

# Invoking the User from Data to Design

Nadaleen Tempelman-Kluit and Alexa Pearce

Personas, stemming from the field of user-centered design (UCD), are hypothetical users that represent the behaviors, goals, and values of actual users. This study describes the creation of personas in an academic library. With the goal of leveraging service-generated data, the authors coded a sample of chat reference transcripts, producing two numeric values for each. The transcripts were plotted on an X/Y graph where X represented the nature of the user's information need and Y represented the nature of the user's motivation. A k-means cluster analysis of the plotted points produced four clusters, which served as the personas' basis.



User personas are increasingly recognized by libraries as a useful and meaningful way to learn about and design services for their user communities. While libraries have made significant progress in adopting a service-oriented and user-centered focus, they remain challenged by the realities of knowing and meeting the needs of diverse and varied clientele. For many academic libraries with service offerings across multiple physical and virtual locations, efforts to serve a generic “user” are insufficient for effective design of services and interfaces. Personas, which come from the field of user-centered design (UCD) and function as archetypes or composites based on real user goals and behaviors, are a tool holding great potential for libraries in understanding and meeting the needs of complex and evolving communities.

The gradual shift in academic libraries' service offerings from a focus primarily on collections to a focus on user-oriented services has received attention throughout all areas of library activities and operations. Walter has argued, “[I]n an era when everything we know about how content is created, acquired, accessed, evaluated, disseminated, employed, and preserved for the future is in flux, the research library must be distinguished by the scope and quality of its service programs in the same way it has long been by the breadth and depth of its locally-held collections.”<sup>1</sup> This shift emphasizes the need for libraries to gain a better understanding of their users. Leanne Bowler et al. have asserted that “(c)onsidering the needs of the user is a core competency of librarianship,” adding that we should “review user-centered design in a critical, reflective, and multilayered manner that reveals the rich array of experiences

---

*Nadaleen Tempelman-Kluit is Discovery & Digital Access Librarian at New York University, e-mail: ntk2@nyu.edu; Alexa Pearce is Librarian for History & American Culture at the University of Michigan Library, e-mail: alexap@umich.edu. ©2014 Nadaleen Tempelman-Kluit and Alexa Pearce, Attribution-NonCommercial (<http://creativecommons.org/licenses/by-nc/3.0/>) CC BY-NC*

in LIS.<sup>22</sup> Such a critical review entails research into the development of UCD methods that guide the design and development of interfaces and services.

For many libraries, an enhanced focus on service design and development has necessitated new and data-driven methods of assessment. Accordingly, we identified personas as a tool to both help us know and design for our users by synthesizing our growing body of service-generated data into meaningful archetypes. To test this supposition, we coded Ask a Librarian (AAL) chat transcripts for criteria that typically make up personas (namely, user need and motivation). In so doing, we developed an evidence-based method of persona creation to address the frequent criticism that they lack rigor and precision.

The NYU Libraries AAL chat service, described in detail below, generates significant quantitative and qualitative data, similar in nature to ethnographic research. Like ethnographic interviews, chat reference transcripts often consist of goal-directed conversations surrounding users' specific needs and their interactions with library tools, services, and resources. In addition to chat, NYU provides in-person reference as well as via text and e-mail. Based on their volume and accessibility, we chose to use chat transcripts for this study; however, reference conversations from any of the other venues would be similarly eligible for this type of analysis.

### User-Centered Design (UCD)

User-centered design is both a design philosophy and a process focused on optimizing interfaces in response to how people work, rather than expecting people to alter their work habits to accommodate the demands of the interface. Gould, in *The Handbook of Human-Computer Interaction*, provides four principles of UCD: "early focus on users and tasks through direct and ongoing contact; empirical measurement, i.e., testing against established nontrivial performance measures; iterative design, in which successive prototypes are tested and refined; and integrated design, or the simultaneous coordination of these principles throughout the design process."<sup>3</sup>

A typical UCD model includes analysis, design, implementation, and deployment phases, with specific methods and techniques employed in each phase. Alan Cooper introduced his goal-directed design method in *The Inmates Are Running the Asylum*.<sup>4</sup> His UCD method questioned the traditional approach of building interfaces and then correcting problems later. Arguing that this approach is inefficient and ineffective, Cooper advocated for a model driven by user research at a project's inception. Personas, data-driven representations of users' goals and behaviors, were the method he proposed to facilitate this model.

Personas quickly gained traction in the design world, where the problem of designing for ambiguous users has been identified as a source of confusion and misunderstanding. Because personas characterize users with specific qualities and needs, they can be easier for designers and developers to connect and identify with than a laundry list of requirements. As Guenther says, "(w)e can understand and draw insights from human characteristics—even composites of characteristics with fictional names—more readily than we can understand sets of tasks."<sup>5</sup> In addition to providing an opportunity for connection, personas have proven helpful in clarifying design decisions, encouraging consensus and engagement within teams, providing a framework for prioritization, and validating development and design.

NYU Libraries has recognized a shortage of proven methods for including user needs at the planning phase of a project or service but has committed to becoming a more data-informed and user-centered institution. After identifying personas as a tool to help bridge this gap, we coded and clustered AAL chat transcripts for user need and motivation, producing four distinct clusters that formed the basis for the final

personas presented in this study. For libraries with similar chat reference services or that wish to incorporate service-generated data from other sources into design and development activities, our method can serve as a template.

### Review of the Literature

A survey of the library literature on the subject of user-centered design indicates usability testing is the most frequently employed method. In a survey of the 113 ARL libraries with a 74 percent response rate, Chen, Germane, and Yang found that 85 percent performed usability testing on some aspect of their website.<sup>6</sup> Ward and Hiller note that, in the past decade, "usability testing has become an integral component of Web design and development in libraries."<sup>7</sup> Defined in a number of ways, usability testing is generally used to assess the design effectiveness of an interface. There are a number of usability testing methods, but formal usability testing is favored in the library literature. This method is usually employed toward the end of the design cycle to validate interfaces. Early studies set the groundwork for the evaluation of library websites through usability methods, largely influenced by writings of Jakob Nielsen, Jared Spool, and Jeffrey Rubin.<sup>8</sup> Battleson, Booth, and Weintrop draw on both Nielsen's *Usability Engineering* and Rubin's *The Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests* as influences for their usability test at the University of Buffalo, and Augustine and Greene used Rubin's book to guide the deployment of their usability tests.<sup>9</sup> Dickstein and Mills employed usability testing in the redesign of their library website, and Brantley, Armstrong, and Lewis endorse Rubin's book as "offering practical and comprehensive instructions for the usability testing process," suggesting that "the numerous usability studies measuring the effectiveness of library Web sites provide templates that researchers can use as models."<sup>10</sup> In her survey on user-centered design for library professionals, Anna Noakes Schulze suggests information professionals should turn to Nielsen's *Usability Engineering* to learn "good usability," which has "often been lacking" in information systems designed by information professionals aspiring to incorporate user-centered design.<sup>11</sup>

Libraries have employed additional UCD methods, including heuristic evaluation, card sorting and focus groups. For example, Ferreira and Pithan coupled the usability methods of Jakob Nielsen with a constructivist model of user study created by Carol Kuhlthau.<sup>12</sup> When redesigning the Carnegie Mellon University Libraries website, George described a number of different user-centered methods used to enhance usability of their website, including Nielsen's heuristic evaluation, which involves expert users navigating and critiquing interfaces.<sup>13</sup> This technique was also used by Dickstein and Mills, as well as by Manzari and Trinidad-Christensen, who noted its rarity, stating, "[u]sability principles have been applied to library Web-site design; however usability studies often do not include the additional heuristic evaluation recommended by Nielsen."<sup>14</sup>

Covey found that the most commonly proposed design changes stemming from usability testing include placement of links, page layout, online help options, and changing the labeling and vocabulary.<sup>15</sup> Brantley, Armstrong, and Lewis noted a user "preference for visual cues such as buttons or icons rather than textual explanations," and provide suggestions for "space-saving techniques and improvements to the layout of the sidebar, the services options, and the list of resource options on the customized pages."<sup>16</sup> Largely cosmetic, these solutions point to a flaw in relying too heavily on methods that incorporate user needs after design completion. Goodwin points out that "usability testing can't make up for a good design methodology," adding that "it's much more effective to do research up front, than to follow a methodology that helps translate your findings into a good design."<sup>17</sup>

By nature, this upfront research must include knowledge of the user populations that a library's interfaces are serving. The library literature reflects a focus on measuring the capacity of an interface to meet its intended purpose rather than understanding the user populations for which that interface is being designed and deployed. Aaron Schmidt and Amanda Etches note, "(u)nderstanding library users is an essential component of creating a user-centered website"; however, they don't acknowledge the necessity of testing for the specific audience(s), asserting, "as long as they're not librarians it pretty much doesn't matter who they are."<sup>18</sup> Laura Hudson, in contrast, states, "negotiating usability can be difficult" without defining target user groups, since "[l]ibraries serve diverse groups of patrons with various needs."<sup>19</sup> In their review of the University of Buffalo Libraries website, Battleson, Booth, and Weintrop acknowledge that users of the website "comprise a very heterogeneous population," and, because support for all of these user groups "was neither practical nor feasible," the primary users were defined as "undergraduate students with little or no experience using the libraries' site."<sup>20</sup> This decision to design for a single population has been identified in the UCD community as a sound design principle, with Cooper arguing, "you will have far greater success by designing for one single person" than you would trying to design for all constituents, and pleasing none.<sup>21</sup>

Personas have the potential to improve design efforts in academic libraries but require more research and investigation. Though in the last few years there has been a noticeable uptake in the use of personas in libraries, efforts to incorporate them have limited representation in the published literature. In 2010, Koltay and Tancheva noted "using personas is a relatively new development and, to our knowledge, has only rarely been applied to an academic library setting."<sup>22</sup> A survey of the library literature discussing personas corroborates this shortage. In 2005, Heather Cunningham wrote about personas she created by synthesizing existing qualitative data in the form of surveys with usability testing by undergraduates in one discipline.<sup>23</sup> For the National Archives, Donald Phillips used diary studies as his ethnographic source for creating personas.<sup>24</sup> In 2007, the Cornell University Libraries (CUL) undertook a web-visioning process, employing personas to "provide insight and communicate the various research practices and processes used by the primary clients in the library." CUL created personas to "formulate our audience's needs and expectations" and serve as a "decision-making tool."<sup>25</sup> To understand the needs and goals of institutional repository users at the University of Colorado, Boulder, 20 graduate students and faculty members were interviewed and four personas were synthesized from these data.<sup>26</sup> At North Carolina State University, personas were created for a website redesign project and "helped keep everyone on the larger team focused on the end user."<sup>27</sup> In a case study, Summerville and Brar discuss their creation of personas to influence the design of their digital library interfaces, and Lage, Losoff, and Maness outline their creation of personas to test the feasibility of library involvement in data curation.<sup>28</sup>

Though an awareness of the benefits of personas is emerging in libraries, more research into their creation process is needed, especially noting that they have been criticized for their subjectivity and lack of scientific rigor. Cooper has acknowledged this criticism, noting, "it is possible to build personas without a lot of research but what happens is the confidence you can have in the validity of those personas shrinks dramatically."<sup>29</sup> Phillips concurs, noting, "personas must be created from behavior that was observed from actual users and not from the dreaded assumptions and rumors closet."<sup>30</sup> The biggest challenge, note Koltay and Tancheva, is "that they are often based on a sample that is not statistically significant."<sup>31</sup> The library literature reflects a number of techniques used to create personas, including analyzing existing qualitative

data, usability testing, diary studies, and ethnographic interviews (the most common approach used in the design field).

With a plethora of service-generated data, libraries are in an enviable position to create sound data-driven personas, though more exploration of methods for data synthesis is needed. To contribute to this new area of research, we chose to analyze existing service-generated data in the form of AAL chat transcripts, to build personas that reflect the needs, goals, and values of our users.

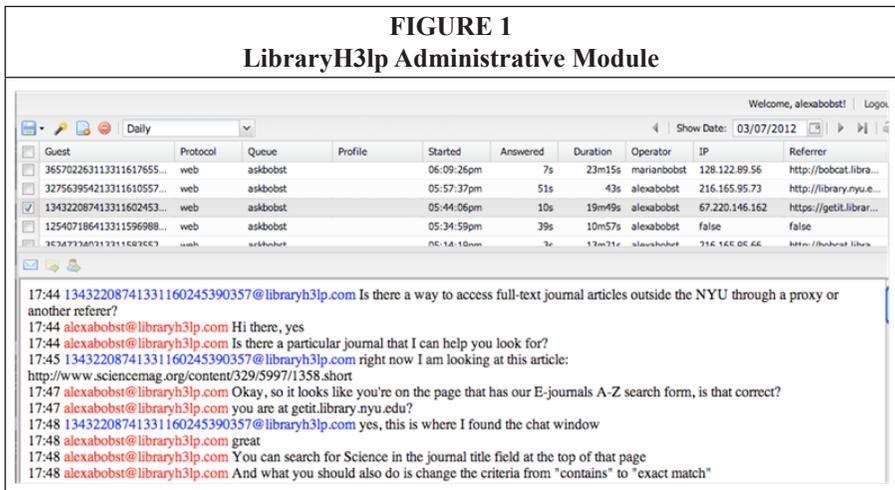
Methodology Part 1: About Our Data

NYU Libraries provides virtual reference services to its users via e-mail, text (SMS), and chat (IM), using two major platforms. The e-mail service runs through OCLC's QuestionPoint platform, while the text and chat services both run through LibraryH3lp. The chat service is the busiest of the three modes of communication, generating approximately 1,700 transactions per month. It is accessible across the libraries website through multiple discovery interfaces and research guides, serving users at all of the university's global sites. With coverage approaching 24 hours per day, the chat service reaches a wide and diverse subset of the NYU Libraries user community.

The LibraryH3lp platform allows us to gather an extensive amount of data about all chat transactions, as shown in figure 1. These data include elements that provide insight into user circumstances, such as IP address and referring URL (in other words, the exact location on the library website from which someone initiated his or her chat). LibraryH3lp also collects complete transcripts that can be downloaded from the platform's administrative module. To build the personas, we used the descriptive data that are collected by LibraryH3lp, as well as qualitative and quantitative data that we produced by coding a random, anonymized sample of 170 transcripts. The sample was drawn from transcripts recorded between November 2011 and February 2012. After removing two transcripts that contained insufficient information, our final sample size was 168.

Methodology Part 2: Developing the Coding Instrument

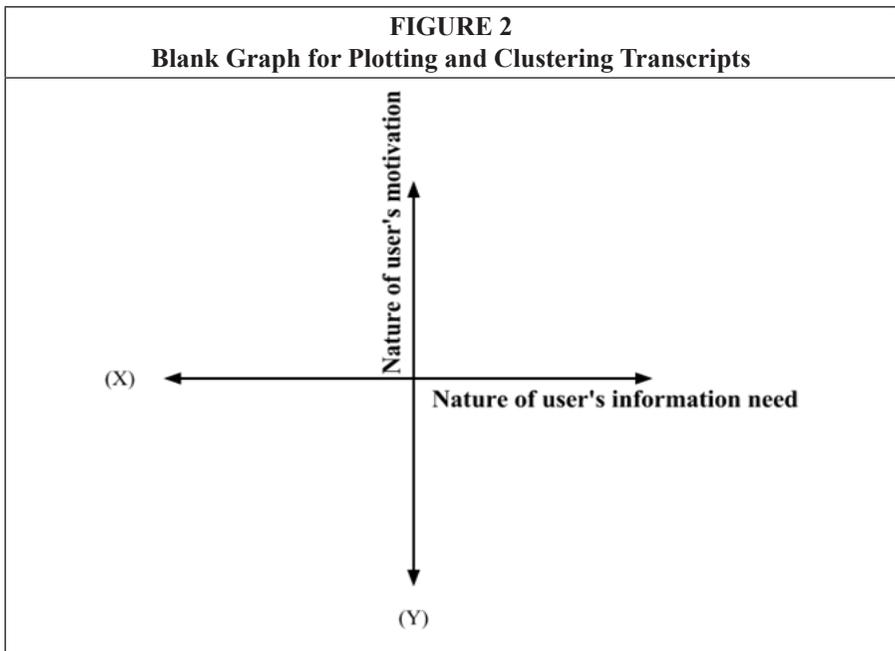
We developed our coding instrument with two goals in mind. The first was to produce a tool enabling us to plot and cluster the transcripts on an X/Y graph. Accordingly, we designed the instrument to produce two numeric values for each chat conversation, one serving as the X coordinate and the other serving as the Y coordinate. The second

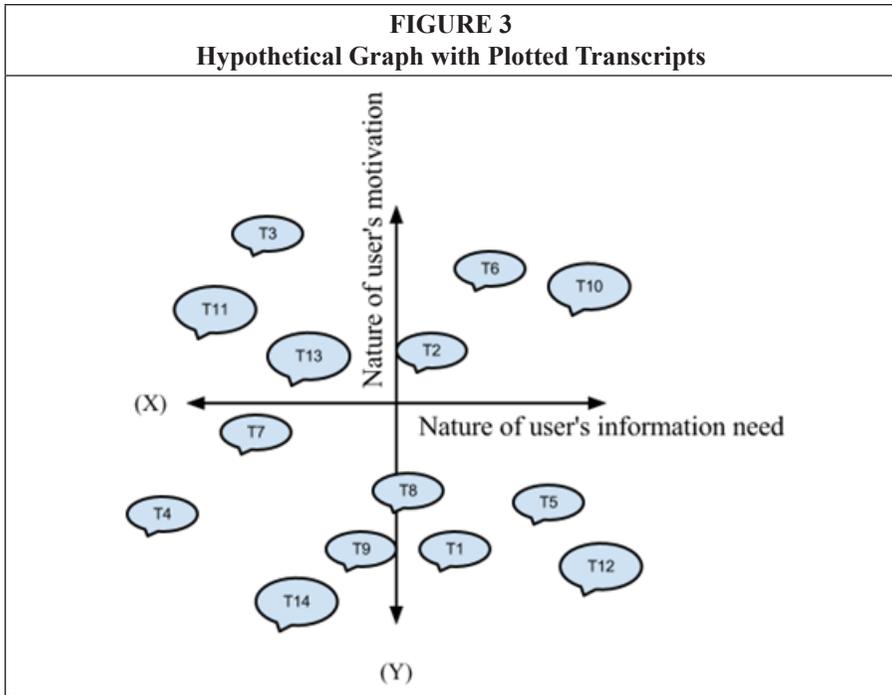


goal was to thematically define the X and Y axes to represent user goals and values. We defined X to represent the nature of the user's information need, while Y would represent the nature of the user's motivation. On the X axis, values could range from very discovery- or content-oriented at the negative end to very delivery- or access-oriented at the positive end. On the Y axis, values could range from very intrinsic at the negative end to very extrinsic at the positive end. Figure 2, below, depicts a blank graph, with these two axes defined and labeled. Figure 3, below, depicts the same graph with hypothetical points plotted, representing coded transcripts from our sample.

To assign meaningful X and Y values to each transcript, we developed a series of questions to help us discern the nature of users' information needs and their levels of motivation. The complete coding instrument is included as Appendix A. We refined these questions through several rounds of iterative testing. Mindful of our level of interrater agreement, we developed a narrative rationale for each question, in which we explained how and why it would allow us to characterize a conversation in a particular way. Appendix B contains the coding rationale. We gathered two preliminary samples of 10 transcripts each to test working drafts of our coding scheme. After coding the first set of 10 transcripts, we examined our results for variation and found that we were in agreement for roughly 50 percent of questions. After discussing our reasoning and revising the wording of the questions, we tested with a second set of 10 transcripts. We improved in agreement considerably after the second test, with matching responses in 70 percent of total questions. Following a final round of language and rationale revisions, we divided our sample of 168 transcripts between us and coded them individually using Qualtrics Survey Software.

The final coding scheme included a total of 8 questions, 6 of which were answerable in a way that would result in numeric X and Y values. Two questions dealt solely with the X axis, with possible values ranging from -2 to +2, and two questions dealt solely with the Y axis, with possible values ranging from -3 to +3. The two questions that did not have numeric values directly assigned to their answers were those recording





the unique transcript identifier and whether a question was reference or directional in nature. If we classified a conversation as reference, we were prompted to select from a list of 6 additional criteria, valued at one point each, positive or negative depending on their nature. These criteria were elements that may have been present in transactions and, if so, would contribute points to the final X or Y value. Five of the 6 possible reference criteria would count toward the X value and 3 would count toward the Y value. If we classified a conversation as being of a directional nature, the survey prompted us with a list of 3 additional criteria, also one point each, all of which were applicable on the X axis and 1 of which was applicable on the Y axis. For these questions, we selected all applicable criteria that may have been present in transactions (for example, a need for subject-specific help, meaningful policy advisement, or research strategies and recommendations). All transactions classified as reference received a default negative point toward the intrinsic end of the Y axis to represent the motivation required of a patron to initiate a conversation with a reference librarian.

While an equal number of questions in the survey could be applied to both the X and Y axes, the coding instrument contained more criteria overall that could potentially describe the characteristic plotted on X (that is, the nature of the information need). While we would have liked to give each characteristic equal consideration, we determined that we could more soundly observe the nature of an information need from a discrete conversation than we could discern a user's level of motivation. While a larger number of points contributed to the X value, the questions that applied to the Y axis lent themselves to greater variation in values (that is, they were answerable with values of up to  $\pm 3$ ), while the greatest range that we could set up for any of the X axis questions was  $\pm 2$ . The nature of a user's motivation was not only more challenging to discern than the nature of the information need, but also more nuanced.

### **Methodology Part 3: Plotting and k-Means Cluster Analysis**

After coding the transcripts, we exported the survey data out of Qualtrics and into SPSS in .csv format to calculate averages for the X and Y values assigned to each transcript. We then used SPSS to plot the transcripts on an X/Y graph before running a specific type of cluster analysis known as k-means. In a k-means cluster analysis, a data set of  $n$  objects is partitioned into  $k$  clusters, also known as Voronoi cells. The number of cells is set prior to the analysis, which divides all data points into regions. All data points in a particular region must be closer to that region's center than to any other regional center. Once the data is partitioned, the cells should exhibit consistent mean distances between the individual regional centers and all the other points in the region. In our case, there were 168 objects that we partitioned first into three, then into four clusters.

### **Methodology Part 4: Enhancing the Base Clusters with Additional Data**

As described below in our results section, the k-means cluster analysis was successful both times we ran it. Following our analysis of both sets of clusters, we decided to use the set of four as the basis for our personas. We then referred to several secondary sources of qualitative data to test the characteristics of the clusters against existing data used to understand user needs. These secondary sources of data included qualitative feedback from the NYU Libraries' 2011 LibQual and LibQual Lite survey results, excerpts and comments from a series of faculty interviews on topics related to digital scholarship, select results of environmental scanning work that the Libraries conducted prior to its most recent round of strategic planning, a report on demographic trends from the libraries' assessment team, and results from a selection of our numerous formal usability tests for a variety of interfaces. These data sources provided useful criteria such as user group affiliation and disciplinary needs, along with specific quotes and comments from library users. We used these sources to inform the development of characteristics and attributes that we could appropriately assign to the clusters in the graph.

## **Results**

### *Cluster Analysis*

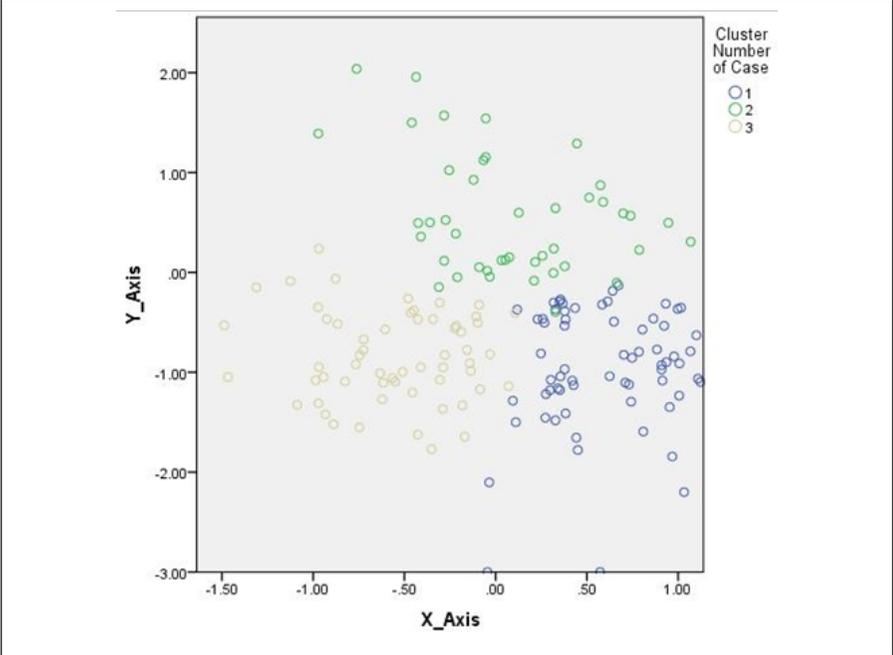
Both rounds of k-means cluster analysis in SPSS were successful. Figures 4 and 5, below, depict the initial results of the analyses, with 3 and 4 clusters, respectively. To better understand and interpret the makeup of each cluster, we drew in X and Y axes over the plotted transcript values, as shown in figures 6 and 7.

The axes helped us visualize our clusters in relation to the graph's quadrants, labeled alphabetically as A, B, C, and D. We also referred back to our SPSS data in spreadsheet form, where we were able to sort the list of transcripts by numeric values and fill in quadrant locations. Within the set of three clusters, two were very closely aligned with individual quadrants on the graph while the third cluster was extremely varied, with presence in all four quadrants. By comparison, the set of four clusters was generally more localized. While all of the clusters straddled boundaries between quadrants, none were extremely varied or extremely homogenous. Rather, they appeared representative of recognizable and realistic levels of complexity and variation in our users' needs. Therefore, we decided that the characteristics we could attribute to the set of four would more accurately capture and depict members of the NYU Libraries user community.

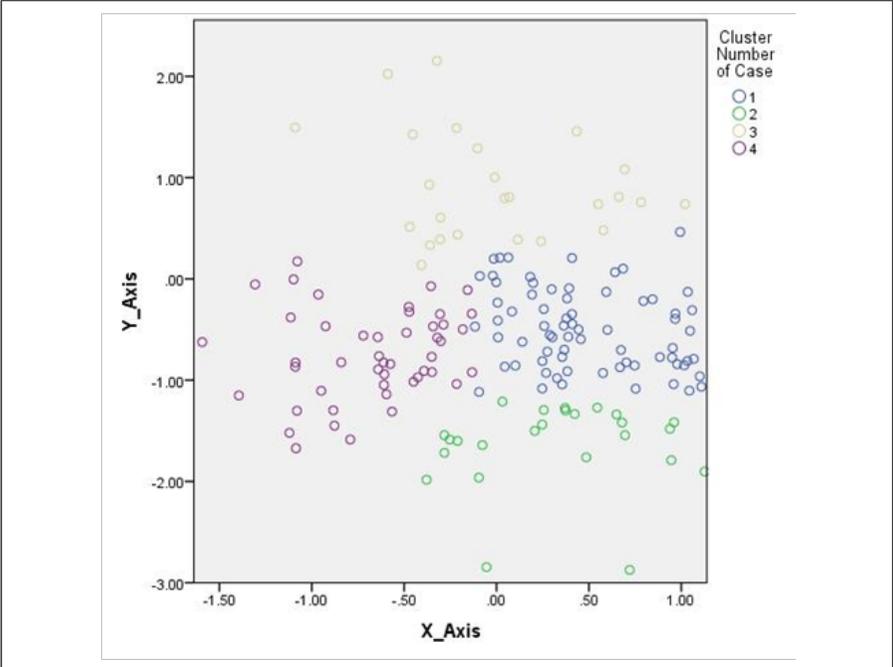
### *Cluster Breakdowns*

Cluster 1, the largest of the four clusters, includes 71 plotted transactions. This equates to 42 percent of the sample. Of the 71 transactions represented in this cluster, 47, or just over two-thirds, fall in Quadrant B, toward the positive, or delivery-focused, end of

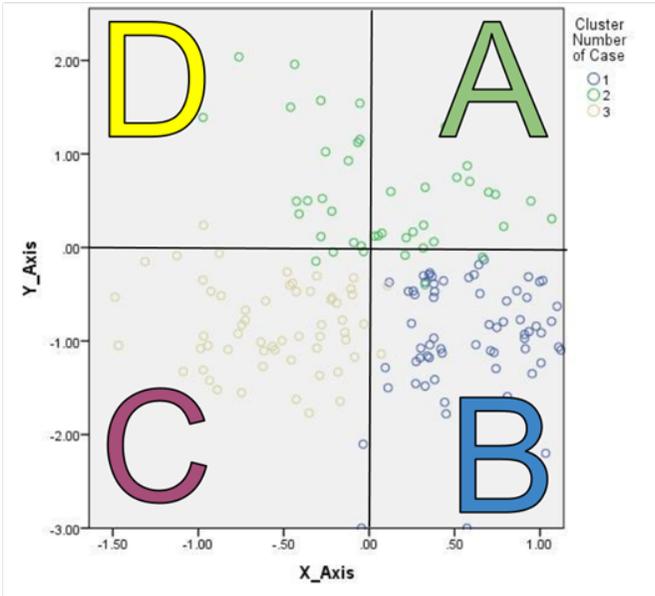
**FIGURE 4**  
**Results of Cluster Analysis for 3 Clusters**



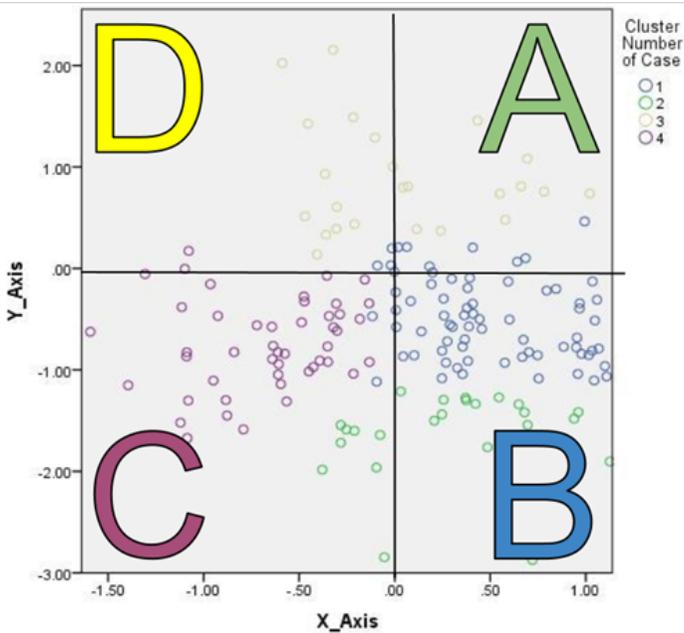
**FIGURE 5**  
**Results of Cluster Analysis for 4 Clusters**



**FIGURE 6**  
**View of 3 Clusters with Axes and Labeled Quadrants**

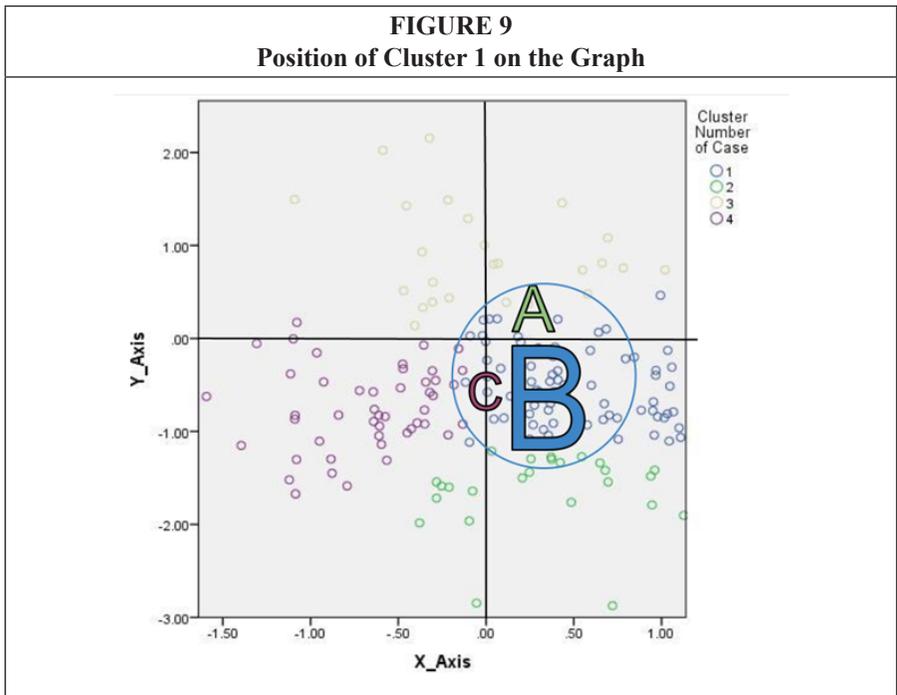
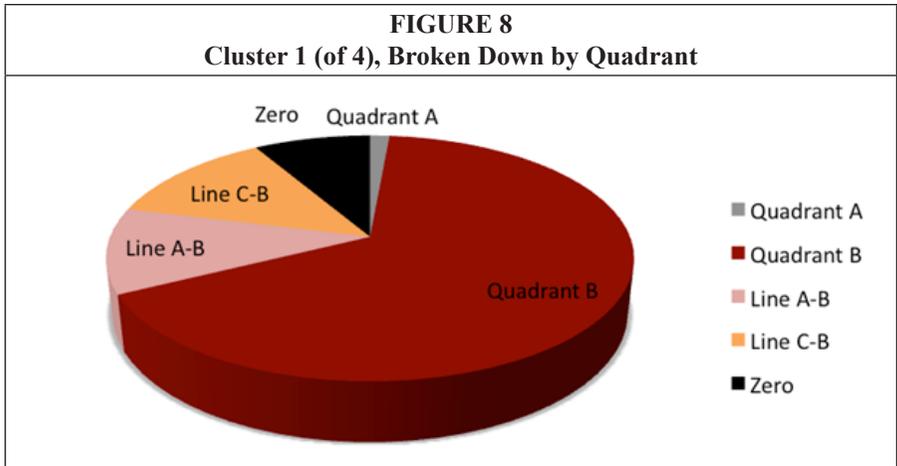


**FIGURE 7**  
**View of 4 Clusters with Axes and Labeled Quadrants**



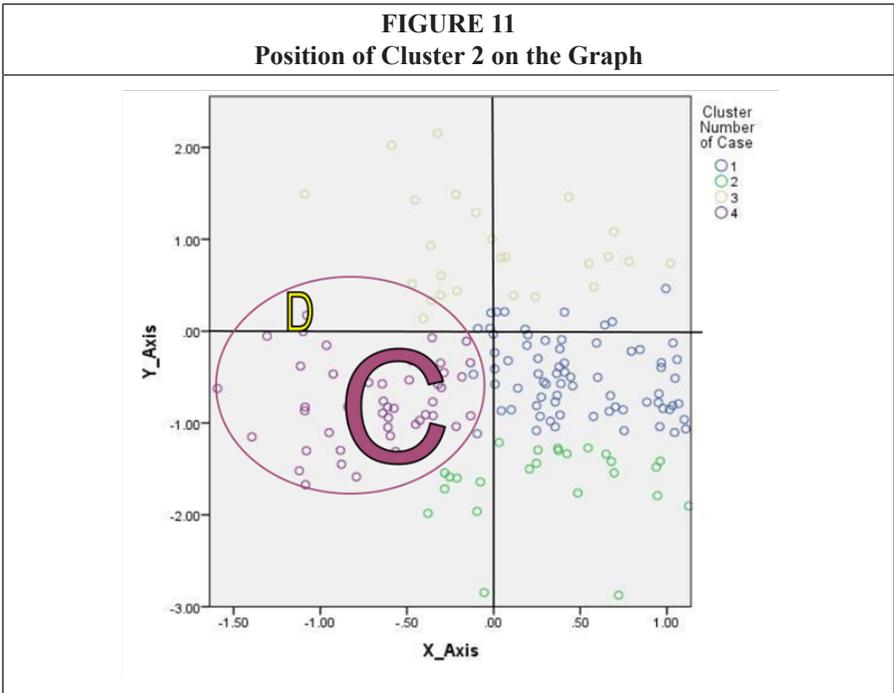
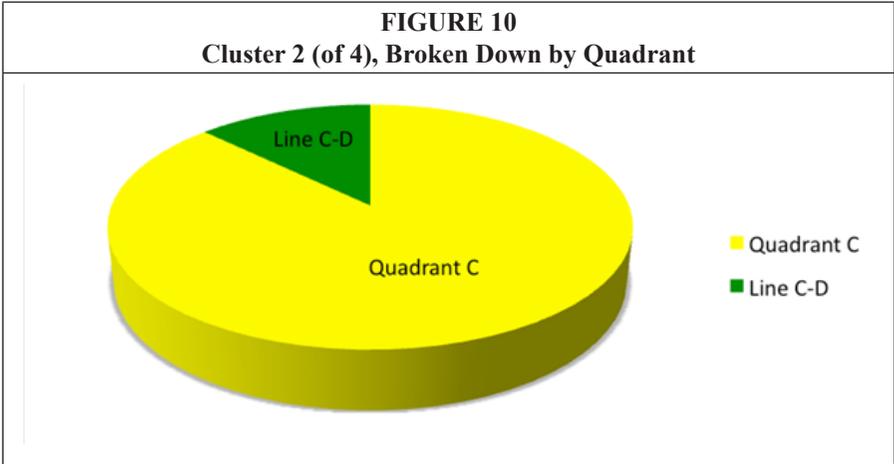
the X axis and the negative, or intrinsic, end of the Y axis. Eight transcripts fell on the line between Quadrants A and B, while nine fell on the line between Quadrants B and C. Six of the transcripts were at zero. Figure 8 represents the breakdown of Cluster 1 by quadrant and figure 9 illustrates the position of Cluster 1 on the graph.

Cluster 2, the next largest in the set of four, includes 47 plotted transcripts, accounting for 28 percent of the sample. With 41 of the 47 transcripts falling in Quadrant C, this cluster is very homogenous and comes close to being coterminous with Quadrant C. The remaining transcripts fall on the boundary between Quadrants C and D. As a whole, the cluster tends toward the intrinsic end of the Y axis and the discovery-oriented

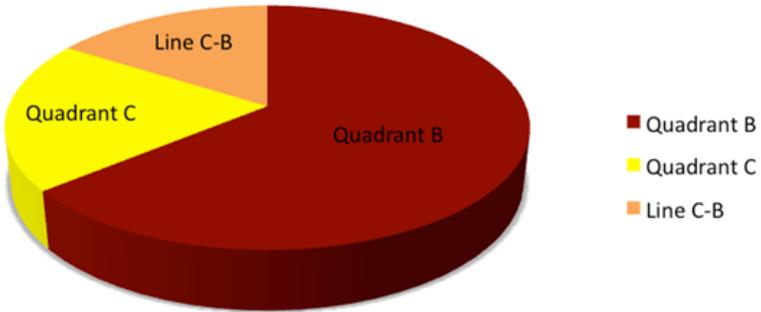


end of the X axis. Figure 10 represents the breakdown of Cluster 2 by quadrant and figure 11 illustrates the position of Cluster 2 on the graph.

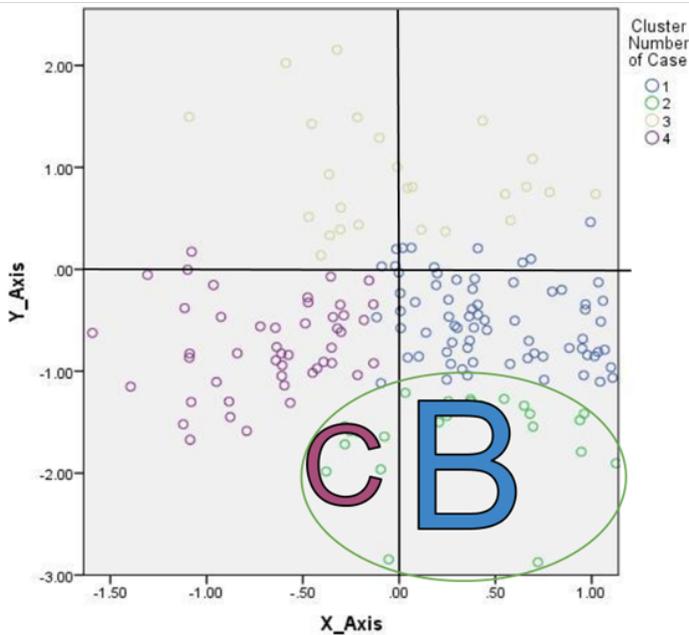
Cluster 3 includes 25 transcripts and accounts for 15 percent of the sample. Fully 16 of these, or 64 percent of the cluster, fall in Quadrant B, with the remainder in Quadrant C or on the line between B and C. This cluster sits toward the negative, or intrinsic, end of the Y axis, with close to two-thirds of its transcripts falling near the positive, or delivery-oriented, end of the X axis. Figure 12 represents the breakdown of Cluster 3 by quadrant and figure 13 illustrates the position of Cluster 3 on the graph.



**FIGURE 12**  
**Cluster 3 (of 4), Broken Down by Quadrant**

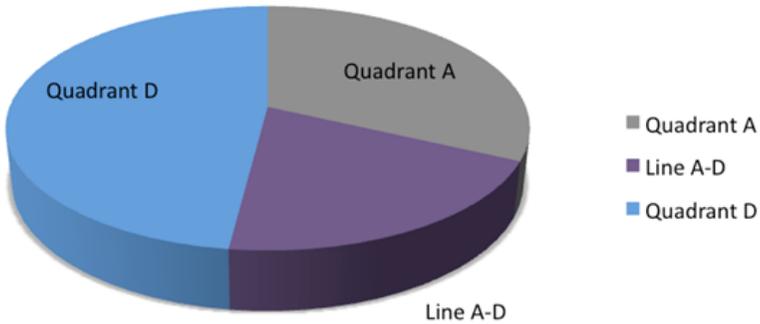


**FIGURE 13**  
**Position of Cluster 3 on the Graph**

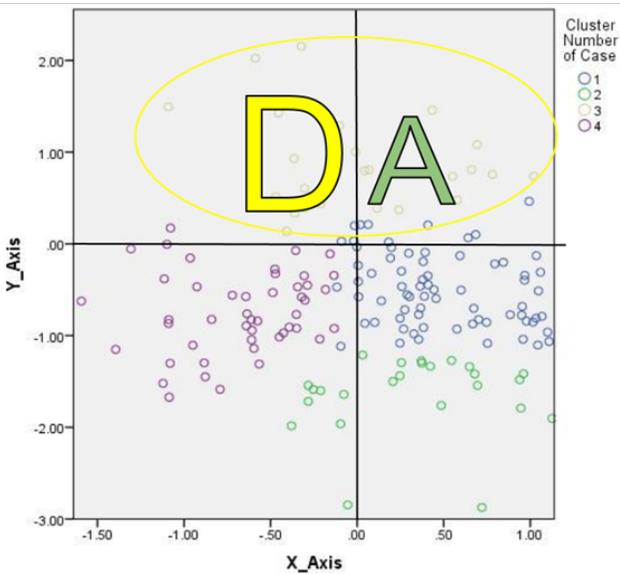


Cluster 4 also comprises 15 percent of the sample, with 25 transcripts. This cluster is situated a little more in Quadrant D than in A, but is split almost evenly between the two. Eight transcripts are in A and 12 are in D, with the remaining points on the line between A and D. As a whole, the cluster is situated toward the extrinsic end of the Y axis, with points spanning the length of the X axis. Figure 14 represents the breakdown of Cluster 4 by quadrant and figure 15 illustrates the position of Cluster 4 on the graph.

**FIGURE 14**  
**Cluster 4 (of 4), Broken Down by Quadrant**



**FIGURE 15**  
**Position of Cluster 4 on the Graph**



**Discussion**

*Understanding the Quadrants and the Clusters*

By plotting each AAL chat transcript on an X/Y graph, we were able to understand them as falling within one of four quadrants, based on the characteristics represented by the two axes. While each plotted point falls within a quadrant or on a line, the resulting clusters span multiple quadrants. That is to say, our data do not cluster in a way that allows us to build homogenous user types based on the four quadrants of our graph. The k-means method instead guides us to form user types whose goals and motivations demonstrate some complexity and nuance in relation to the two themes

that the axes represent, the nature of the user's information need and the nature of the user's motivation.

Any point that was plotted above zero on the Y axis, in Quadrant A or Quadrant D, represented a conversation in which a user's level of motivation was discernibly extrinsic. For example, if a user mentioned that he or she was assigned to complete a specific homework task or was working on behalf of someone else, the conversation would have received positive points toward its overall Y value. Additionally, if a user resisted instruction, expressed a sense of immediacy, or demonstrated that he or she was engaged in a discrete, rather than long-term, project, we were likely to have given a positive Y value to the transcript. As mentioned above, discerning motivation proved to be more challenging than determining the nature of the information need. In cases where motivation was indiscernible, a transcript's Y value was zero.

Any point that was plotted below zero on the Y axis, in Quadrant B or Quadrant C, represented a conversation in which the user's level of motivation was discernibly intrinsic. For example, if a user engaged in a substantial conversation about his or her research process and demonstrated curiosity or understanding about how and why library tools and services factored in, we tended to give the conversation negative points toward its overall Y value. Additionally, we gave a negative Y point to any reference question to acknowledge the level of motivation required for a user to initiate a reference transaction. Accordingly, the sample as a whole is situated relatively low on the Y axis.

Points that were plotted in quadrants A or B, to the right of zero on the X axis, represent conversations in which users needed help accessing content. For example, if users were engaged in known-item searches or wanted to verify journal or database subscriptions, we gave the conversation positive points toward its X value. We also tended to give positive X values to conversations about site navigation or that included any troubleshooting components, generally characterizing these information needs as mechanical, or delivery based, in nature.

Points that were plotted to the left of zero on the X axis, in quadrants C or D, represent transactions in which users needed help finding, identifying, and evaluating content. In general, we characterized these needs as topical or discovery-oriented and gave negative X values to their occurrences. These included things like users asking for help selecting an appropriate database for research in a specified subject area, discussing the terms of research paper assignments, or asking for feedback on the suitability of particular resources. Additionally, if users engaged in tandem searching and evaluating with librarians, we tended to give those conversations negative X values.

With attention to the individual quadrants, we see a further degree of specificity, and are able to characterize the conversations from our sample in terms of need as well as motivation. For example, any point in Quadrant B represents a conversation in which an intrinsically motivated user required help with access to specific content. As mentioned above, the four clusters that we produced through our k-means analysis are not coterminous with the four quadrants on the graph. However, the quadrants provide a useful mechanism for characterizing and interpreting the clusters, as described below.

## **Translating the Clusters into Personas**

### *Cluster 1*

This cluster is located mostly below zero on the Y axis, in territory characterized by intrinsic motivation and access-related needs. However, the cluster remains fairly close to the X axis and has some representation in Quadrant A. Thinking of this cluster as a composite user type, we might say that it represents someone who is moderately invested in his research or other academic work. He uses library tools and resources for projects that reflect his real interests and that he has some ability to shape, but these

projects likely do not amount to career-defining scholarship. While he occasionally requests assistance with discovery, as indicated by the portion of the cluster located in Quadrant C, he is much more likely to approach library staff for assistance with delivery. We don't know from these data alone if this user type relies on library interfaces for discovery of content, but it may be the case that he relies on library tools much more heavily in the latter stages of his research process.

### *Cluster 2*

Similar to Cluster 1, Cluster 2 is situated mostly below zero on the Y axis, in territory that may be characterized as moderately intrinsic, regarding motivation. Similar to the user type derived from Cluster 1, the Cluster 2 user is engaged in projects or research activities that she cares about and has some ability to shape. In contrast to the first cluster, she represents users who ask for discovery-related help (that is, finding, identifying, and evaluating content). While this user type may also ask for and receive assistance with delivery, we know that she initiates her engagement with library tools and services during the discovery stages of her research process.

### *Cluster 3*

Cluster 3 is similar to Cluster 1, with about two-thirds of its transcripts falling in Quadrant B, but it has no presence in A and is situated much closer to the negative end of the Y axis. As the only cluster in the set of 4 that is located entirely below the X axis, we can see Cluster 3 as a type that represents our most intrinsically motivated users. His use of library tools is likely to be connected to projects that are self-directed or entail close collaboration. His research activities are likely very scholarly in nature or may be related to career goals or long-term personal interests. His needs are largely, though not exclusively, mechanical, representing users who need help connecting final dots more often than they need help with getting started or identifying useful resources. But like the Cluster 1 user, his presence in Quadrant C is not insignificant, serving as a reminder that many users in our community need help at all stages, even if they are more likely to initiate reference transactions for delivery help than they are for discovery help.

### *Cluster 4*

In opposition to Cluster 3, Cluster 4 is the only grouping located entirely above zero on the Y axis and is therefore representative of our most extrinsically motivated users: that is, those who are using library tools and resources as means to assigned ends, without a larger sense of engagement or personal investment in their work. With its near-even split between Quadrants A and D, this cluster represents users who need help with content as well as access and navigation. She incorporates library tools and resources at all stages of her work and may not necessarily understand the substantive differences between discovery and delivery stages. Her goals tend to be limited to completing her task or tasks at hand, which are highly unlikely to be connected to any deeper long-term or scholarly engagement.

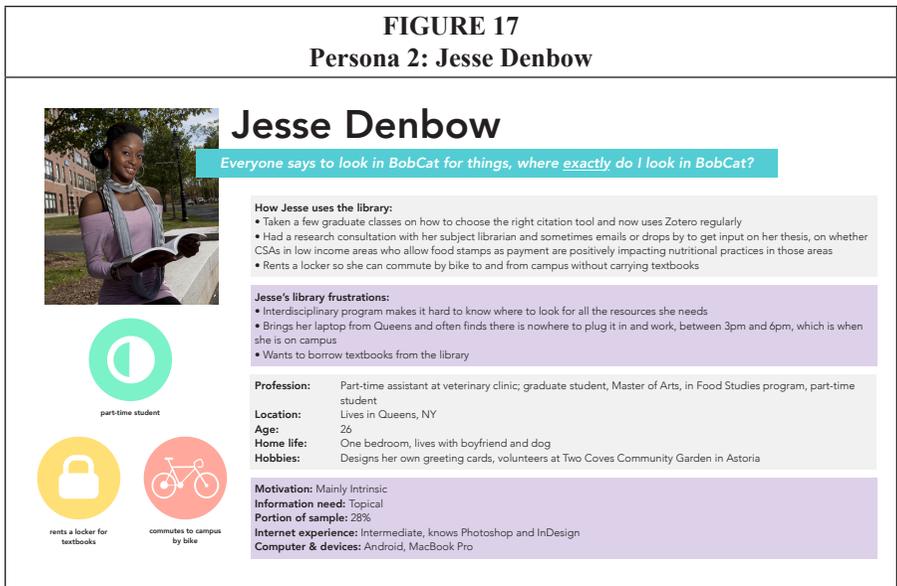
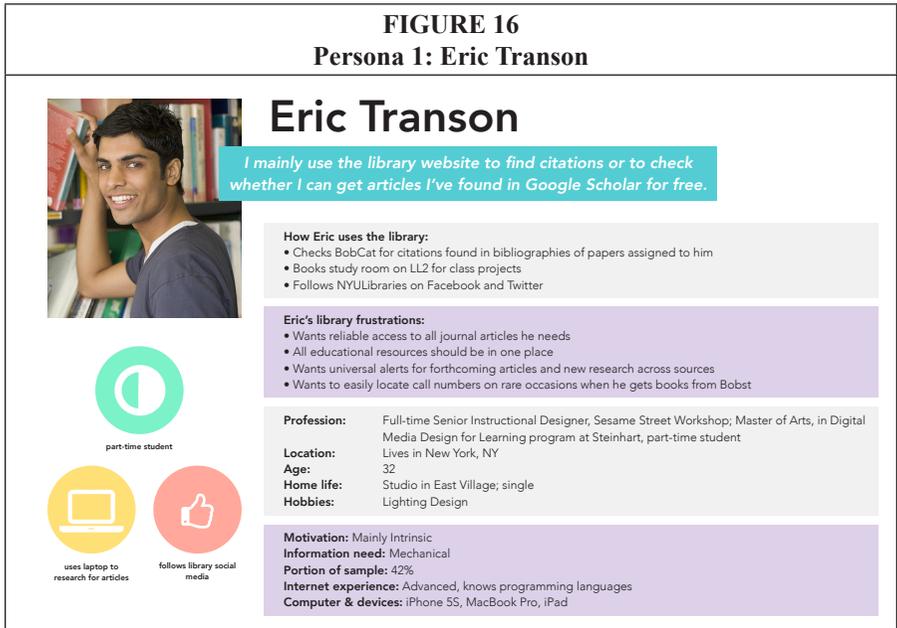
### *Completed Personas*

The preliminary personas described above represent groupings found in our initial data set, based on similar levels of motivation and information needs, as characterized by the two axes on our graph. While these go a long way toward helping us define representative user types, they still required additional enhancements to fully humanize them. Accordingly, the last stage of our process entailed incorporating qualitative data from recent outreach and assessment activities, as detailed above. From these

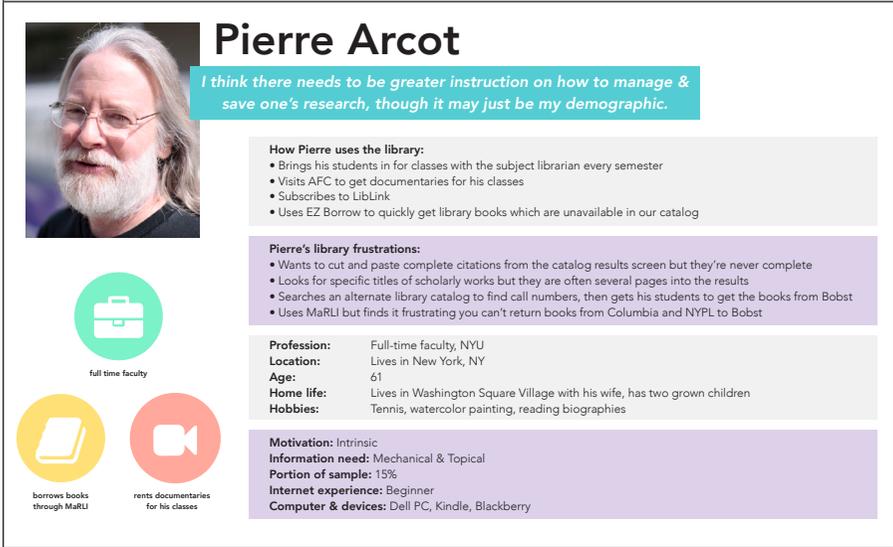
data sources, we were able to extract quotes and concerns expressed by real users and match them with the groupings above. The final synthesized personas are depicted in figures 16, 17, 18, and 19 below and represent four types of users reflecting the needs, goals, and values of the NYU Libraries community.

**Limitations**

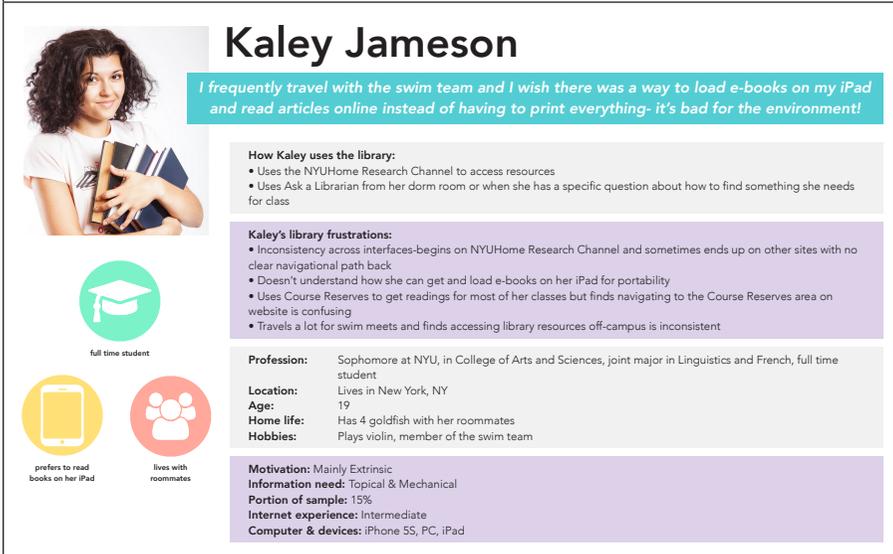
Several aspects of this study have not been tested elsewhere or have not been tested in the manner described here and merit additional discussion and investigation.



**FIGURE 18**  
**Persona 3: Pierre Arcot**



**FIGURE 19**  
**Persona 4: Kaley Jameson**



First, the task of measuring user motivation presented a challenge, not least because it is not well represented in previous studies. By contrast, the user's information need is a more tested and established theme in the library literature. Further research is needed in developing methods for observing and understanding motivation.

Similarly, the process used for mapping the secondary qualitative data to the clusters was not based on any formal or previously tested method. The additional qualitative

data used to finalize the personas ensured that these fictional characters were given traits and attributes reflective of members of our user community.

In determining our sample size, we gathered transcripts from busy periods during the fall and spring semesters of the most recent academic year. We did not attempt to place any other chronological parameters around the sample, though we do acknowledge that life cycle and seasonality may play a role in the construction and utility of personas. Investigation of this potential role would be best addressed in a separate study.

The sample size employed is comparable to those used in published transcript analyses in the library literature. However, this study is experimental and a departure from transcript analyses primarily intended to analyze and assess reference services. We found no clear formula for defining an appropriate sample size for a pilot study such as this one.

Our decision to use Ask a Librarian transcripts as the primary source of data invites potential criticism related to response bias. Namely, the sample represents library users who voluntarily initiate reference transactions. It necessarily excludes other library users who may not ask for help or be aware of the chat service. For these reasons, we gave an additional intrinsic point to every transcript analyzed to acknowledge the level of motivation required in initiating a conversation. This study, focused on developing a methodology for creating personas from service-generated data, did not attempt to create personas for the entire potential user community. Personas that aim to represent all potential users and even nonusers would be best investigated in further studies.

Finally, this study does not assess the use or application of these personas in any particular design project. However, NYU Libraries is currently incorporating personas in multiple projects, including the large scale re-envisioning of the libraries website, and the evaluation of web scale discovery tools.

## Conclusion

The final personas portray significant diversity in the needs, goals, and behaviors of the NYU Libraries user community. This project reaffirms that such a level of diversity, as well as nuance, can be expected even within a defined subset of the user community (namely, Ask a Librarian chat service users). Arguably, our decision to use data from the chat service as the foundation for these personas also explains why we are able to discern some similarities among them, such as the tendency toward intrinsic motivation and pronounced need for delivery-related help, as indicated by two of the four clusters that are mostly located in Quadrant B. While the ability to perform this kind of pattern detection is an important outcome of the cluster analysis described above, our development of the clusters into complete personas enables us to move beyond simply reinforcing broad similarities and differences in user characteristics. By providing the clusters with names, faces, needs, and frustrations, we are able to more fully characterize, define, and humanize the users whose goals and behaviors we seek to understand and serve.

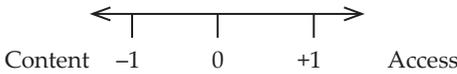
By building personas out of coded, service-generated data, we have attempted to address the critique that persona development tends to lack rigor. Rather than gathering data to fill in standard user templates (such as the "typical undergraduate") we analyzed and described the patterns in our primary set of data before assigning traits and user statuses, which we mapped against additional data sources. It is certainly the case that our decisions to describe the clusters as faculty, graduate students, or undergraduates likely reflect prevailing assumptions and mental models about those categories, but our data-driven approach has tilted the balance toward a grounded theory model for persona creation.

# Appendix A: Coding Instrument

## 1. Transcript number

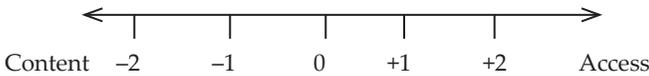
### 2. Referring URL (X-axis)

- -1: User is on search page but has not conducted a search
- 0: User is not on a search page or URL unknown
- +1: User has done a search



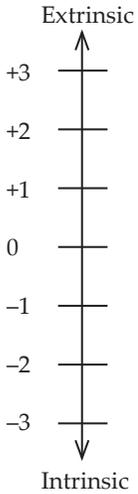
### 3. Familiarity with resources and services (X-axis)

- -2 User doesn't know where to begin; needs help getting oriented as well as content-based recommendations
- -1 User has a sense of how to engage with the site but requires assistance discerning appropriate resources
- 0 User's level of familiarity with resources and services is not discernible
- +1 User demonstrates awareness of library catalog and well-known databases
- +2 User demonstrates advanced awareness/facility with our tools (cited reference analysis; bibliographic management tools; dissertations, etc)



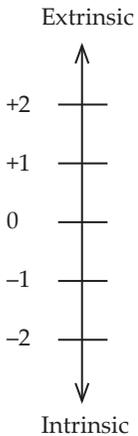
### 4. Capability of accessing and understanding resources and/or willingness to engage with process (Y-axis)

- -3: User asks for or enthusiastically accepts instruction; demonstrates strong desire to understand how tools/resources work
- -2: User accepts instruction willingly or gratefully but didn't seek it out explicitly (or know that it was an option)
- -1: User accepts instruction but with little indication that he/she will make use of new knowledge in the future
- 0: User's attitude toward instruction/knowing how things work is not discernible
- +1: User accepts instruction as a means to an end, possibly with reluctance
- +2: User demonstrates some reluctance to engage in research process or learn how tools work; may exhibit impatience
- +3: User asks for librarian to do research or gather documents on his/her behalf; demonstrates no interest in understanding tools or process



**5. Origin of user's information needs, if discernible (Y axis)**

- **-2:** self-directed, comes from within; user demonstrates strong degree of interest in project with no indication that it is connected to an assignment or task or to goals related to p & t
- **-1:** project is self-directed to a degree; user is very engaged but work is related to standard task like thesis, dissertation, or p & t goals
- **0:** Origin is not discernible
- **+1:** project has been assigned by someone else (boss/prof/etc) but user shows mild interest in process and/or subject/topic
- **+2:** project has been assigned by someone else (boss/prof/etc) and user makes clear that he/she is doing it because he/she has to



**6. Reference or directional**

- Reference
- Directional

**7. Answer if Reference is selected**

If Reference

- Research strategies and recommendations (-1 on X axis)
- Research strategies and recommendations (-1 on Y axis)
- Find known item (+1 on X axis)
- Meaningful policy advisement (-1 on Y axis)
- Meaningful policy advisement (+1 on X axis)
- Technical troubleshooting (+1 on X axis)
- Subject specific help (-1 on X axis)
- Default point toward intrinsic (-1 on Y axis)

**8. Answer if Directional is selected**

If Directional

- Physical space/library as place (+1 on X axis)
- Quick basic question re hours/access/etc (0 on X axis)
- Quick basic question re hours/access/etc (0 on Y axis)
- Transfer/referral to person or point of service (-1 on X axis)

## Appendix B: Rationale for Coding Instrument

**1. Transcript number**

Enter transcript number and coder initials

**2. Referring URL**

**Axis:** X (Content-Access)

**Observable criteria:** User’s location on library website coupled with presence or absence of search at the time the AAL conversation is initiated.

**Assumptions:** Users who have spent time searching on their own will be likely to have access questions; that is, they will initiate reference transactions when they need assistance accessing a particular resource or set of resources. Users who have not spent time searching on their own when they ask for help are likely to need content-based assistance (for example, database recommendations or help forming search strategies).

**Exceptions:** Some users may request content-based help after working on their own; these users may be reluctant to ask for help initially or just used to finding what they need on their own. Some users who need help with access to a particular tool will not say so when they initially phrase their questions. Instead, they will present a generic need (asking, for instance, “How do I find books about art?”) when they are looking for a particular resource.

**3. Level of Familiarity with Resources and Services**

**Axis:** X (Content-Access)

**Observable criteria:** User’s awareness of and facility using library resources.

**Assumptions:** Users who are not very familiar with library resources will be likely to need content-based help, such as database recommendations or assistance forming search strategies. Users who are more advanced or experienced will likely be asking for access-related assistance to connect the dots of their research.

*Exceptions:* Some advanced users ask for content-based help, either out of learned appreciation for librarian expertise or openness/willingness to collaborate. Some users who are able to easily navigate the site may not discern differences between certain tools or resources.

#### 4. User's capability for accessing/understanding library resources & services

**Axis:** Y (Intrinsic-Extrinsic)

**Observable criteria that we are coding:** User's level of interest in learning to use and engage with library resources (that is, susceptibility to instruction).

**Assumptions:** Users who demonstrate interest and curiosity about functionality and use of library resources are likely to be more intrinsically motivated; they will be conducting research in a long-term way or for a project they are invested in and will want to increase their fluency with tools and systems. These users will want to feel confident in their engagement with library tools. Users who demonstrate less interest in knowing how library tools work or who make clear that they'd like to obtain research materials but do not want or need to know how to do it themselves are more likely to be extrinsically motivated. These users are likely trying to complete assigned tasks and are not necessarily invested in the tasks at hand.

*Exceptions:* Some users are very motivated and invested in their work but are not interested in knowing how to use library tools and resources. Some users may demonstrate interest or curiosity about the library search environment even if they are working on projects that are not of special or long-term value to them.

#### 5. Origin of user's information needs, if discernible

**Axis:** Y (Intrinsic-Extrinsic)

**Observable criteria:** Origin of the user's project: that is, self-selected and designed by the user or assigned by an external person or group.

**Assumptions:** Users who need library resources to complete assigned tasks are not likely to be very interested in "behind-the-scenes" information about how or why tools work the way they do. They will ask for help only when necessary and only to complete immediate tasks at hand. Users who are engaged in longer-term research projects or activities that they created or designed will be more inclined to enhance their understanding of how tools and resources can contribute to their success.

*Exceptions:* Some users who are working on assigned projects will be very motivated to do their best work possible and will be interested to learn details about library tools and resources. Some users who are working on projects of their own design may consider knowledge of library resources to be extraneous or unimportant.

## 6. Reference or Directional

- Choose according to these definitions, developed by the NYU Libraries Virtual Reference Services Subcommittee:
  - **Reference:** any content-based question pertaining to research within a subject area or the use of a tool or process (for instance, how to use a component of RefWorks, looking up a title in BobCat, assisting with the formulation of a search strategy in a database); meaningful advisement on policies or procedures, as opposed to simple pointing to services or forms (for example, should someone use ILL or make a trip up to NYPL, given their needs and time frame); troubleshooting technical issues (ezproxy, broken links, wrong sfx targets, and so on).
  - **Directional:** non-content-based transactions, including simple questions about services such as ILL, circulation, reserves, and the like; simple questions about navigating the physical space in Bobst (such as, matching a call number with a floor, locating a restroom, checking out a laptop, library hours); intake of complaints, comments, or suggestions about the library environment

## Notes

1. Scott Walter, "Distinctive Signifiers of Excellence: Library Services and the Future of the Academic Library," *College & Research Libraries* 72, no. 1 (2011): 6–8.
2. Leanne Bowler, Sherry Koshman, Jung Sun Oh, Daqing He, Bernadette G. Callery, Geof Bowker, and Richard J. Cox, "Issues in User-Centered Design in LIS," *Library Trends* 59.4 (2011): 721–52.
3. John D. Gould, Stephen J Boies, and Jacob Ukelson, "How to Design Usable Systems," in *Handbook of Human-Computer Interaction*, eds. M. Helander, T.K. Landauer, P Prabhu (New York: North-Holland Publishing Company, 2007), 231–53.
4. Alan Cooper, *The Inmates Are Running the Asylum* (Indianapolis, Ind.: Sams), 1999.
5. Kim Guenther, "Developing Personas to Understand User Needs," *Online* Sept./Oct. (2006): 49–51.
6. Yu-Hui Chen, Carol Anne Germain, and Huahai Yang, "An Exploration into the Practices of Library Web Usability in ARL Academic Libraries," *Journal of the American Society for Information Science and Technology* 60, no. 5 (2009): 953–68.
7. Jennifer L Ward and Steve Hiller, "Usability Testing, Interface Design, and Portals," *Journal of Library Administration* 43, no. 1/2 (2005): 155–71.
8. Ruth Dickstein and Vicki Mills, "Usability Testing at the University of Arizona Library," *Information Technology and Libraries* 19, no. 3 (2000): 144–51; Brenda Battleson, Austin Booth, and Jane Weintrop, "Usability Testing of an Academic Library Web Site: A Case Study," *Journal of Academic Librarianship* 27, no. 3 (2001): 188–98; Anna Noakes Schultz, "User-Centered Design for Information Professionals," *Journal of Education for Library and Information Science* 42, no. 2 (2001): 116–22; Roslyn Rawald, "Academic Library Website Design Principles: Development of a Checklist," *Australian Academic & Research Libraries* 32, no. 2 (2001): 123–36; Susan Augustine and Courtney Greene, "Discovering How Students Search a Library Web Site: A Usability Case Study," *College & Research Libraries* 63, no. 4 (2002): 354–65; Steve Brantley, Annie Armstrong, and Krystal M. Lewis, "Usability Testing of a Customizable Library Web Portal," *College & Research Libraries* 67, no. 2 (2006): 146–63.
9. Battleson, Booth, and Weintrop, "Usability Testing of an Academic Library Web Site"; Augustine and Greene, "Discovering How Students Search a Library Web Site."
10. Dickstein and Mills, "Usability Testing at the University of Arizona Library"; Brantley, Armstrong, and Lewis, "Usability Testing of a Customizable Library Web Portal."
11. Anna Noakes Schultz, "User-Centered Design for Information Professionals," *Journal of Education for Library and Information Science* 42, no. 2 (2001): 116–22.
12. Sueli Mara Ferreira and Denise Nunes Pithan, "Usability of Digital Libraries: A Study Based on the Areas of Information Science and Human-Computer-Interaction," *OCLC Systems*

and *Services* 21, no. 4 (2005): 311–25.

13. Carole A. George, "Usability Testing and Design of a Library Website: An Iterative Approach," *OCLC Systems and Services* 21, no. 3 (2005): 167–81.

14. Dickstein and Mills, "Usability Testing at the University of Arizona Library"; Laura Manzari and Jeremiah Trinidad-Christensen, "User-Centered Design of a Web Site," 15. Brantley, Armstrong, and Lewis, "Usability Testing of a Customizable Library Web Portal."

15. Diane Covey, "Usage and Usability Assessment: Library Practices and Concerns," *University Libraries Research Paper* 61 (2002): 1–93.

16. Brantley, Armstrong, and Lewis, "Usability Testing of a Customizable Library Web Portal."

17. Matthew Klee, "Personas and Goal-Directed Design: An Interview with Kim Goodwin," *User Interface Engineering* (blog), Nov. 22, 2012 (12:22 p.m.), available online at [www.uie.com/articles/goodwin\\_interview/](http://www.uie.com/articles/goodwin_interview/) [accessed 21 August 2014].

18. Aaron Schmidt and Amanda Etches, *User Experience (UX) Design for Libraries* (London, U.K.: Facet), 2012.

19. Laura Hudson, "From Theory to (Virtual) Reality," *NetConnect* (2001): 12–15.

20. Battleson, Booth, and Weintrop, "Usability Testing of an Academic Library Web Site."

21. Cooper, *The Inmates Are Running the Asylum*.

22. Zsuzsa Koltay and Komelia Tancheva, "Personas and a User-Centered Visioning Process," *Performance Measurement and Metrics* 11, no. 2 (2010): 172–83.

23. Heather Cunningham, "Designing a Web Site for One Imaginary Persona That Reflects the Needs of Many," *Computers in Libraries* 25, no. 9 (2005): 15–19.

24. Donald Phillips, "How to Develop a User Interface That Your Users Will Really Love," *Computers in Libraries* 32, no. 7 (2012): 6–15.

25. Koltay and Tancheva, "Personas and a User-Centered Visioning Process."

26. Jack Mannes, Tomasz Miaskiewicz, and Tamara Sumner, "Using Personas to Understand the Needs and Goals of Institutional Repository Users," *DLib Magazine*, 2008, available online at [www.dlib.org/dlib/september08/maness/09maness.html](http://www.dlib.org/dlib/september08/maness/09maness.html) [accessed 21 August 2014].

27. Angela Fullington Ballard and Susan Teague-Rector, "Building a Library Web Site: Strategies for Success," *C&RL News*, 72, no. 3 (2011): 132–35, available online at <http://crln.acrl.org/content/72/3/132.full> [accessed 21 August 2014].

28. Kathryn Lage, Barbara Losoff, and Jack Mannes, "Receptivity to Library Involvement in Scientific Data Curation: A Case Study at the University of Colorado Boulder," *portal: Libraries and the Academy* 11, no. 4 (2011): 915–37.

29. Gerry Gaffney, "Personas and Outrageous Software: An Interview with Alan Cooper," *Information & Design* (blog), Nov. 22, 2012 (2:12 p.m.), available online at <http://uxpod.com/personas-and-outrageous-software-an-interview-with-alan-cooper/> [accessed 21 August 2014].

30. Phillips, "How to Develop a User Interface That Your Users Will Really Love," 6–15.

31. Koltay and Tancheva, "Personas and a User-Centered Visioning Process."