

Findings from structured questions concerning Canadian census data use and discovery, and future directions

DLI National Training, Nov. 24, 2021
Susan Mowers, Shannon Leitch, Julien Doris
March 17, 2022

Background

This report presents findings from two sets of structured-questions (two surveys) distributed to the Canadian DLI community concerning Canadian census data use and discovery. Participants at a recent Historical Census Project presentation which took place on November 24, 2021 during DLI national training were Canadian library specialists (generally data librarians) who provide data services to their university or college community, primarily students and researchers. Susan Mowers, representing the [Canadian Census Data Discovery Partnership \(CCDDP\) project](#), reported on the project in the [Historical Census Project presentation](#), and by means of two online surveys, she consulted the participants about their experiences supporting the finding, use and understanding of Canadian census data on their campuses about their experiences. Links to the online surveys were provided to participants during the presentation. As participants were provided with time to answer the two surveys during the presentation itself, they did not receive any advance notice about the survey topics. Thus, the participants' answers are based on their views and recollected knowledge of the survey topics. Because there was no opportunity to consult records, their survey answers were not based on recorded statistics they may have had regarding their data consultations and teaching sessions.

Following a summary, findings are presented for each question topic below and the number of respondents is provided below, ranging from 25-36 respondents per question.

Contents

| | |
|--|-------------------------------------|
| - Summary | 3 |
| 1. Disciplines who use Canadian census data according to data service providers (data librarians) in postsecondary libraries | 4 |
| Combined Census Years | 4 |
| Use of most recent Canadian Censuses (after 2006) | 5 |
| Use of Censuses from 1996 to 2006 | 5 |
| Use of Censuses from 1665 to 1996 | 6 |
| 2. By level of data users, the estimated percentage of users who do not initially know that the Census covers their topic(s) of interest | 7 |
| 3. Obstacles to the use of census data | 8 |
| 4. Ranking Data Discovery Features | 9 |
| Data discovery features for novice census data users | 10 |
| Data discovery features for Intermediate census data users | 11 |
| 5. Teaching sessions involving finding and using census data | 13 |
| 6. Sum up and opportunities | |
| Sources consulted | Error! Bookmark not defined. |

Summary

Time was provided to a group of Canadian data librarians and specialists during a presentation to provide their views on a series of questions about Canadian Census data. The group of data librarians and specialists who were surveyed are responsible for giving data consultations and teaching sessions about finding and using Census data.

The survey addressed five broad topics, ranging from the use of census data, including census awareness, obstacles to use, and rating potential features for improving Canadian Census of Population data discovery on their campuses. A summary of the findings follows.

(1) Use of Census data by discipline: for 8 out of 9 disciplines, the most recent censuses were the most used censuses, with use decreasing for years 1996 - 2006, then dropping significantly for most disciplines for the pre-1996 years. History and Sociology (including Demography) were the biggest users for this time period, followed by Geography and Economics.

(2) Awareness of the subject matter of the Canadian Census of Population was measured as lack of awareness. Respondents reported that at the beginning of data consultations, 40% of advanced data users, and 80% of novice data users, did not initially know that the census covered their topic(s) of interest.

(3) Obstacles to the use of Census data were viewed as significant. 50% or more data librarians attested that there were 6 (out of 7) obstacles to the use of Census data. The top 2 obstacles at 78% of data librarians, were "Difficulties knowing how to use census platforms to get needed data" and "Difficulties knowing where to search", followed closely by "Lack of awareness .. for courses", "Lack of awareness ... for research", and "conceptual barriers (census jargon)".

(4) Respondents reported considerable differences by the level of the data users in their ranking of the usefulness of 9 listed Census data discovery features. In comparing the three discovery features most highly ranked in each data user group, these were for novice users: keyword searching, user friendly topics, and visual icons, for intermediate users: keyword searching, faceted searching and standardized geography lists, for expert users: faceted searching, detailed metadata, and emphasis on discovery of census documentation.

(5) The most common estimate about the proportion of data teaching sessions which incorporated finding and using Census data was 20%.

The report also addresses two follow-up questions: the meaning of possible gaps identified in the survey, and how to think about user needs around Census data discovery.

Findings

1. Disciplines who use Canadian census data according to data service providers (data librarians) in postsecondary libraries

There were 33 responses to the question:

Which disciplines, in your view, use Census data (by three time periods)?¹

Data librarians were also asked to identify any other disciplines who use Canadian census data - further to the nine listed disciplines (as shown in Figures 1-4). The additional seven write-in disciplines were: Anthropology, Architecture, Education, Environmental studies, Industrial Relations, Social Work, and Urban Planning/Planning. In addition, respondents reported that their institution's Institutional planning and research departments also used Canadian Census data.

Combined Census Years

Figure 1. Number of times data librarians identified disciplines as regular users of Canadian census data for any time period (most recent Censuses, Censuses between 1996 and 2006, and Censuses before 1996).

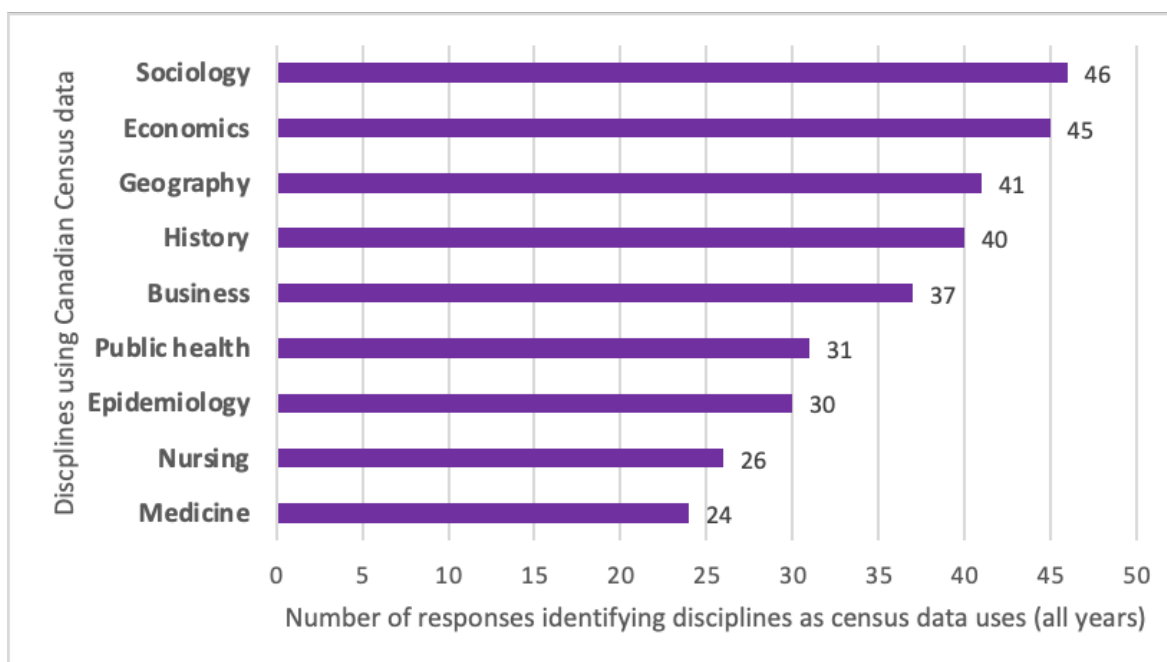


Figure 1 combines responses for 3 time periods (namely data use for most recent Censuses, Censuses between 1996 and 2006, and Censuses before 1996), thus there are a maximum of 99 possible responses per discipline.

¹ The phrasing and layout of the question on the survey was such that it “may” have led 4 out of 34 respondents to reduce the numbers of disciplines that they selected, as 4 out of 34 respondents chose two or fewer disciplines.

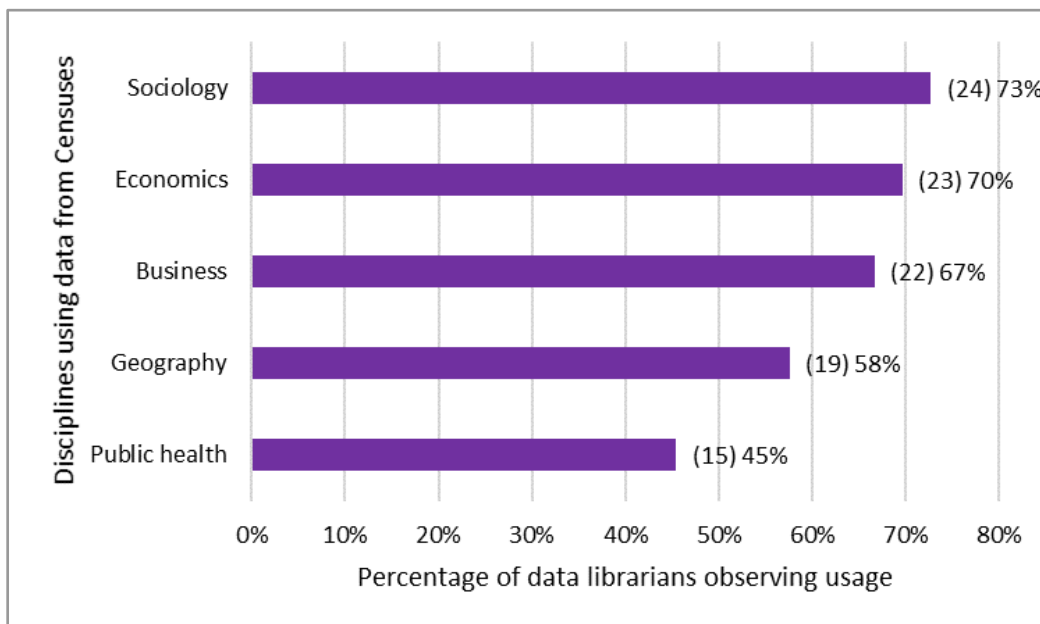
Out of the nine disciplines shown here from the survey, data librarians observed that Sociology was the biggest user of Census data (46 responses total), followed by Economics (45 responses) and Geography (41 responses). It is understandable that data librarians ranked Sociology as either the top or second highest user of Canadian Census data users among the 9 listed disciplines, as Demography falls under Sociology in most schools.

Data librarians identified health-related fields to be the disciplines that made the least use of Census data, with Medicine as the least reported to use Census data (24 responses), followed by Nursing (26 responses) and Epidemiology (30 responses). However, as this is only a general overview of responses, further analysis was done to identify the patterns of use (both percentages and frequencies) by disciplines according to the time period for which there is Census data (see Figures 2-4 below).

Use of most recent Canadian Censuses (after 2006)

Figure 2.

Top five disciplines using data from the most recent Censuses as observed by data librarians and ranked by percentage of data librarians observing usage.

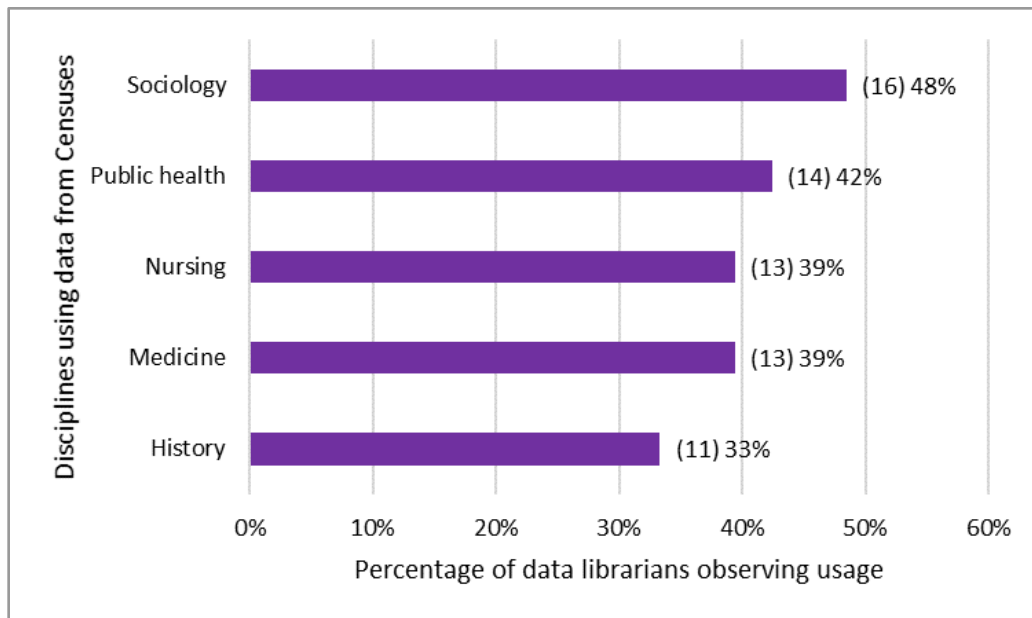


Based on their experience, data librarians also identified Sociology as the biggest user of the most recent Censuses (73% of data librarians). Economics (70% of data librarians) and Business (67% of data librarians) were the second and third most often observed users of the most recent Censuses.

Use of Censuses from 1996 to 2006

Figure 3.

Top five disciplines using data from Censuses from 1996 to 2006, as observed by data librarians and ranked by percentage of data librarians observing usage.

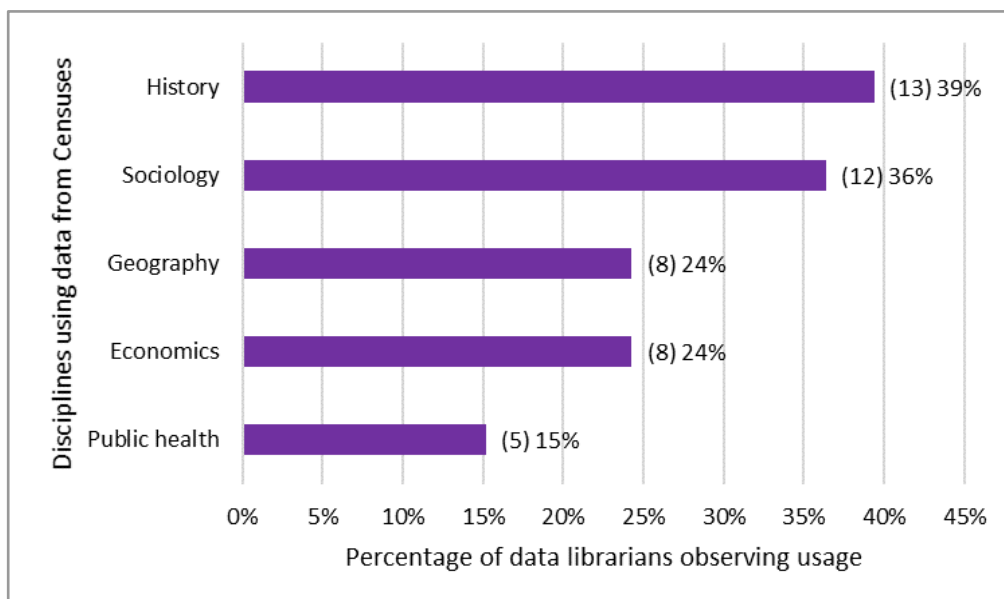


Once again, Sociology was observed to be the biggest user of Census data from the 1996 to 2006 Censuses of Population of Canada (48% of data librarians' responses). However, the second and third disciplines most observed using this time period of Census data changed from Economics and Business (Figure 1) to Public Health (42% of responses) and Nursing and Medicine (both 39% of responses).

Use of Censuses from 1665 to 1996

Figure 4.

Top five disciplines using data from Censuses before 1996 as observed by data librarians and ranked by percentage of data librarians observing usage.



For the older Censuses, the use of census data by most disciplines plummeted dramatically, but increased for History (from last place at 33%, to first place at 39% of respondents). Despite the overall decline in observed usage for this time period, Sociology remained one of the top users of Census data (36% of respondents).

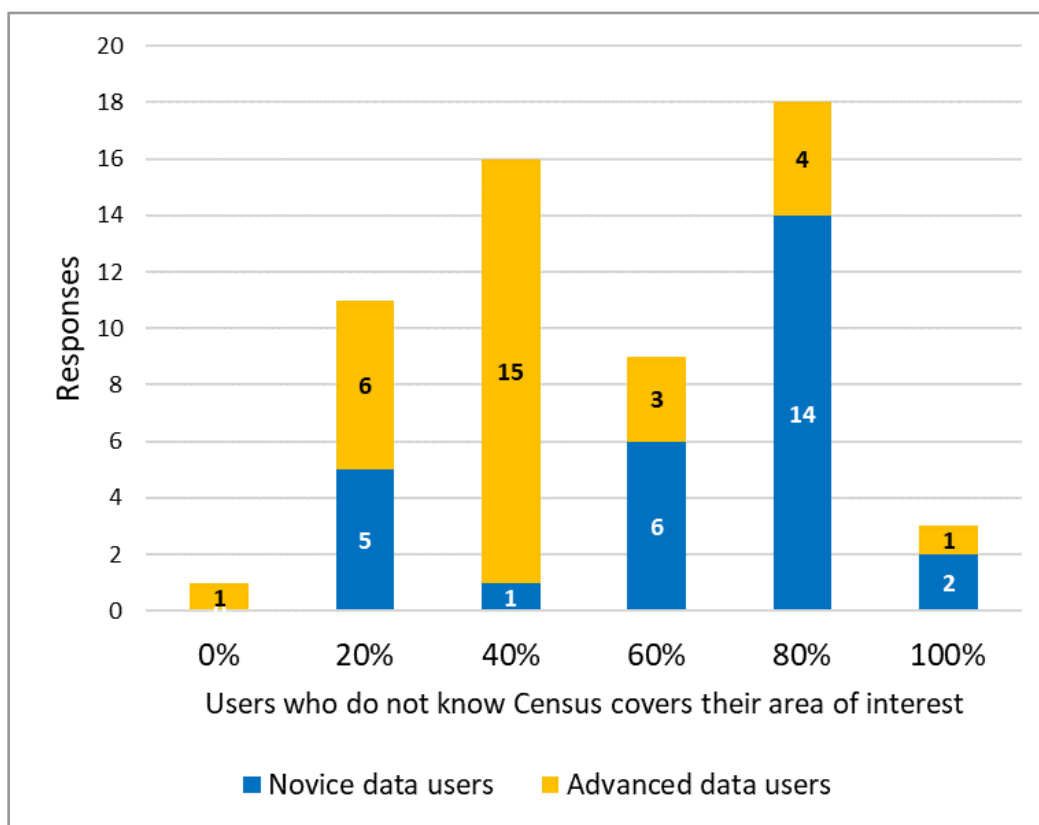
2. By level of data users, the estimated percentage of users who do not initially know that the Census covers their topic(s) of interest

There were 31 responses to the question:

Of your user consultations that involve census data, please estimate the percentage of users who don't initially know that the census covers their topic(s) of interest?

Figure 5.

Percentage of users who do not initially know that the Census covers their topic(s) of interest, broken down between novice and advanced data users, as observed by data librarians.



The majority of data librarians estimate that 40% of advanced data users do not know that the Census covers their topic(s) of interest (15 responses). For novice data users, the majority of data librarians estimate that 80% of these data users do not know that the Census covers their topic(s) of interest (14 responses).

3. Obstacles to the use of census data

There were 36 responses to the question:

Considering your answers on census data use, what would you say are the obstacles (if any) that apply to census data use on your campus ? (Select all relevant answers).

Figure 6.

Obstacles to the usage of Census data on campus, ranked in order of total number (and percentage) of responses.

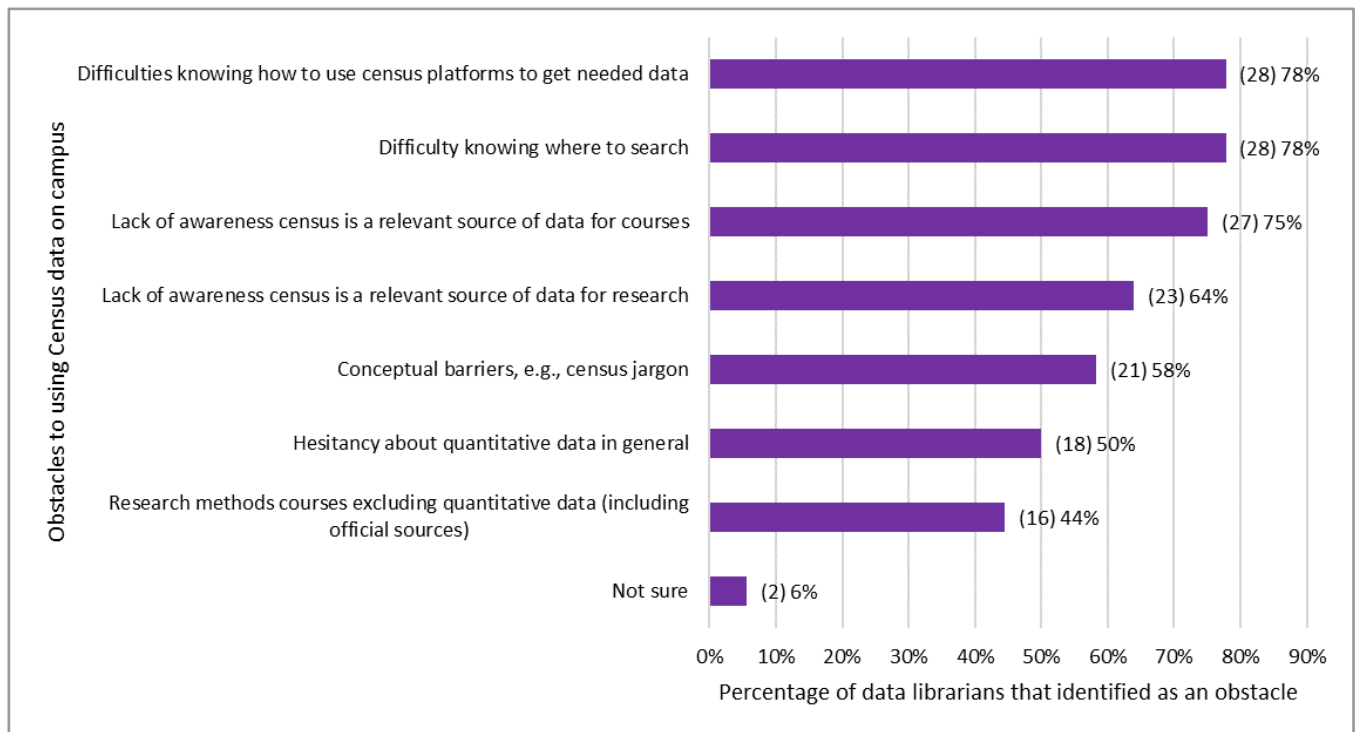


Figure 6. There was a high selection of obstacles to the use of census data existing on postsecondary campuses. As shown above, 6 out of the 7 obstacles were identified by at least half of the data librarian respondents. Only 4 out of the 36 respondents gave single answer responses.

The top two obstacles to the use of census data were identified by 78% of respondents were: “Difficulties knowing how to use census platforms to get needed data” and “Difficulties knowing where to search”.

Difficulties knowing where to search do not seem to exist in isolation. **On average, data librarians each selected 5 obstacles in response to this question, suggesting that, overall, there are many perceived barriers to access.** These first two obstacle responses were followed closely at 75% of responses for Lack of awareness of the census as a relevant source of data for courses. Other notable results were: Lack of awareness of the census as a relevant source of data for research (64% of responses), Conceptual barriers, e.g., census jargon (58% of responses), Hesitancy about quantitative data in general (50% of responses) and to a

somewhat lesser extent, Research methods courses excluding quantitative data (including official sources) (44% of responses).

4. Ranking Data Discovery Features

There were 25 data librarians who responded to Question 4:

Please indicate WHO, if anyone, would benefit from access to these data discovery features to retrieve good historical Census data and documentation results? (Novice, Intermediate and Expert users).

The data users referred to in this question reflect their level of experience with data. Novice data users tend to be new to finding and using sources of quantitative data, whether for coursework or research. Intermediate data users, in contrast, would have used some sources of quantitative data and be comfortable with basic functions for finding and retrieving data. Expert data users tend to be *more* knowledgeable about *more* data sources, have some awareness of data documentation, can navigate more advanced data portals, and can better prepare data for analysis.

The list of 12 census data discovery features, as shown in figure 7, for this question was not intended to be exhaustive, but rather to get a sense of how data librarians (who provide consultation services and teaching on finding and using Canadian census data, current and historical), view the usefulness of types of Census data discovery features for a single, envisioned Canadian Census data discovery portal.

This section refers to the results of the second survey, concerning Census data discovery.

Figures 7-10 present the results for the question above, according to the level of data users, novice, intermediate and expert.

Combined (all levels of data users)

Figure 7.

Ranking useful Census data discovery features for novice, intermediate, and expert-level users of Census data.

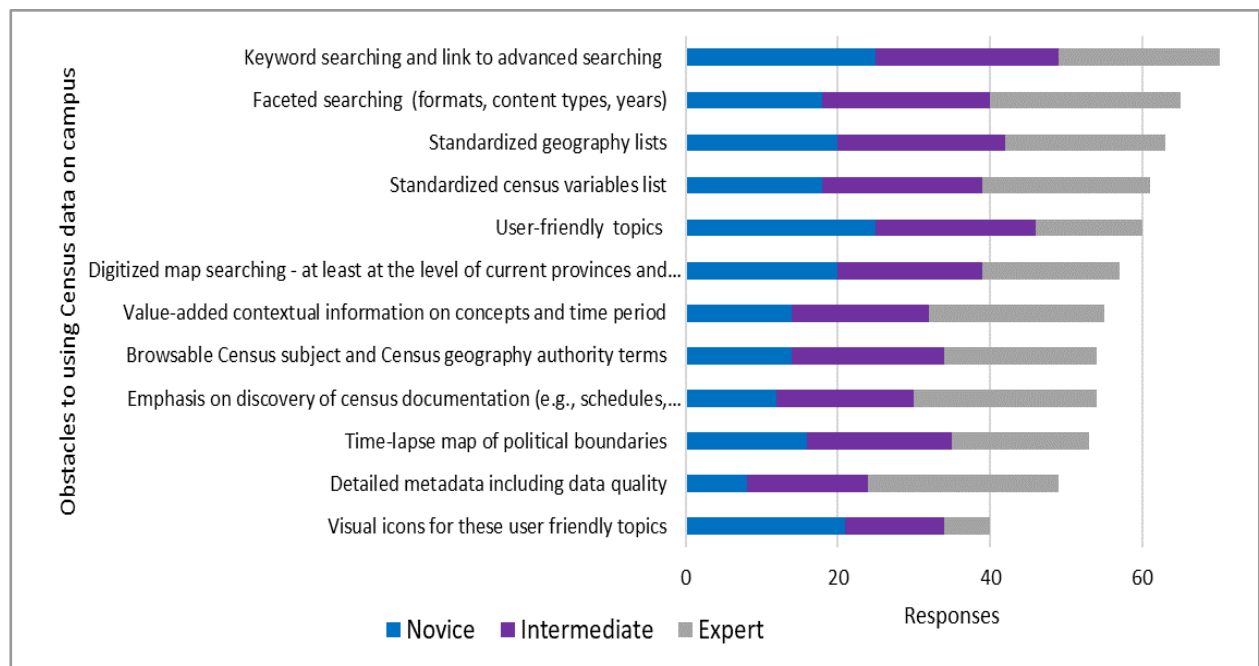


Figure 7. Given that 25 responses were received on the usefulness of the above for each of three (3) levels of data users, there are a maximum total of 75 responses for each census data discovery feature.

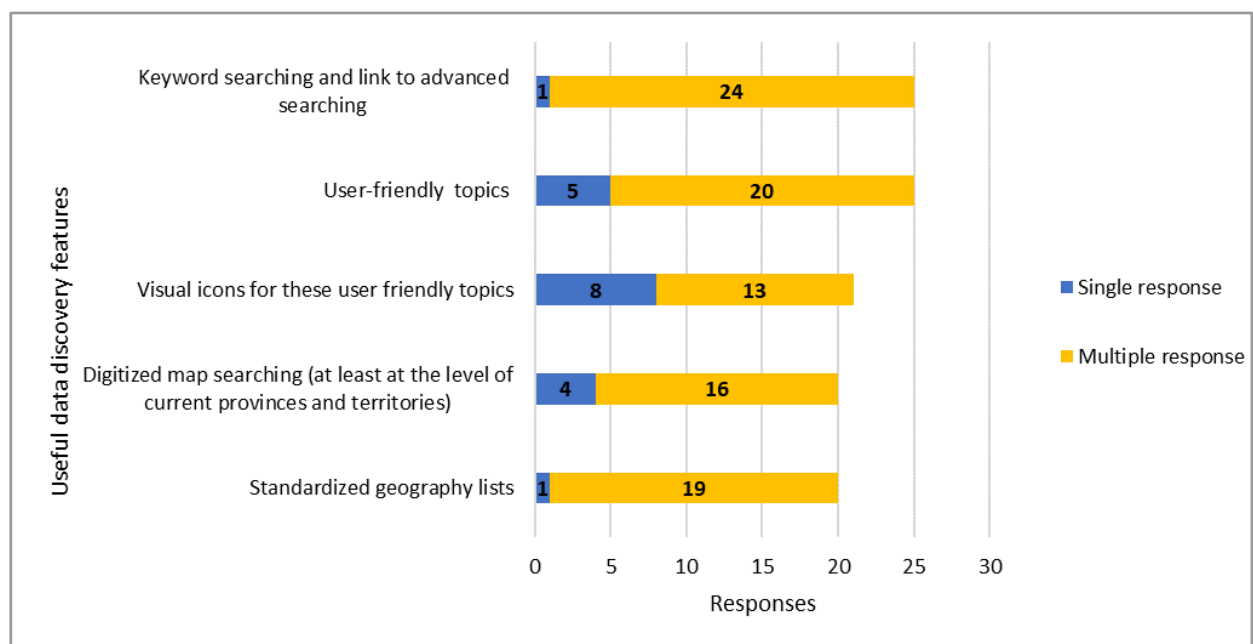
Noting this limitation, it is interesting that the relevance of the features ranged from 45 responses to 72 responses. It is understood that by extension, having supporting metadata or geospatial boundary maps would be necessary for these features to be possible.

All user groups were indicated to have a high level of relevance for: keyword (total 72 out of 75) and faceted searching (total 65 out of 75), standardized lists of geographies (total 65 out of 75) and of census variables (total 63 out of 75).

Data discovery features for novice census data users

Figure 8.

Top five data discovery features identified by data librarians as most useful for novice users of Census data.



Compared with the next two figures, two out of the top ranked data discovery features are unique to this figure - not found on figures 9 and 10 below. These two features are: (a) Related visual icons for these user-friendly topics and (b) Digitized map searching. These two features further standout from figure 9, for intermediate users, for having been selected for novice users only (8 or 32% of the time and 4 or 16% of the time respectively) and not for other levels.

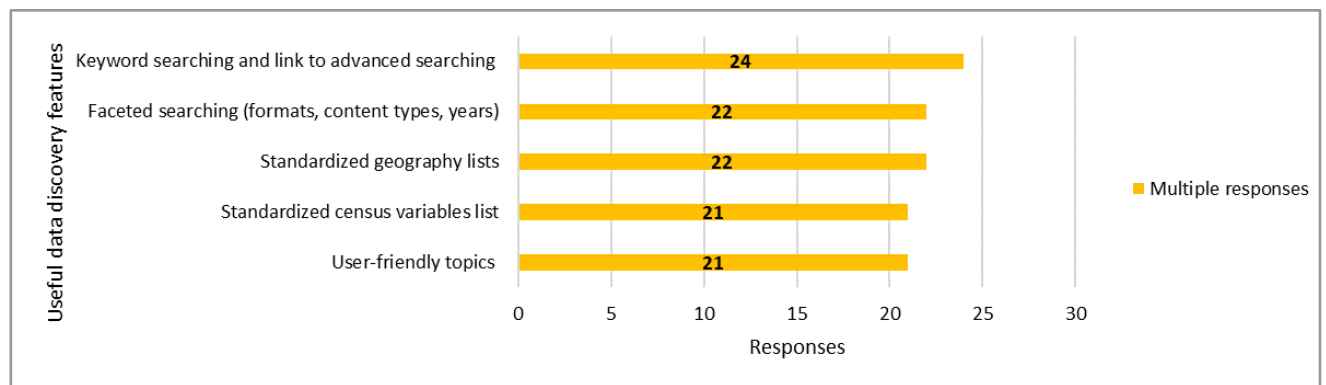
Following these 5 top ranked features for novice users, the next four highest ranked census data discovery features are, (a) Faceting searching (18 or 72%), (b) Standardized Census variable list (18 or 72%), (c) Time-lapse map of political boundaries (16 or 64%), and (c) Standardized census variables list (16 or 64%).

These types of features are important in the light of three noted obstacles to the use of Canadian census data that appear to be particularly relevant for novice users. See figure 6 on page 6. These are, (a) Lack of awareness of the census as a relevant source of data for courses (75%), (b) Conceptual barriers, e.g., census jargon, (58%), and (c) Hesitancy about quantitative data in general (50%).

Data discovery features for Intermediate census data users

Figure 9.

Top five data discovery features identified by data librarians as being most useful for intermediate users of Census data.



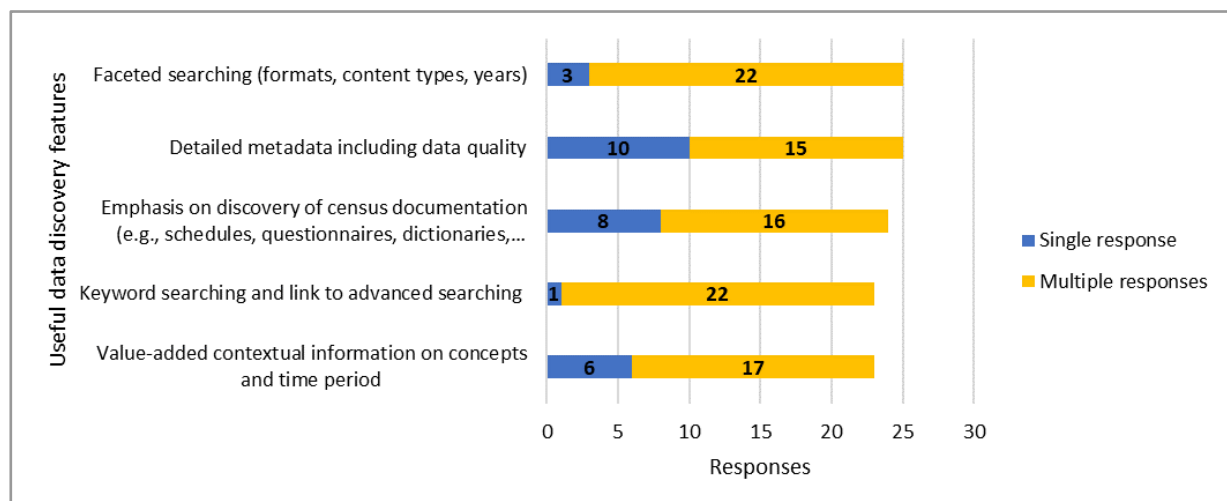
As with novice users (Figure 8), Keyword searching is the top ranked useful census data discovery feature for intermediate users (24 responses or 96%). However, as intermediate users would have greater knowledge of Census topics, concepts and language when compared to novice users, the next most useful data discovery features for intermediate users were identified to be those that facilitated more concise searching: (a) Faceted searching and (b) Standardized geography lists (both 22 responses or 88% each); as well as more in-depth searching or contextual information: (c) Standardized census variables list and (d) User-friendly topics (both 21 responses or 84% each).

Following these 5 top ranked features for intermediate users, the next highest ranked census data discovery features are, (a) Browsable census subject and census geography authority terms (20 responses or 80%), (b) Digitized map searching (19 responses or 76%), (c) Time-lapse map of political boundaries (19 responses or 76%) and (d) Emphasis on discovery of census documentation (e.g., schedules ..) (18 responses or 72%).

Data discovery features for expert census data users

Figure 10.

Top five data discovery features identified by data librarians as being most useful for expert users of Census data.



The top five data discovery features identified for expert users of Census data are again different from those identified for novice or intermediate users.

A greater focus was put on in-depth Census data searching and accessing detailed data and information with top features identified as: (a) Faceted searching (25 responses or 100%); (b) Detailed metadata, including metadata quality (25 responses or 100%); (c) Emphasis on discovery of Census documentation (24 responses or 96%); (d) Keyword searching (23 responses or 92%); and (e) Value-added contextual information on concepts and time periods (23 responses or 92%).

The remaining data discovery features deemed to be significant for this group were: Keyword searching and link to advanced searching (22 responses), Standardized census variables list (21 responses), Standardized geography lists (20 responses), Browseable Census subject authority terms and Census geography authority terms (19 responses), Time-lapse map of political boundaries (17 responses), Digitized map searching - at least at the level of current provinces and territories] (17 responses), User-friendly topics (13 responses).

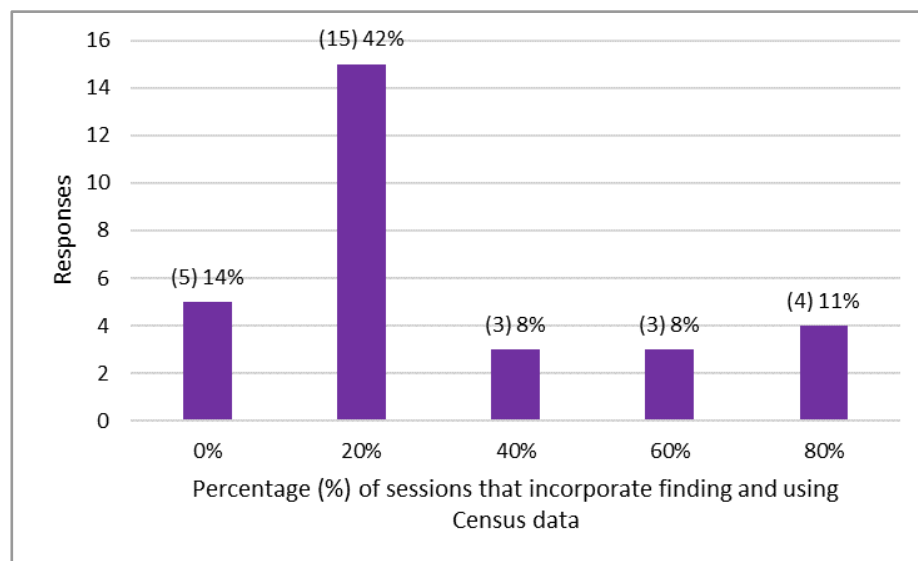
5. Teaching sessions involving finding and using census data

There were 36 responses to question 5:

What percentage of your teaching sessions incorporate finding and using census data?

Figure 11.

Percentage (in 20% intervals) of teaching sessions held by data librarians that incorporate the finding and using of census data.



42% of data librarians said that only 20% of their teaching sessions incorporate finding and using Census data. Taken together with data librarians who said that none of their teaching sessions incorporate finding and using Census data, a small majority of data librarians (56% or 20 responses) their teaching sessions incorporated little or no finding and using Census data.

The distribution of other responses: 40%, 60%, and 80% of teaching sessions incorporating Census data was fairly even, with data librarians who responded that 80% being the next highest category at 11% or 4 responses. No data librarians said that 100% of their teaching sessions incorporated finding and using Census data. 16% (or 6 responses) of respondents chose “not applicable” as their response to the question.

Issues for further exploration

1. Understanding census data use and challenges to data use

Given the high response rates for the series of obstacles (see Figure 6) to the use of Canadian census data, more research is suggested. Are these obstacles more pronounced for some disciplines, or for types of data users and beyond? More research would be helpful to confirm the findings and the implications of the gaps observed. First, in the “high” number of data consultations where initially, users do not initially know that the census is a relevant source of data for their questions (80% for novice data users against 40% for advanced data users. Second, in the highest selection for percentage (20% percentage category) of teaching sessions that incorporate finding and using census data.

There is an interesting possibility that because of knowledge gaps about what census data covers, that there are real differences in doing consultations on census data. It would be useful to know through further investigation with data librarians whether there are different levels of complexity and time required for both answering data consultations, or giving teaching sessions, involving census data. It would also be useful to know from the point of view of end-users, if they themselves feel that finding and using census data is more complex and time-consuming than other data types.

Other explanations for these gaps could be poor access to reliable and understandable census data tools of different types, compared to tools for some other data sources. By understanding such differences, strategies can be identified on a broader level, rather than local level, to facilitate the findability and use of Canadian census data.

2. Understanding users’ census data discovery needs

As a framework for further research into user needs for Canadian census data discovery, it is useful to step back and consider a more fluid and refined definition of users than simply novice, intermediate and expert data user levels, as defined in the survey.

As follows, here is an important, more refined approach to understanding census data discovery user needs from Callum Foster’s presentation for the Disseminating Census Data: Learning from Experience workshop on February 16, 2022. As the first step, Foster presents four broad types of census end-products which people could be interested in finding and using. The author presents these distinct categories: **data**, **commentary**, **exploratory** and **explanatory**. Overall, most users would likely “prioritize” one category over others.

Callum Foster (2022) presents a framework for how to think about user needs using four categories of census products. These products present four ‘usable’ information facets revealed by the census and represent what it is that the user is looking for:

1. Numeric information

Data and **Commentary**

Data are primarily “numeric”, but they will include labels (generally textual) for the numbers. Data may also be spatial, e.g., boundaries of municipalities, provinces or territories, or latitude, longitude

Commentary is primarily textual commentary or information about the data, and often includes topic-specific graphs, maps, and tables.

- a. Users may seek presentation-ready data or commentary. Such sources can be readily presented and cited in a report and may also be able to be graphed or mapped.
- b. Users may also want to analyse census data across a region or Canada to answer questions not answered in presentation-ready census data or commentary. Typically, users want to download this data for their own use, e.g., statistical analysis or mapping.

2. Interactive.

Exploratory. Exploratory sources may allow the user to investigate themes like census topics, places, peoples, and times. Typically, these sources provide interactive tools to provide summarized census information ².

3. What was going on? How did it come about?

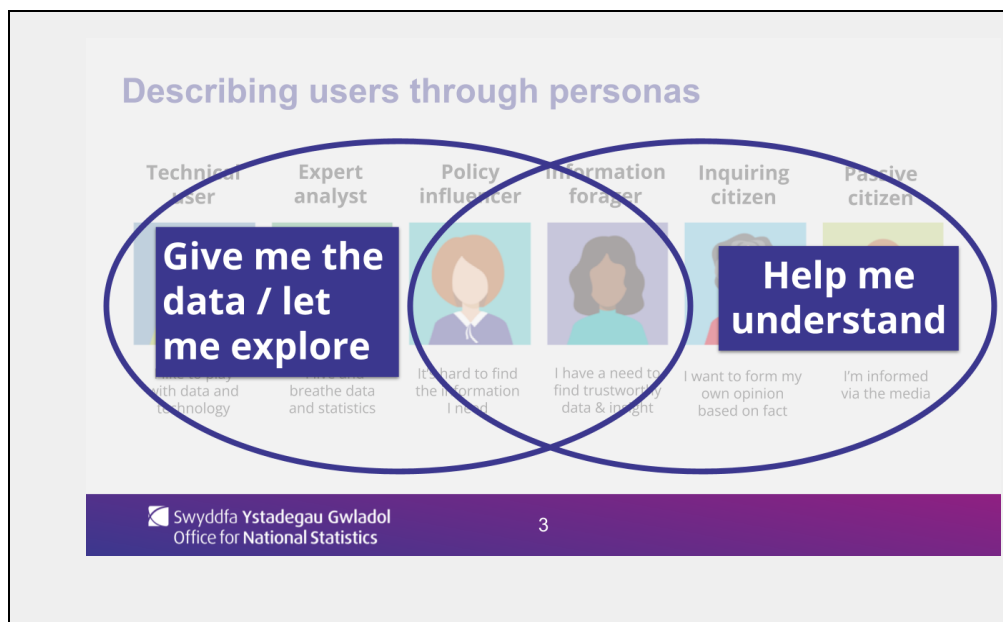
Explanatory.

- a. Explanatory sources may provide synopses, definitions, and visualisations to explain concepts and trends. This may also cover information about the geography of the time and the censuses including how the censuses were carried out including the questions asked.

In the following **figure 12**, Foster (February 16th, 2022), helps us to approach how to think about meeting user needs on a scale from “Give me the data / let me explore”, to “Help me understand “. We can see from the articulation of these two broadly stated needs that they are quite different. Therefore, the presentation of supporting online tools to meet one or the other of these needs would need to be quite different.

² The range of census topics is potentially the same whether from data, commentary, or exploratory tools and sources. Generally, census topics (or census characteristics) revolve around peoples’ cultures, lives, and demography. Examples of cultural topics may include languages, ancestral background, Indigenous status, place of birth, educational background, or religious affiliation. Examples of census topics about people’s lives are types of households and dwellings, work (e.g., field, occupation, and place of work), income as applicable, and relative level of rural versus urban. Demographic variables include age, sex, family size, and types of relationships making up family and family members’ ages.

Figure 12



Second, by Callum Foster's having visualized these two needs as overlapping circles, we can see that, while people may prioritize one type of need over another, neither of these needs is mutually exclusive. Therefore these needs are actually a spectrum from one to the other.

To better understand this framework, by recognizing that the users themselves are quite diverse, the advantage of Foster's framework is that he makes no assumptions about the degree to which users may or may not have used census information or different presentations types before.

Figure 13

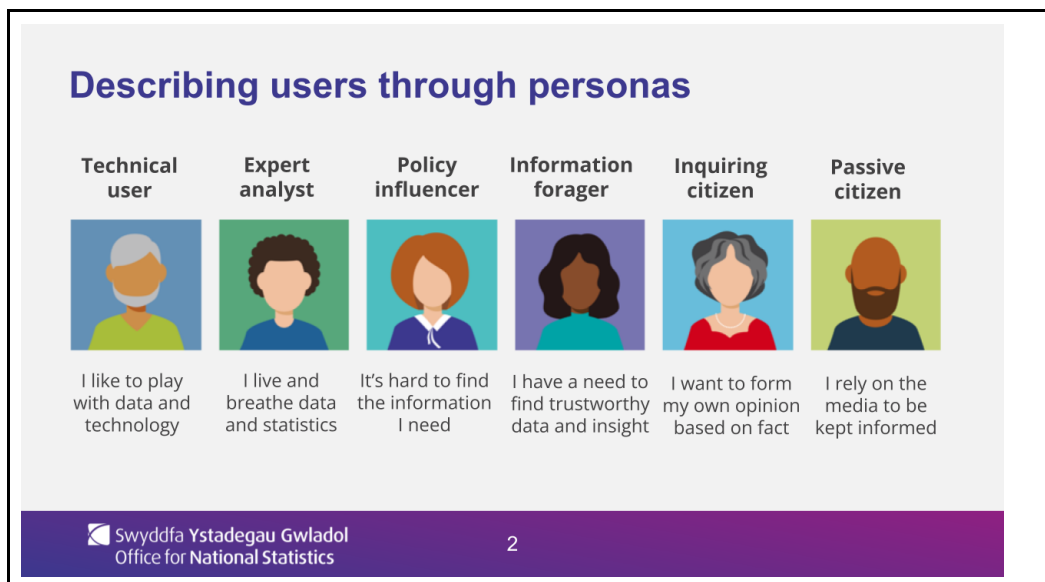
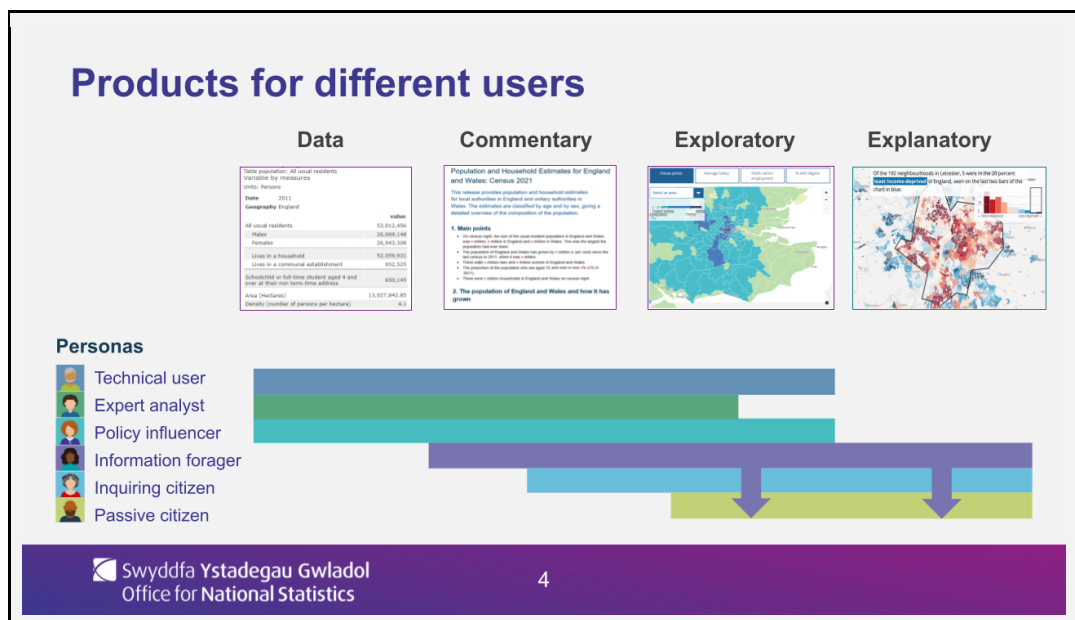


Figure 14



Callum Foster’s census personas add a level of insight about census discovery needs to that provided in the survey responses which looked at “levels” of data-users as: novices, intermediate and expert. According to **figures 12, 13 and 14** shown above, I have adapted Callum Foster’s analysis of “Data/information end products and Personas”. I have mapped his personas to “Data discovery end products” and I present an analysis of these by the 12 features as discussed above under section 4: Ranking Census Discovery Features (pp. 9-13).

It is possible that one person may take on different personas in different contexts. One factor affecting this may be due to their understanding of one census topic versus another, or their understanding of, or interest in finding out about, census concepts like geography. A second factor is their own tendencies to explore the different available formats and presentations available of census information. A final factor is access to support, like peers, other experts, or user tools.

Therefore, in contrast to the approach taken by the survey with its three levels of data users, novice, intermediate and expert, we need not make assumptions about users’ level of familiarity with census data overall. A more nuanced approach is to consider user needs from the point of view of census end-products, in other words: the users’ self-identified census information goals. To that end, assessing the usability of data discovery features can be better assessed by thinking about them according to these end-product families below.

| Foster’s personas | Technical users and policy analysts | | (Information forager) | Engaged and inquiring citizens | |
|---|---|--|--|--|---|
| | Give me the data / let me explore | | | Help me to understand | |
| <i>Data discovery end-products</i> | <i>Metadata about data measures (e.g., codebooks)</i> | <i>Census reference documents (how the census was carried out and the questions)</i> | <i>Census concepts, geographic boundaries, analytical reports; or commentary</i> | <i>The times and places + other contextual information</i> | <i>Graphical displays of census information</i> |
| Browsable Census subject authority terms and Census geography authority terms (e.g. geographic units) | | | | | |
| Detailed metadata including data quality | | | | | |
| Digitized map searching - at least at the level of current provinces and territories | | | | | |
| Emphasis on discovery of census documentation (e.g., schedules, questionnaires, dictionaries, instructions, maps) | | | | | |
| Faceted searching (filters): formats, content types, years | | | | | |
| Keyword searching and link to advanced searching | | | | | |
| Standardized census variables list | | | | | |
| Standardized geography lists (e.g., places) | | | | | |
| Time-lapse map of political boundaries | | | | | |
| User friendly topics | | | | | |
| Value-added contextual information on concepts and time period | | | | | |
| Visual icons for these user-friendly topics above | | | | | |

Sources consulted

- Bethel, A. C., Rogers, M., & Abbott, R. (2021). Use of a search summary table to improve systematic review search methods, results, and efficiency. *Journal of the Medical Library Association*, 109(1), 97–106. Library, Information Science & Technology Abstracts.
- Dixit, R., Rogith, D., Narayana, V., Salimi, M., Gururaj, A., Ohno-Machado, L., Hua Xu, Johnson, T. R., & Xu, H. (2018). User needs analysis and usability assessment of DataMed—A biomedical data discovery index. *Journal of the American Medical Informatics Association*, 25(3), 337–344. Library, Information Science & Technology Abstracts.
- Foster, C. (2022). (Presentation title). Disseminating Census data: Learning from experience. Virtual workshop
- Hecht, S. (2019). *Improving UX of Open Government Data Platforms*. ISPIM Conference Proceedings, Manchester. https://ocul-uo.primo.exlibrisgroup.com/permalink/01OCUL_UO/ap3p9v/cdi_proquest_journals_2297095177
- Lamb, I., & Larson, C. (2016). Shining a Light on Scientific Data: Building a Data Catalog to Foster Data Sharing and Reuse. *Code4Lib Journal*, 32, 1–1. Library, Information Science & Technology Abstracts.
- Magnus, R. (2021). Finding and Evaluating Census Data: A 360-Degree Perspective. *Online Searcher*, 45(3), 26–31. Library, Information Science & Technology Abstracts.
- Mathys, T. (2005). THE GO-CEO! SPATIAL DATA PORTAL: A DATA DISCOVERY AND RESEARCH TOOL FOR THE SOCIAL SCIENCES IN UK ACADEMIA. *Assignment*, 22(4), 15–20. Library, Information Science & Technology Abstracts.
- Mattison, D. (2003). Counting Heads Around the World. *Searcher*, 11(1), 32. Library, Information Science & Technology Abstracts.
- McKay, D., Chang, S., Smith, W., & Buchanan, G. (2019). The Things We Talk About When We Talk About Browsing: An Empirical Typology of Library Browsing Behavior. *Journal of the Association for Information Science & Technology*, 70(12), 1383–1394. Library, Information Science & Technology Abstracts.
- Remy, L., Ivanović, D., Theodoridou, M., Kritsotaki, A., Martin, P., Bailo, D., Sbarra, M., Zhao, Z., & Jeffery, K. (2019). Building an integrated enhanced virtual research environment metadata catalogue. *Electronic Library*, 37(6), 929–951. Library, Information Science & Technology Abstracts.
- Smith, M. (2010). Statistics from the ground up. *School Libraries in Canada (17108535)*, 28(1), 21–23. Library, Information Science & Technology Abstracts.
- Wicaksono, A. F., & Moffat, A. (2021). Modeling search and session effectiveness. *Information Processing & Management*, 58(4), N.PAG-N.PAG. Library, Information Science & Technology Abstracts.