al.'s [12] observation that researchers developed "different ways of fostering identification (…) the use of keywords, headlines, and quotes give a quick understanding of the kernel of the persona description" (p. 6). Our own analysis found that a variety of visual cues across persona layouts, from color coding, use of icons, visual scales, and even data charts [47] (see Fig. 4 for an example).

The field is in dire need of empirical user studies. Nascent work shows promise in applying methodologically diverse methods such as eye-tracking, think-aloud, observer notes [34, 35], and examining multiple aspects of persona design, such as photos, text vs. numbers, and so on [36, 51, 52]. Yet, there is a lack of systematic research that would incrementally advance the design practice of personas into a more *optimal* state. Currently, some of the empirical findings are conflicting (such as those by Hill et al. [35] and Salminen et al. [34] regarding the use of multiple photos). We surmise that this is due to variations in the implementation of persona templates – both small and large variations can affect user experiences in crucial ways. In other words, the persona templates tested by different research "look and feel" different and thus are perceived as and engaged with in different ways. The only way, it thus appears, to produce consistent research insights that are generalizable across the nuanced implementations of DDPs, is to include more design variations in these user studies. This would, consequently, require the use of large-scale data collection, potentially prompting for more scalable data collection such as persona crowd experiments. There are already existing

examples of using crowdsourcing for data collection in persona studies [51, 52]; however, more efforts are needed.

6 Conclusion

An empirical analysis shows that DDP layouts draws heavily from both quantitative and qualitative data. Some persona information (e.g., lifestyle, personality) is difficult to obtain using purely quantitative methods, thus requiring qualitative insights to realistically portray this information. We also find that graphical complexity and information richness do not necessarily correlate. The range of information categories is high, as the most information-rich persona template has more than 300% more information categories than the least information-rich template. Furthermore, the choice of the persona development methodology may carry over to the information design of DDPs, with quantitative data typically presented as scores, metrics, or tables and qualitative data as text-rich narratives. We did not find one "general template" for DDPs; this cannot be defined easily if at all, due to the variety of the outputs of different methods as well as differences in the information needs of the persona users.

References

1. Brickey, J., Walczak, S., Burgess, T.: Comparing Semi-Automated Clustering Methods for Persona Development. IEEE Transactions on Software Engineering. 38, 537–546 (2012). https://doi.org/10.1109/TSE.2011.60.
2. Tu, N., Dong, X., Rau, P.P., Zhang, T.: Using cluster analysis in Persona development. In: 2010 8th International Conference on Supply Chain Management and Information. pp. 1–5 (2010).
3. Stevenson, P.D., Mattson, C.A.: The Personification of Big Data. Proceedings of the Design Society: International Conference on Engineering Design. 1, 4019–4028 (2019). https://doi.org/10.1017/dsi.2019.409.
4. Jung, S., Salminen, J., An, J., Kwak, H., Jansen, B.J.: Automatically Conceptualizing Social Media Analytics Data via Personas. In: Proceedings of the International AAAI Conference on Web and Social Media (ICWSM 2018). , San Francisco, California, USA (2018).
5. Jung, S., Salminen, J., Kwak, H., An, J., Jansen, B.J.: Automatic Persona Generation (APG): A Rationale and Demonstration. In: Proceedings of the 2018 Conference on Human Information Interaction&Retrieval. pp. 321–324. ACM (2018).
6. Jung, S., Salminen, J., Jansen, B.J.: Personas Changing Over Time: Analyzing Variations of Data-Driven Personas During a Two-Year Period. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. p. LBW2714:1–LBW2714:6. ACM, Glasgow, UK (2019). https://doi.org/10.1145/3290607.3312955.
7. Aoyama, M.: Persona-and-scenario based requirements engineering for software embedded in digital consumer products. In: Proceedings of the 13th IEEE International Conference on Requirements Engineering (RE'05). pp. 85–94. , Washington, DC, USA (2005). https://doi.org/10.1109/RE.2005.50.
8. Dhakad, L., Das, M., Bhattacharyya, C., Datta, S., Kale, M., Mehta, V.: SOPER: Discovering the Influence of Fashion and the Many Faces of User from Session Logs using Stick Breaking Process. In: Proceedings of the 2017 ACM on Conference on Information and

Knowledge Management - CIKM '17. pp. 1609–1618. ACM Press, Singapore, Singapore (2017). https://doi.org/10.1145/3132847.3133007.
9. Holden, R.J., Kulanthaivel, A., Purkayastha, S., Goggins, K.M., Kripalani, S.: Know thy eHealth user: Development of biopsychosocial personas from a study of older adults with heart failure. International Journal of Medical Informatics. 108, 158–167 (2017). https://doi.org/10.1016/j.ijmedinf.2017.10.006.
10. An, J., Kwak, H., Jung, S., Salminen, J., Jansen, B.J.: Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data. Social Network Analysis and Mining. 8, (2018). https://doi.org/10.1007/s13278-018-0531-0.
11. Denzin, N.K., Lincoln, Y.S.: Introduction: entering the field of qualitative research. In: Strategies of Qualitative Inquiry. pp. 1–34. Thousand Oaks: Sage (1998).
12. Nielsen, L., Hansen, K.S., Stage, J., Billestrup, J.: A Template for Design Personas: Analysis of 47 Persona Descriptions from Danish Industries and Organizations. International Journal of Sociotechnology and Knowledge Development. 7, 45–61 (2015). https://doi.org/10.4018/ijskd.2015010104.
13. Dang-Pham, D., Pittayachawan, S., Nkhoma, M.: Demystifying online personas of Vietnamese young adults on Facebook: A Q-methodology approach. Australasian Journal of Information Systems. 19, (2015). https://doi.org/10.3127/ajis.v19i0.1204.
14. Salminen, J., Jansen, B.J., An, J., Kwak, H., Jung, S.-G.: Automatic Persona Generation for Online Content Creators: Conceptual Rationale and a Research Agenda. In: Nielsen, L. (ed.) Personas - User Focused Design. pp. 135–160. Springer London, London (2019). https://doi.org/10.1007/978-1-4471-7427-1_8.
15. Cooper, A.: The Inmates Are Running the Asylum: Why High Tech Products Drive Us Crazy and How to Restore the Sanity. Sams - Pearson Education, Indianapolis, IN (1999).
16. Nielsen, L.: Personas - User Focused Design. Springer, New York, NY, USA (2019).
17. Anvari, F., Tran, H.M.T.: Persona ontology for user centred design professionals. In: The ICIME 4th International Conference on Information Management and Evaluation. pp. 35–44. , Ho Chi Minh City, Vietnam (2013).
18. Nielsen, L.: Engaging personas and narrative scenarios, http://personas.dk/wp-content/samlet-udgave-til-load.pdf, (2004).
19. Bornet, C., Brangier, E.: La méthode des personas: principes, intérêts et limites. Bulletin de psychologie. 115–134 (2013).
20. Floyd, I.R., Jones, M.C., Twidale, M.B.: Resolving Incommensurable Debates: A Preliminary Identification of Persona Kinds, Attributes, and Characteristics. Artifact. 2, 12–26 (2008). https://doi.org/10.1080/17493460802276836.
21. Jones, M., Marsden, G.: Mobile Interaction Design. Wiley (2006).
22. Mulder, S., Yaar, Z.: The User is Always Right: A Practical Guide to Creating and Using Personas for the Web. New Riders (2006).
23. Pruitt, J., Adlin, T.: The Persona Lifecycle: Keeping People in Mind Throughout Product Design. Morgan Kaufmann, Boston (2006).
24. Goodman, E., Kuniavsky, M., Moed, A.: Observing the User Experience: A Practitioner's Guide to User Research. Morgan Kaufmann (2013).
25. Pichler, R.: A template for writing great personas. (2012).
26. Seiden, J., Gothelf, J.: Lean UX: Applying Lean Principles to Improve User Experience. Oreilly (2003).
27. Miaskiewicz, T., Kozar, K.A.: Personas and user-centered design: How can personas benefit product design processes? Design Studies. 32, 417–430 (2011).

28.Dantin, U.: Application of personas in user interface design for educational software. In: Proceedings of the 7th Australasian conference on Computing education-Volume 42. pp. 239–247. Australian Computer Society, Inc., Newcastle, New South Wales, Australia (2005).
29.LeRouge, C., Ma, J., Sneha, S., Tolle, K.: User profiles and personas in the design and development of consumer health technologies. Int J Med Inform. 82, e251-268 (2013). https://doi.org/10.1016/j.ijmedinf.2011.03.006.
30.Antle, A.N.: Child-personas: fact or fiction? In: Proceedings of the 6th conference on Designing Interactive systems. pp. 22–30. ACM (2006).
31.Loitsch, C., Weber, G., Voegler, J.: Teaching accessibility with Personas. In: International Conference on Computers Helping People with Special Needs. pp. 453–460. Springer (2016).
32.Sedlmayr, B., Schöffler, J., Prokosch, H.-U., Sedlmayr, M.: User-centered design of a mobile medication management. Informatics for Health and Social Care. 44, 152–163 (2019).
33.Al-Qirim, N.: Personas of e-commerce adoption in small businesses in New Zealand. Journal of Electronic Commerce in Organizations (JECO). 4, 18–45 (2006).
34.Salminen, J., Nielsen, L., Jung, S.-G., An, J., Kwak, H., Jansen, B.J.: "Is More Better?": Impact of Multiple Photos on Perception of Persona Profiles. In: Proceedings of ACM CHI Conference on Human Factors in Computing Systems (CHI2018). ACM, Montréal, Canada (2018). https://doi.org/10.1145/3173574.3173891.
35.Hill, C.G., Haag, M., Oleson, A., Mendez, C., Marsden, N., Sarma, A., Burnett, M.: Gender-Inclusiveness Personas vs. Stereotyping: Can We Have it Both Ways? In: Proceedings of the 2017 CHI Conference. pp. 6658–6671. ACM Press, Denver, Colorado, USA (2017). https://doi.org/10.1145/3025453.3025609.
36.Salminen, J., Liu, Y.-H., Sengun, S., Santos, J.M., Jung, S.-G., Jansen, B.J.: The Effect of Numerical and Textual Information on Visual Engagement and Perceptions of AI-Driven Persona Interfaces. In: Proceedings of the ACM Intelligent User Interfaces (IUI'20). ACM, Cagliary, Italy (2020).
37.Salminen, J., Jung, S. gyo, Jansen, B.J.: The future of data-driven personas: A marriage of online analytics numbers and human attributes. In: ICEIS 2019 - Proceedings of the 21st International Conference on Enterprise Information Systems. pp. 596–603. SciTePress, Heraklion, Greece (2019).
38.Anvari, F., Richards, D., Hitchens, M., Babar, M.A.: Effectiveness of Persona with Personality Traits on Conceptual Design. In: Proceedings of the 37th International Conference on Software Engineering - Volume 2. pp. 263–272. IEEE Press, Piscataway, NJ, USA (2015).
39.Radjenović, D., Heričko, M., Torkar, R., Živkovič, A.: Software fault prediction metrics: A systematic literature review. Information and software technology. 55, 1397–1418 (2013).
40.Zhu, E., Hadadgar, A., Masiello, I., Zary, N.: Augmented reality in healthcare education: an integrative review. PeerJ. 2, e469 (2014).
41.Tempelman-Kluit, N., Pearce, A.: Invoking the User from Data to Design. CRL. 75, 616–640 (2014). https://doi.org/10.5860/crl.75.5.616.
42.Dupree, J.L., Devries, R., Berry, D.M., Lank, E.: Privacy Personas: Clustering Users via Attitudes and Behaviors Toward Security Practices. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. pp. 5228–5239. ACM, New York, NY, USA (2016). https://doi.org/10.1145/2858036.2858214.

43.Zhu, H., Wang, H., Carroll, J.M.: Creating Persona Skeletons from Imbalanced Datasets - A Case Study using U.S. Older Adults' Health Data. In: Proceedings of the 2019 on Designing Interactive Systems Conference - DIS '19. pp. 61–70. ACM Press, San Diego, CA, USA (2019). https://doi.org/10.1145/3322276.3322285.
44.dos Santos, T.F., de Castro, D.G., Masiero, A.A., Junior, P.T.A.: Behavioral persona for human-robot interaction: a study based on pet robot. In: International Conference on Human-Computer Interaction. pp. 687–696. Springer (2014).
45.Tu, N., He, Q., Zhang, T., Zhang, H., Li, Y., Xu, H., Xiang, Y.: Combine Qualitative and Quantitative Methods to Create Persona. In: 2010 3rd International Conference on Information Management, Innovation Management and Industrial Engineering. pp. 597–603 (2010). https://doi.org/10.1109/ICIII.2010.463.
46.Ford, D., Zimmermann, T., Bird, C., Nagappan, N.: Characterizing Software Engineering Work with Personas Based on Knowledge Worker Actions. In: Proceedings of the 11th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. pp. 394–403. IEEE Press, Piscataway, NJ, USA (2017). https://doi.org/10.1109/ESEM.2017.54.
47.Miaskiewicz, T., Luxmoore, C.: The Use of Data-Driven Personas to Facilitate Organizational Adoption–A Case Study. The Design Journal. 20, 357–374 (2017).
48.Brooks, C., Greer, J.: Explaining Predictive Models to Learning Specialists Using Personas. In: Proceedings of the Fourth International Conference on Learning Analytics And Knowledge. pp. 26–30. ACM, Indianapolis, Indiana, USA (2014). https://doi.org/10.1145/2567574.2567612.
49.Kanno, T., Ooyabu, T., Furuta, K.: Integrating Human Modeling and Simulation with the Persona Method. In: Stephanidis, C. (ed.) Universal Access in Human-Computer Interaction. Users Diversity. pp. 51–60. Springer Berlin Heidelberg (2011).
50.Goodman-Deane, J., Waller, S., Demin, D., González-de-Heredia, A., Bradley, M., Clarkson, J.P.: Evaluating Inclusivity using Quantitative Personas. Presented at the Design Research Society Conference 2018 June 28 (2018). https://doi.org/10.21606/drs.2018.400.
51.Salminen, J., Santos, J.M., Jung, S.-G., Eslami, M., Jansen, B.J.: Persona Transparency: Analyzing the Impact of Explanations on Perceptions of Data-Driven Personas. International Journal of Human–Computer Interaction. 0, 1–13 (2019). https://doi.org/10.1080/10447318.2019.1688946.
52.Salminen, J., Jung, S.-G., Santos, J.M., Jansen, B.J.: Does a Smile Matter if the Person Is Not Real?: The Effect of a Smile and Stock Photos on Persona Perceptions. International Journal of Human–Computer Interaction. 0, 1–23 (2019). https://doi.org/10.1080/10447318.2019.1664068.

Appendix 1: Full coding results

				Background					Design-related						Business	Graphics			Miscellaneous			
Year	**Title**	**Only quantitative**	**Mixed quant-qual**	**Name**	**Age**	**Gender**	**Personality and psychographics**	**Urbanicity**	**Work related issues**	**Daily life or work context**	**Product related issues**	**Product goals**	**Scenarios**	**A day in the life**	**Market size, segment**	**Color-coding**	**Use of facial picture**	**Use of cartoon picture**	**Reference to sources**	**Disabilities**	**International considerations**	**Explanations**
2005	Persona-and-scenario based requirements engineering for software embedded in digital consumer products		X	X	X	X	X	X	X	X	X	X						X				
2007	Persona-scenario-goal methodology for user-centered requirements engineering	X		X	X	X	X	X	X	X	X	X						X				
2008	A Latent Semantic Analysis Methodology for the Identification and Creation of Personas	X		X	X	X	X	X	X	X	X	X	X				X					
2008	Data-driven persona development	X		X			X	X	X			X		X	X		X					
2009	Developing and validating personas in e-commerce: A heuristic approach	X		X			X	X			X	X				X						
2010	Combine Qualitative and Quantitative Methods to Create Persona		X	X	X	X	X	X	X	X		X	X	X			X						
2010	Using cluster analysis in persona development		X	X	X		X	X	X	X	X	X	X				X						

				Background					Design-related						Business	Graphics			Miscellaneous			
Year	**Title**	**Only quantitative**	**Mixed quant-qual**	**Name**	**Age**	**Gender**	**Personality and psychographics**	**Urbanicity**	**Work related issues**	**Daily life or work context**	**Product related issues**	**Product goals**	**Scenarios**	**A day in the life**	**Market size, segment**	**Color-coding**	**Use of facial picture**	**Use of cartoon picture**	**Reference to sources**	**Disabilities**	**International considerations**	**Explanations**
2011	Integrating Human Modeling and Simulation with the Persona Method			X	X	X	X	X	X		X		X	X			X					
2011	Behavioral Persona for Human-Robot Interaction: A Study Based on Pet Robot	X	X	X	X	X	X	X	X		X	X	X					X				
2012	Basic senior personas: a representative design tool covering the spectrum of European older adults	X		X		X	X	X	X	X	X	X	X			X	X			X		
2013	Learning Latent Personas of Film Characters	X		X	X	X	X	X											X		X	
2014	Personas in the Middle: Automated Support for Creating Personas As Focal Points in Feature Gathering Forums	X		X		X	X			X		X	X				X					X
2014	Explaining predictive models to learning specialists using personas		X	X		X	X	X		X							X				X	
2014	Invoking the User from Data to Design	X		X	X	X		X	X	X	X	X	X		X		X					
2015	Demystifying online personas of Vietnamese young adults on Facebook: A Q-methodology approach		X				X		X		X					X						
2016	A Factor Analysis Approach to Persona Development using Survey Data	X			X			X	X	X	X											

				Background					Design-related						Business	Graphics			Miscellaneous			
Year	**Title**	**Only quantitative**	**Mixed quant-qual**	**Name**	**Age**	**Gender**	**Personality and psychographics**	**Urbanicity**	**Work related issues**	**Daily life or work context**	**Product related issues**	**Product goals**	**Scenarios**	**A day in the life**	**Market size, segment**	**Color-coding**	**Use of facial picture**	**Use of cartoon picture**	**Reference to sources**	**Disabilities**	**International considerations**	**Explanations**
2016	Privacy Personas: Clustering Users via Attitudes and Behaviors Toward Security Practices	X		X			X		X	X	X	X										
2017	Animal personas: representing dog stakeholders in interaction design		X	X	X	X	X	X		X		X					X				X	
2017	ID3P: Iterative Data-driven Development of Persona Based on Quantitative Evaluation and Revision	X		X	X	X			X		X	X					X					X
2017	Know thy eHealth user: Development of biopsychosocial personas from a study of older adults with heart failure	X		X	X	X	X	X	X	X	X	X	X				X			X	X	
2017	SOPER: Discovering the Influence of Fashion and the Many Faces of User from Session Logs using Stick Breaking Process	X						X				X	X			X						
2017	The Use of Data-Driven Personas to Facilitate Organizational Adoption–A Case Study		X	X		X	X	X	X	X	X	X	X	X		X	X		X			X
2017	Characterizing Software Engineering Work with Personas Based on Knowledge Worker Actions		X	X		X	X		X	X	X	X			X							
2018	Analysis of Regional Group Health Persona Based on Image Recognition	X				X	X	X	X	X								X				
2018	Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data	X		X	X	X		X	X	X	X	X				X	X					

				Background					Design-related						Business	Graphics			Miscellaneous			
Year	**Title**	**Only quantitative**	**Mixed quant-qual**	**Name**	**Age**	**Gender**	**Personality and psychographics**	**Urbanicity**	**Work related issues**	**Daily life or work context**	**Product related issues**	**Product goals**	**Scenarios**	**A day in the life**	**Market size, segment**	**Color-coding**	**Use of facial picture**	**Use of cartoon picture**	**Reference to sources**	**Disabilities**	**International considerations**	**Explanations**
2018	Evaluating Inclusivity using Quantitative Personas	X		X		X		X	X	X	X				X		X					X
2018	From 2,772 segments to five personas: Summarizing a diverse online audience by generating culturally adapted personas		X	X	X	X		X	X	X	X	X				X	X					
2018	Imaginary People Representing Real Numbers: Generating Personas from Online Social Media Data		X	X	X	X		X	X	X	X	X				X	X					
2018	Research on the Annual Reading Report of Academic Libraries Based on Personas		X			X		X	X	X	X	X										
2019	Creating Persona Skeletons from Imbalanced Datasets - A Case Study using U.S. Older Adults' Health Data		X	X	X	X	X	X	X	X	X		X				X			X		X