

KnitBI: Stitching Together Library Data Using Power BI

Cairo Sanders, University of Victoria

Introduction

In 2022, the University of Victoria (UVic) Libraries had a “data everywhere” problem – we had a lot of data, in many storage locations, generated by many units and generations of employees. We are not the first academic library to have faced this problem or to seek a more centralized solution. For instance, at the 2016 Library Assessment Conference, Hurst et al. presented preliminary research and planning for a library assessment dashboard and toolkit for a “data in one place solution”.¹ Their work showed a common desire across libraries for a set of standards and a simpler solution for accessing, visualizing, and analyzing their operational data.¹

Building on this foundation, we went to work building our one-stop solution for library data, with the goals of:

1. Centralizing data
2. Improving the data interface experience
3. Minimizing the reporting burden for data
4. Using the data to make data driven decisions
5. Developing data literacy skills across the library

Motivation

We inventoried UVic Libraries’ data holdings in 2022 across ten library departments and found over 34 distinct locations where data were being stored, and which were being captured or created using more than 37 tools. With many sources of data come several user interfaces. We graphed this tangled mess of data from the unit to the storage location in a Sankey chart and shared our findings within the library (Figure 1).

When data is mismanaged, information is mismanaged. When there is data duplication, various outlets for access, and no protocols, data can be used to draw inaccurate conclusions. Thus, data collected is more often than not left unused for more important tasks. When data is analyzed, each instance may require significant work pulling, cleaning, transforming, and visualizing the data. This takes time, and with messy, disjointed data, that time is tenfold.

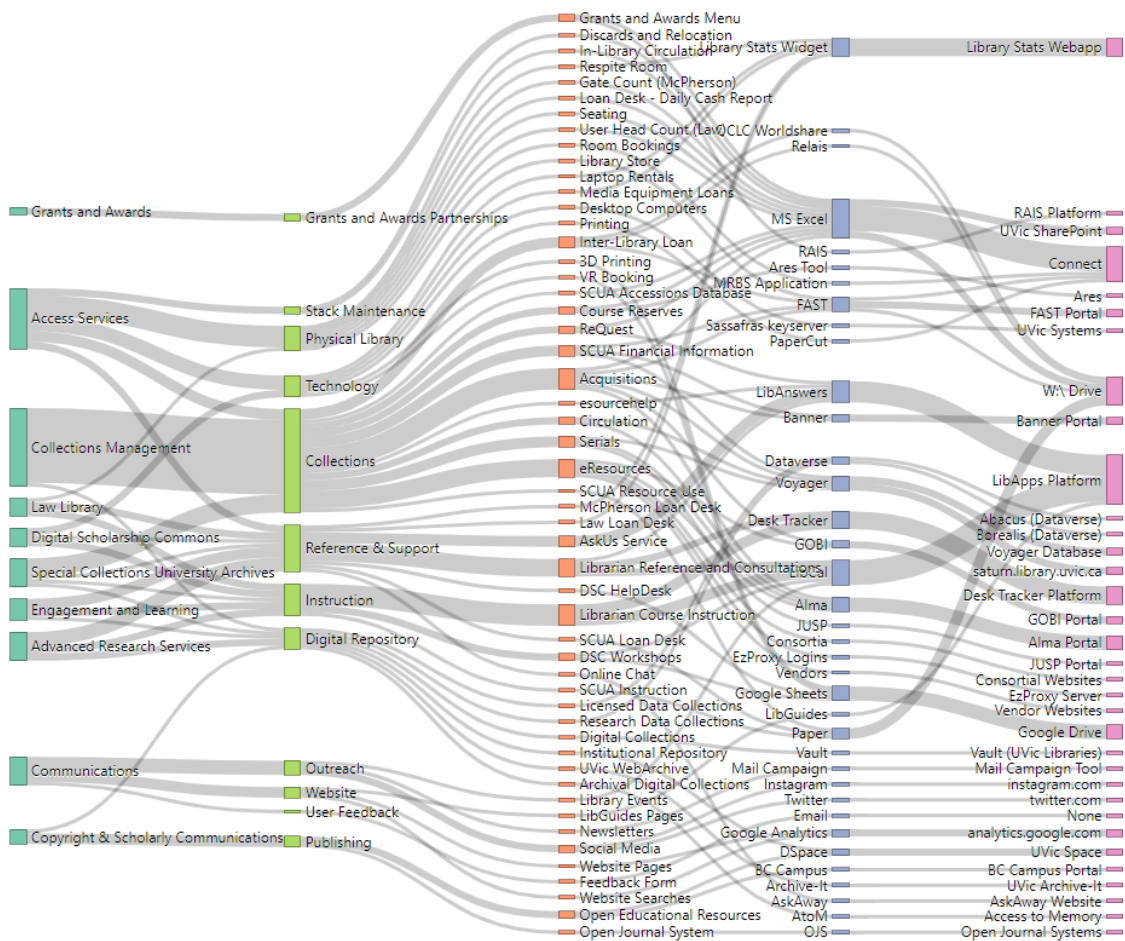


Figure 1. Sankey Chart Connecting University of Victoria Libraries' Data (2022)

Casting On: How the Project Started

Starting the project involved several key steps to ensure a solid foundation. We started by reviewing other libraries doing similar work. These previous examples of centralized solutions by other institutions provided valuable inspiration. For instance, the University of Pennsylvania Libraries has developed a bespoke library activity data farm written in the Ruby scripting language which they titled MetricDoc.² that can pull data from an array of sources. Portland State University Library completed a similar project in which they collated their data into a centralized database and created their dashboard using C3, a JavaScript library.³ Both these solutions involved high code interfaces.

Although one of our tested tools would involve high-code solutions with D3.js we ultimately decided that a bespoke interface was not feasible for the University of Victoria Libraries, which would require a dedicated developer and a mass undertaking of change management. We hoped to reduce these burdens via an existing business intelligence tool.

Choosing a Tool

When done well, data visualization is faster to read and easier to interpret than tabular data.⁴ To ensure a wider audience of data, and to help produce out of the box visualizations for demonstrating and sharing, we sought our solution in the form of data dashboards. There are a variety of such 3rd party solutions to creating dashboards. We tested three well established software solutions: Tableau, D3.js (a JavaScript library), and Power BI.

Tableau is a low code solution to designing flexible and customizable data visualizations; however, in the testing stage, we found that the data transformation was cumbersome a very important aspect of the project. Additionally, the load time, even on small amounts of data, was laggy, which is not ideal for user experience. To share Tableau data in a secure manner for more sensitive data, our organization would need to purchase and maintain a Tableau server, causing to additional labour and cost burden.⁵

D3.js is a library in JavaScript. Since it is written in a scripting language, the possibilities for custom visualizations are virtually endless. With this coding solution, the time it took to develop each visualization including molding the data and custom building each chart was too high. Additionally, the dashboard would need to be hosted and maintained on a local server, which would require more money and labour.

Power BI is a Microsoft solution for data visualizations and dashboards, which pairs easily with other Microsoft tools (e.g. Excel and SharePoint). Our institution already had a Microsoft suite license and culture of using the suite of tools. Also built into Power BI is Power Query, which allows users to query and transform data from their source, without editing the source file; a useful addition given our needs. Power Query is a robust solution to data transformation, for which UVic Libraries data requires a lot of. It is less powerful for data cleaning, which, when data is particularly messy, we chose to use Python to clean.

Power BI was selected as the solution for this project because it allows for:

1. easy transformations
2. fast visualization building
3. direct and flexible connections to data sources
4. integration with other Microsoft tools

Power BI's flexibility for creating data visualizations is the most rigid of the reviewed solutions. However, research in human computer interaction shows that comprehension of data depends greatly on what graph is used. The graphs we so often see, such as charts and bar graphs, are used because they have been proven to most accurately represent continuous data, and categorical data, respectively.⁴ Whereas creativity and embellishment in data visualization draws the audience's eye and helps them remember the message.⁴ Since the goal of our interface is to communicate information accurately,

simple, best practice visualizations was our optimal design choice - and those are the types of visualizations that are easy to create in Power BI.

Inventorying Data

The centralized data dashboard project was not completed overnight. As a faster solution in the interim, the data inventory that was used to create the Sankey chart (Figure 1) was more consciously built out to provide library staff access to their data. We added additional fields to help users access, pull, and understand their data. Such as contact information, links to documentation, and available data types. This was done first simply in an Excel workbook. It was then translated into a custom user interface in Power BI so that user could easily search, filter, access, and understand existing data sources in the library (Figure 2). Even after the centralized data dashboard was developed, the inventory was still available and maintained for UVic libraries as an alternative to the dashboard.

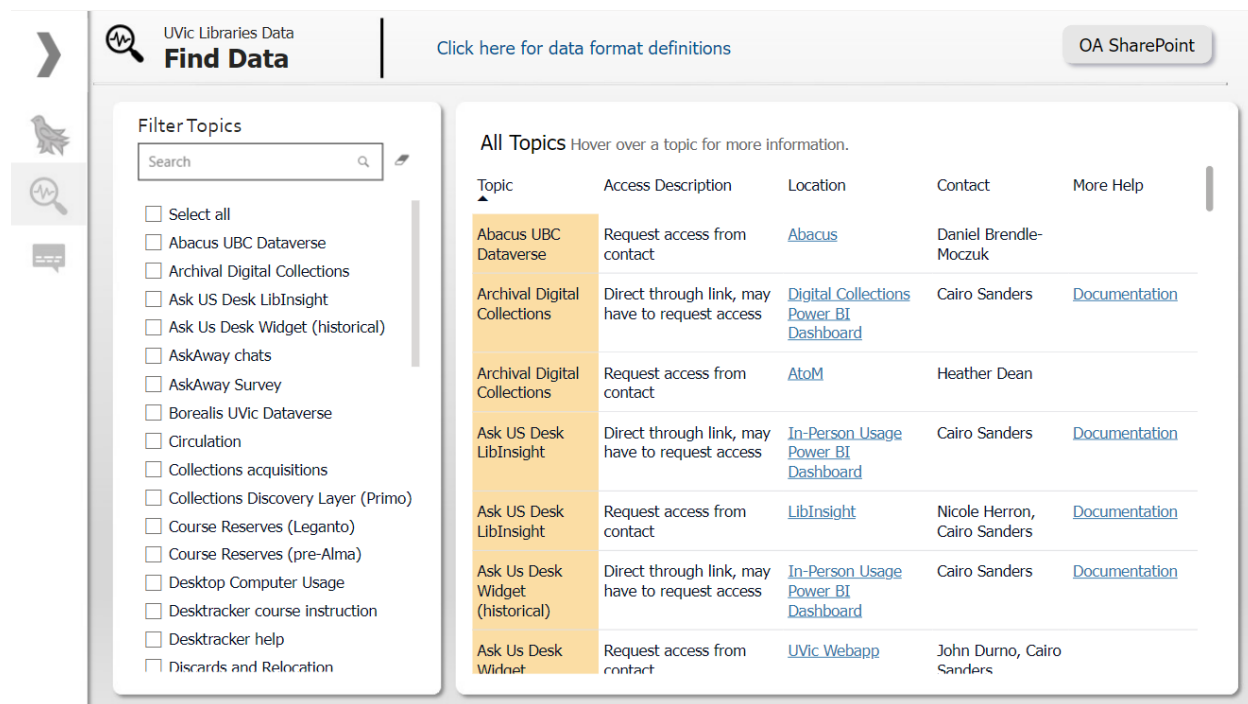


Figure 2. University of Victoria Libraries Data Inventory

Collaboration with Units

It was important to collaborate and receive buy-in from data owners early on in this project. At the start of the project, some units were already invested in their data and recognized its potential for improvement. By working with these units first, we were able to develop a proof of concept. We then shared these data dashboards with other library units. Figure 3 shows a workflow that naturally developed with unit collaborations and communication. The content and structure of reports is reliant on:

1. The unit needs
2. Institutional reporting needs

3. Data driven decision making from management
4. Change management

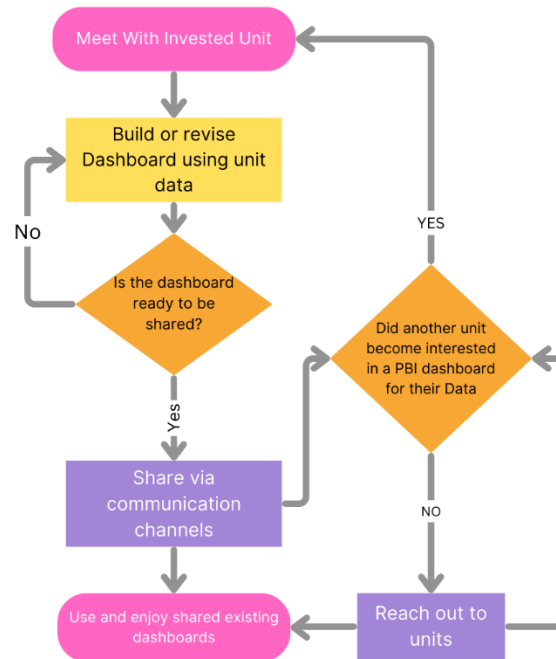


Figure 3. Unit Collaboration Workflow for Building Data Dashboards at the University of Victoria Libraries

A Sweater: Research Help Data as an Example Dashboard

Data Sources

A project to improve collection of statistics on research help consultations illustrates the process. This data was being collected in many different locations (Table 1). Data was pulled in various ways, prioritizing automation to save time. However, some sources had no option to access the data from external sources apart from manually downloading the data tables (Table 1).

Table 1. Data Sources and Their Connections in UVic Libraries' Research Help Data Dashboard

Data Source	Content	Data access method
LibCal	Consultation Appointment Bookings	APIs
LibAnswers	Email Tickets	APIs
Excel	Local Storage Research Help	Direct connection to PBI
DeskTracker	Manual Input Research Help	Manual download
AskAway	Live Chat	Manual download

This data was very messy, mainly due to the names of the librarians associated with a research help transaction, and the qualitative input regarding the nature of the research help. Because Power Query's limited ability for data cleaning, this is done with Python. Figure 4 shows the flow of data into the Research Help Dashboard. It is possible to use Python within Power Query Editor.⁶ The data sources that are accessible via APIs are also easily queried through Power BI. There are a few other ways to query APIs in Power BI: with the Web Connector, or with Power Automate, another Microsoft Power tool.^{7 8}

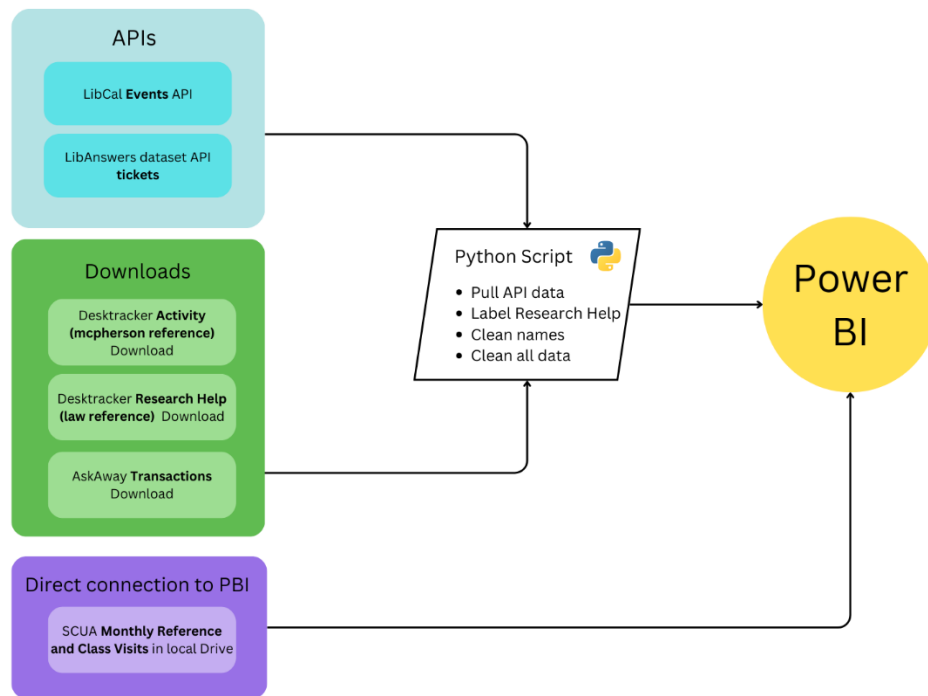


Figure 4. Data Flow Diagram for the Research Help Dashboard at the University of Victoria Libraries

User Interface

The home page of each dashboard is the help page, which has a brief explanation of what the dashboard is (Figure 5-1), what data are included (Figure 5-1), and a list of variables in the report (Figure 5-2). We found that users required this explanation before diving into the

data, or they felt lost or overwhelmed. The variable definition tables we created were inspired by data dictionaries, which are commonly used alongside shared datasets.⁹ The variables are explained by their source, examples, and potential warnings. The warnings communicate where there might be missing or incorrect data. On the home page, there are links at the top which take the user to other helpful resources: a tour of the specific dashboard, a hands-on workshop, a LibGuide on general Power BI usage, and accessibility tips for built in Power BI accessibility tools (Figure 5-3).

Colour is used on this intro page to draw the user's eye to the additional resources, as most other elements are grey or white. On the left is page navigation which can expand and close (Figure 5-4). Navigation is the same on every page but starts expanded only on the help/start page, so that users are aware of it, but as to not claim as much space.

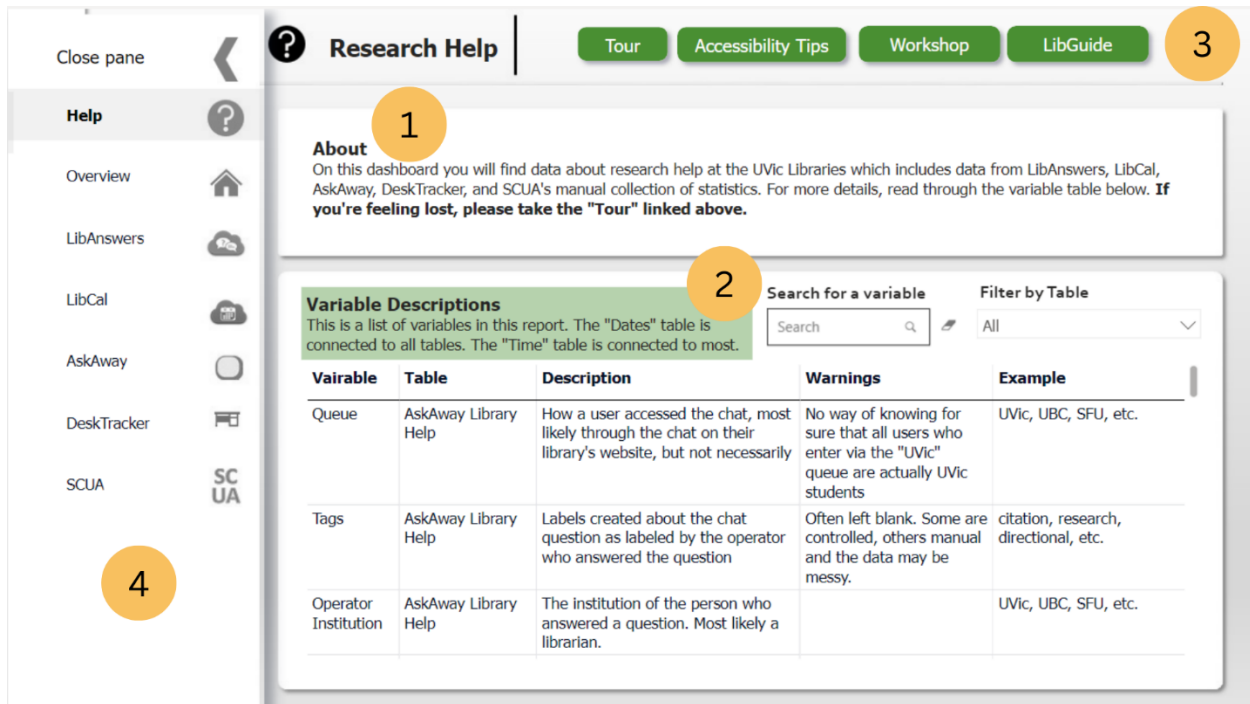


Figure 5. The Home/Help Page of the Research Help Dashboard at the University of Victoria Libraries

Other pages in the report show data visualizations on the topic. The pages are split into an overview page, which combines all the data in the report, and into pages by data source.

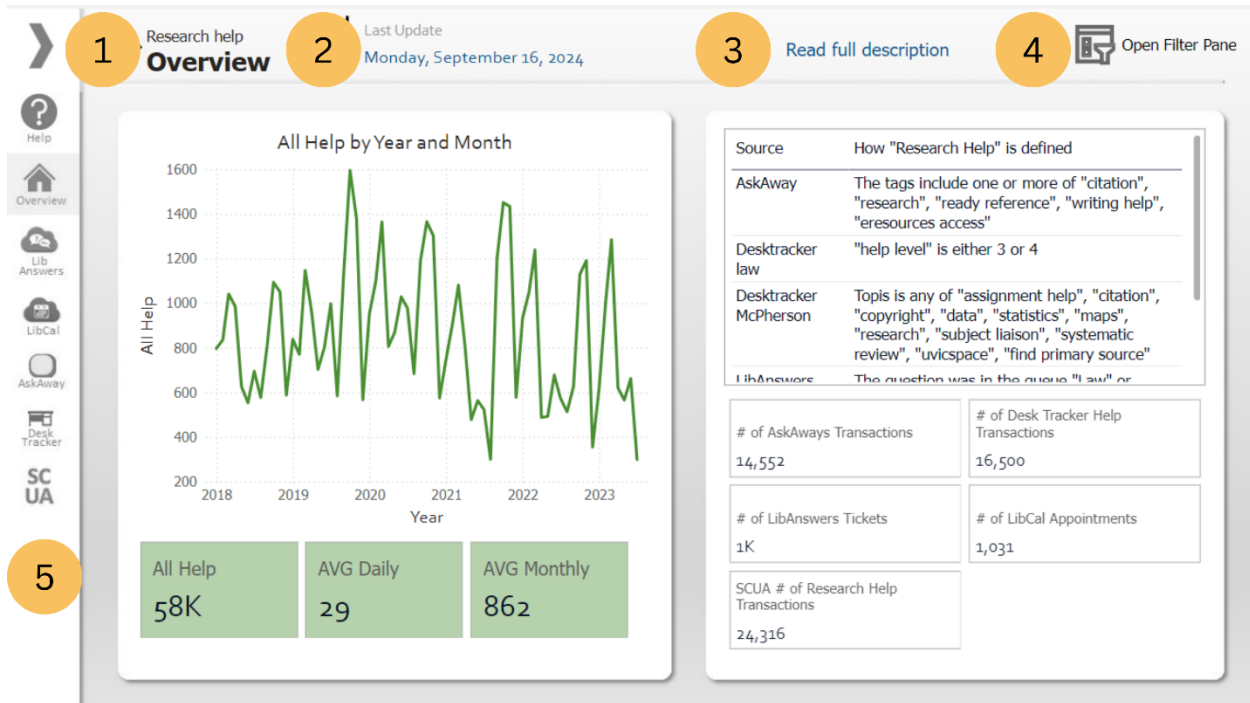


Figure 6. The Overview Page of the Research Help Dashboard at the University of Victoria Libraries

Each page in the dashboard has some consistent elements (Figure 6):

1. Dashboard name and page name
2. Date of last update
3. Description drop-down pane
4. Filter drop-down pane
5. Expanding navigation pane

Colour was used again as a feature drawing the eye to the most important parts of the page. The choice of each visualization is based on established best practices in the field of data visualizations.

Since dashboards are not just visualizations in isolation, but part of a larger user interface, there are some other best practices we followed. To ensure there is sufficient white space, pages have no more than 3 visualizations per page. For accessibility, and general readability, we follow the rule of minimal contrast ratio for text to background 4.5:1 and avoid bright colours, so that text is readable. Lastly, the pages automatically snap to full width so users do not have to scroll. Studies show that the top section of a webpage, or “above-the-fold”, gets the most attention.¹⁰

The Closet: The Dashboard Hub

The Dashboard Hub is a list of all dashboards that we have developed through Power BI. The list is made with Microsoft Lists. With the Grid View in Microsoft lists we can show a

preview thumbnail of the dashboard, add a title and a description as well as tags such as who the audience is for a dashboard (Figure 7). This becomes the one-stop-shop for UVic libraries data dashboards – in contrast to the original 34+ storage locations (Figure 1).

Dashboards

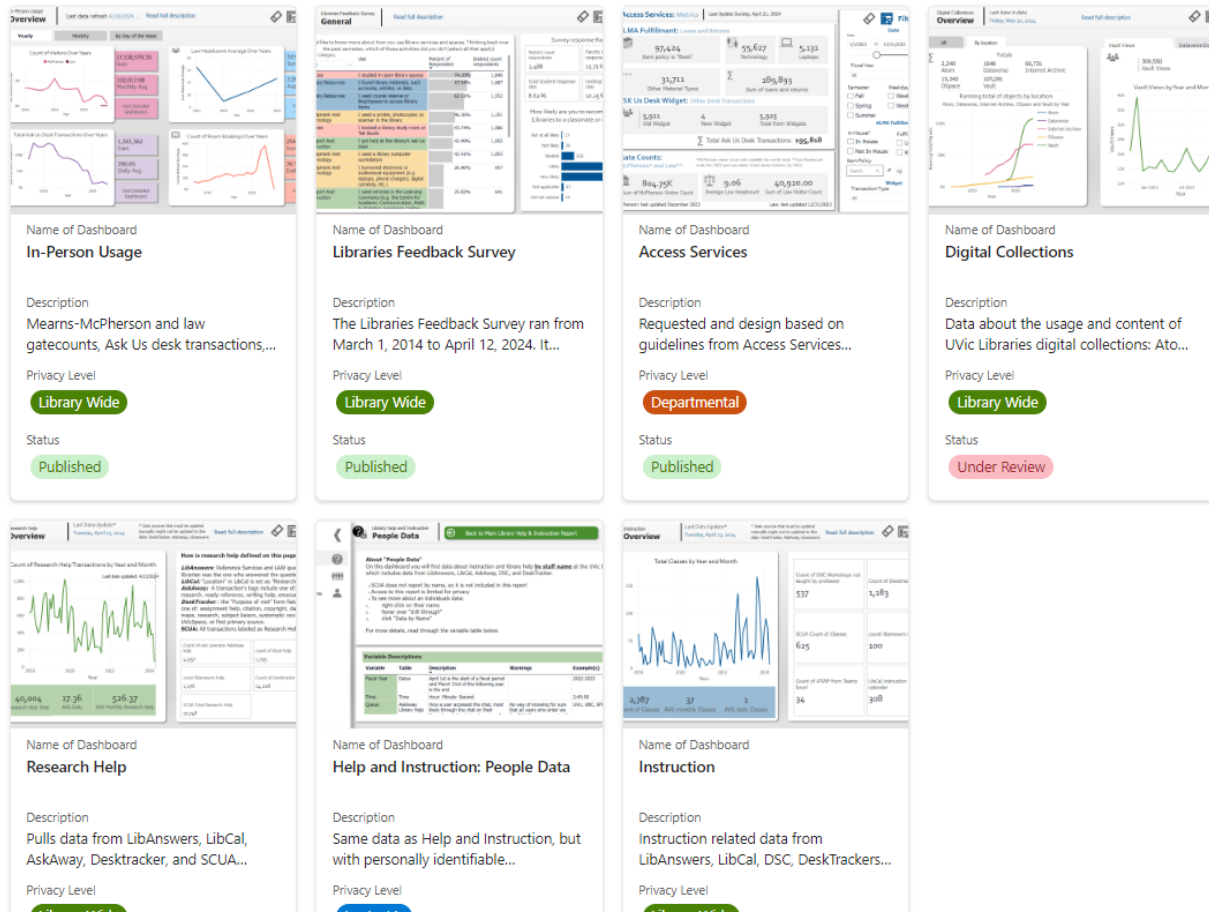


Figure 7. University of Victoria Libraries' Dashboard Hub for Organizational Data

Each dashboard has the same base for its user interface. The help page has the same elements, and each data page has the filters pane, drop down-description, date updated, and expanded navigation in the same locations. The consistent user interface freed users from learning a new interface completely every time they used a new dashboard.

Dropped Stitches: Mistakes and Lessons

Some aspects of our original plan did not work out as expected. We learned from those challenges and developed a new strategy that led to a much more successful product.

Inconsistent User Interfaces

The user interface of our dashboards was not always consistent. Figure 8 illustrates some of the interfaces that were developed before the transition to the consistent user interface. The colours, navigation, and shape of the reports were inconsistent, among other user interface elements. After observing users and consulting with experts, we realized that a unified interface would be much simpler for users to learn and understand.

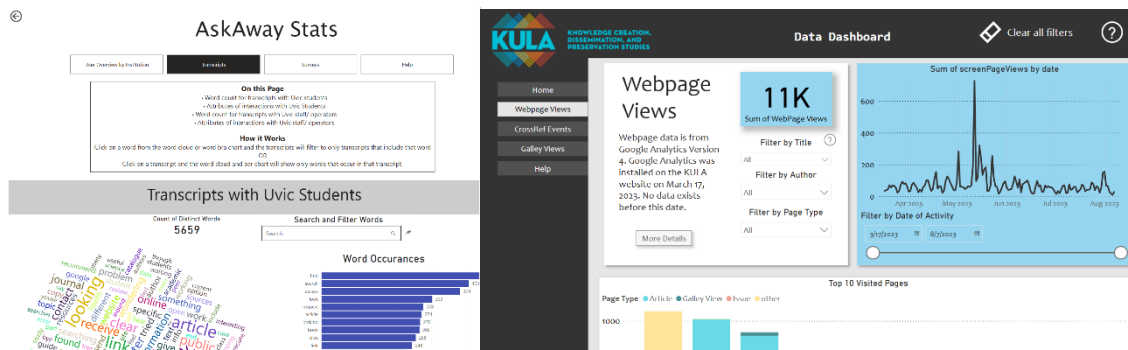


Figure 8. AskAway and Kula Data Dashboard User Interfaces Before Consistent Dashboard Redesign

Overly Complicated Data Visualizations

There were also some more experimental approaches to data visualizations, such as using a spiral bar graph and box and whisker charts (Figure 9). While neither of these charts are new inventions, they are also not generally recognized and ended up causing more confusion than expressing information. For these out of the box solutions, we realized choosing the well-known, and, according to data visualization research, best-practice visualizations to convey information.

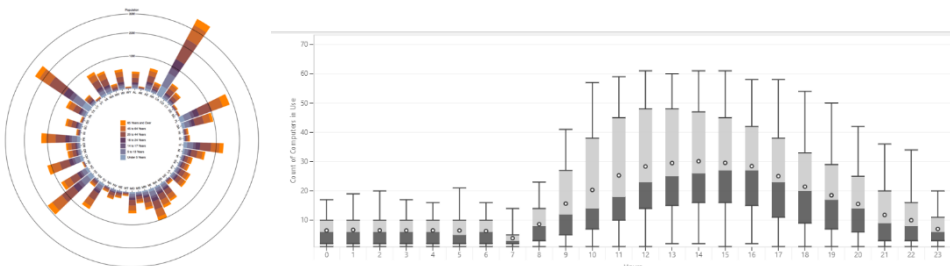


Figure 9. Spiral Chart and Box and Whisker Plot

Using simple visualizations is also important because not all the dashboard users will have the same data literacy. Still, there is no perfect solution to user interfaces that will work for everyone out of the box.

Cold Releasing Without Training

To mitigate some initial confusion, we started developing training and tutorials along with the reports. Each Dashboard has its own tutorial which users can click through. It uses speech bubbles, circles, and highlighting to walk through a dashboard (Figure 10).

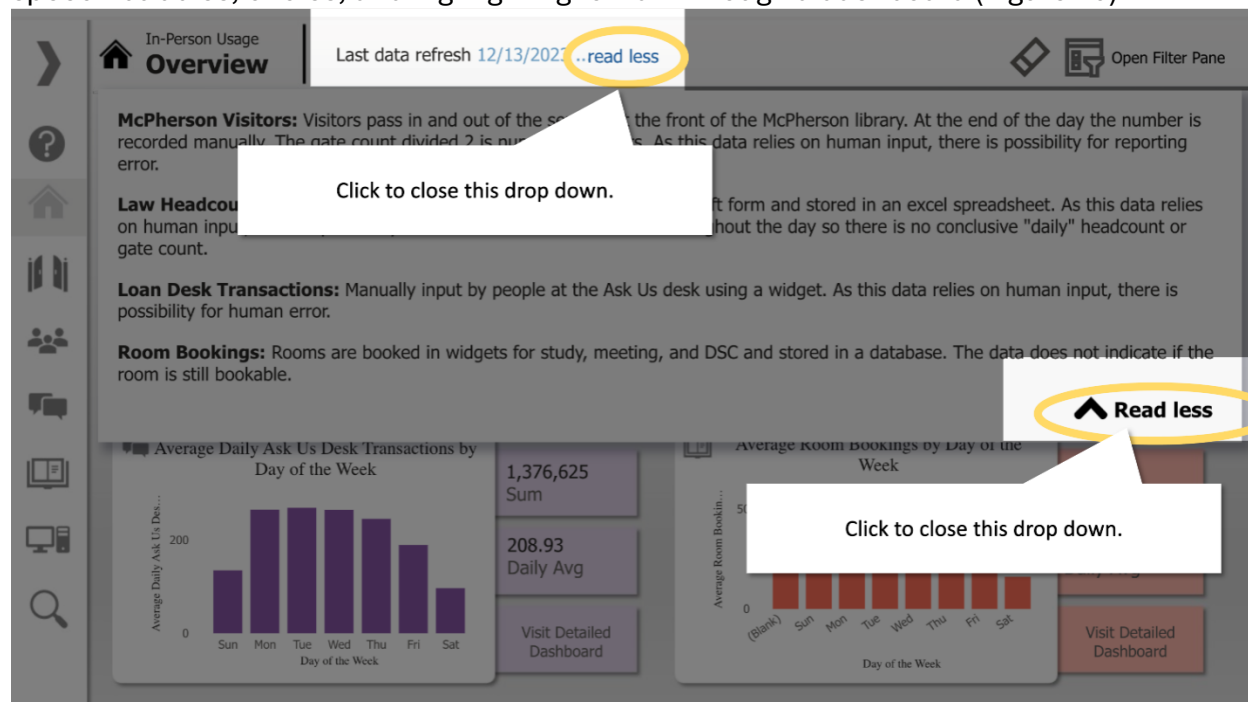


Figure 10. Example of a Page from a Dashboard Tutorial Made in PowerPoint

Power BI, while flexible, still has its own set of user interface features that need to be learned. These are generally consistent between reports, so we developed a [LibGuide on Power BI end-user features](#), since most existing online guidance is geared towards Power BI developers. We also created a workshop with hands on experience for Power BI, available to anyone, through our digital scholarship platform. This is run synchronously once a term, but is also possible to do asynchronously, as it is developed with the hyflex method in mind.¹¹

The Knitter: Subject Matter Expertise and Tools

This is not something that can be done off the side of someone's desk, or even a percentage of the role dedicated to the project. The UVic libraries has a dedicated full time professional employee dedicated almost entirely to this project. The expertise requires a technical background and knowledge of many technical tools and topics. The tools and topics that were mentioned in this paper and are used in the project over all are listed in Table 2.

Table 2. Subject Matter Expertise and Tools Required Used to Create the Data Dashboards for UVic Libraries

Resource	Knowledge Type	Specifics
Power BI	Software	Power query, Dax
Python	Software	Pandas, requests
APIs	Software	REST
Relational Databases	Software	MySQL, Solr
JavaScript	Software	D3.js
Data Visualization	Topic	
User Interface Design	Topic	

Next Steps

User Testing and Feedback

All our feedback so far comes from natural conversations and asking directly for feedback from users. We would like to run more official user tests like unmoderated usability tests where we ask participants questions and watch them complete tasks.

Advanced Analysis

As data becomes more important in proving our success as an institution, we would like to branch out from the out of the box reports and create more intricate and intertwined data analysis using the knitted data. This likely would not have as wide as an audience but be made for management. At this point, these kinds of analysis come at request, but we would like to have a larger ongoing and updating project. Luckily, once the data is added to Power BI service (the cloud version of Power BI) it is easy to access, like a database. It is already clean and in the right format to be used again in another Power BI report.

Conclusion

In conclusion, integrating data from multiple sources into a single interface using Power BI is a complex yet highly rewarding endeavor. This project underscores the critical importance of adhering to best practices in user interface and visualization design to ensure clarity and usability. Collecting feedback from user interface, user experience, and accessibility experts, as well as end-users, is essential for refining the interface and enhancing the overall user experience. The technical demands of this project are significant, requiring a deep understanding of data integration, visualization techniques, and Power BI capabilities. Given the complexity and scope, this initiative necessitates a dedicated full-time position to manage and execute effectively. By focusing on these key areas, libraries can leverage what we learned to create powerful, user-friendly data interfaces that drive informed decision-making and organizational success.

Endnotes:

- ¹ Megan Hurst, Christine Madsen, Frankie Wilson, MacKenzie Smith, and William F. Garrity, “All Your Data Displayed in One Place: Preliminary Research and Planning for a Library Assessment Dashboard and Toolkit,” in *Proceedings of the 2016 Library Assessment Conference: Building Effective, Sustainable, Practical Assessment, October 31- November 2, Arlington, Virginia*, ed. Sue Baughman, Steve Hiller, Katie Monroe, and Angela Pappalardo (Arlington, VA: Association of Research Libraries, 2016), 568, <https://www.libraryassessment.org/wp-content/uploads/bm~doc/proceedings-2016.pdf>.
- ² “(PDF) Business Intelligence Infrastructure for Academic Libraries.” *ResearchGate*, Oct. 2024, <https://doi.org/10.18438/B83G75>.
- ³ Mealey, Nathan. “Building a Scalable and Flexible Library Data Dashboard.” *The Code4Lib Journal*, no. 35, Jan. 2017, <https://journal.code4lib.org/articles/12152>.
- ⁴ Albers Szafir, Danielle, et al., editors. *Visualization Psychology*. Springer International Publishing, 2023, <https://doi.org/10.1007/978-3-031-34738-2>.
- ⁵ *Business Intelligence and Analytics Software | Tableau*. <https://www.tableau.com/>. Accessed 20 Jan. 2025.
- ⁶ Microsoft. *Use Python in Power BI Power Query Editor - Power BI*. 13 Feb. 2023, <https://learn.microsoft.com/en-us/power-bi/connect-data/desktop-python-in-query-editor>.
- ⁷ Microsoft. *Power Query Web Connector - Power Query*. 16 Feb. 2024, <https://learn.microsoft.com/en-us/power-query/connectors/web/web>.
- ⁸ Microsoft. *Custom Connectors Overview*. 13 Aug. 2024, <https://learn.microsoft.com/en-us/connectors/custom-connectors/>.
- ⁹ *Getting Started Creating Data Dictionaries: How to Create a Shareable Data Set*. <https://doi.org/10.1177/2515245920928007>. Accessed 20 Jan. 2025.
- ¹⁰ *Chapter 4: Building Above the Fold*. https://learning.oreilly.com/library/view/webflow-by-example/9781801075398/B17015_04_Final_JC_ePub.xhtml. Accessed 20 Jan. 2025.
- ¹¹ Beatty, B. J. “Designing a Hybrid-Flexible Course.” *Designing a Hybrid-Flexible Course*, 2019, pp. 35–47, https://edtechbooks.org/hyflex/hyflex_design.