Developing Strategies for the Harmonization of AHRQ Data Tools (Data Tools)

Prepared for:

Agency for Healthcare Research and Quality
5600 Fishers Lane
Rockville, MD 20857

Prepared by:

Atlas Research
805 Fifteenth Street NW, Suite 910
Washington, DC 20005

Developing Strategies for the Harmonization of AHRQ Data Tools (Data Tools)

Table of Contents

Foreword

This Working Paper is a compilation of the findings from six tasks—conducted between 2018 and 2019—that constituted the major steps in assessing the possibility of harmonizing AHRQ data tools. Minor changes have been made to the documents from their original versions. Specifically, links that are no longer available, temporary usernames and passwords, and photographs with identifiable persons have been removed. These changes have been noted at the beginning of each task.

This page intentionally left blank.

RFTO #10: Developing Strategies for the Harmonization of AHRQ Data Tools (Data Tools)

Project Report

July 15, 2019

CONTENTS

REVISION HISTORY

Date	Version	Description	Author(s)
June 14, 2019	Draft	Draft of the project report summarizing the results of the project tasks and outcomes	Atlas Team
June 28, 2019	Final	Updated with feedback received from AHRQ and added new content in Section 2.6.2	Atlas Team
July 15, 2019	Updated Final	This is updated version of the Final Project Report, incorporating the feedback received from AHRQ on July 9	Atlas Team

NOTE: The following document was produced as part of AHRQ's Data Tools Harmonization Project Assessment Phase. The original document contained links to wire frames and prototypes that are not available for viewing at this time. Those links have been replaced with the following highlighted text:

(link unavailable at this time)

1 INTRODUCTION

The Agency for Healthcare Research and Quality (AHRQ) provides several data tools and summary data tables that have dedicated user groups, but currently these tools are siloed and lack a common user experience (UX) and branding among them. The AHRQ data tools included in this project are:

- Consumer Assessment of Healthcare Providers and Systems (CAHPS)
- The Health Cost and Utilization Project's (HCUP) HCUPnet
- HCUP Fast Stats
- Medical Expenditure Panel Survey (MEPS) Summary Tables: Household Component (HC)/MEPSnet (HC)
- MEPS Summary Tables: Insurance Component (IC)/MEPSnet (IC)
- National Healthcare Quality and Disparities Reports Data Query Tool (NHQDR)

The project's goal is to evaluate the potential for harmonizing AHRQ statistical data tools so that they use a common or limited set of user interfaces, are clearly branded as AHRQ, are less costly to expand and maintain, and are more easily transferrable between vendors.

To accomplish this goal, the project executed 6 main task areas, each with its own set of deliverables.

Table 1. Project Deliverables

Task	Deliverable	Date Submitted
Task 1 – General Requirements		
B.1.1	Minutes of kick-off meeting and monthly meetings	Minutes of the kick-off meeting - submitted on October 24, 2018 Monthly Meeting Minutes were submitted on: - November 15, 2018 - December 14, 2018 - January 18, 2019 - February 15, 2019 - March 15, 2019 - April 19, 2019 - May 20, 2019 - June 14, 2019 - July 12, 2019 In addition, the project team held Bi-Weekly Meetings with AHRQ project leads during the following dates (minutes were also submitted for each): November 9, December 7, December 21, January 4, February 1, March 1, March 29, April 12, April 26, May 10, May 22, June 7, and June 28

B.1.2	Monthly progress reports	In coordination with the AHRQ project leads, Monthly Progress Reports were submitted on: - December 14, 2018 - January 16, 2019 - February 13, 2019 - March 13, 2019 - April 17, 2019 - May 16, 2019 - June 13, 2019 - The final monthly progress report will be submitted prior to the end of the period of performance (POP), which is July 15, 2019
B.1.3	Draft final report	June 14, 2019
B.1.4	Final report	July 15, 2019 (this is the most up-to-date Final Report based on feedback on the interim version of the final report that was submitted on June 28, 2019)
Task 2 – General Specifications		
B.2.1	Draft specifications for each AHRQ Tool	November 26, 2018
B.2.2	Final specifications for each AHRQ tool	An updated final version was submitted on December 28, 2018 after receiving AHRQ's feedback on the original final version that was submitted on December 14, 2018
Task 3 – Environmental Scan		
B.3.1	Draft plan for environmental scan	November 6, 2018
B.3.2	Final plan for environmental scan	November 28, 2018
B.3.3	Draft results of environmental scan	January 11, 2019
B.3.4	Final results of environmental scan	An updated final version was submitted on February 4, 2019 after receiving AHRQ's feedback on the original final version that was submitted on January 25, 2019
Task 4 – Product Assessment		
B.4.1	Draft list of products to assess	November 30, 2018
B.4.2	Final list of products to assess	December 17, 2018
B.4.3	Draft product assessment	January 25, 2019
B.4.4	Final product assessment	February 8, 2019
Task 5 – Style Examples		
B.5.1	Alpha version of style examples	February 8, 2019 (in addition, an accompanying report to supplement the deliverable was submitted)
B.5.2	Beta version of style examples	February 22, 2019 (in addition, an accompanying report to supplement the deliverable was submitted)

B.5.3	Final version of style examples	March 15, 2019 (in addition, an accompanying report to supplement the deliverable was submitted)
Task 6 – Prototype Tools		
B.6.1	Protype tools webpages Alpha	April 9, 2019 (in addition, an accompanying report to supplement the deliverable was submitted)
B.6.2	Prototype tools webpage Beta	May 3, 2019 (in addition, an accompanying report to supplement the deliverable was submitted)
B.6.3	Prototype tools webpage - final	May 31, 2019 (in addition, an accompanying report to supplement the deliverable was submitted on May 31, 2019 and an updated version of this report was submitted on June 14, 2019 based on AHRQ's final feedback)

2 PROJECT TASK AREAS

2.1 Task 1: General Requirements

During the kick-off meeting, the project team met with the Task Order Officer (TOO) and the AHRQ project leads to review the scope of work, validate assumptions, and confirm the project delivery schedule.

The project team then had regular formalized communications with AHRQ, supplemented with ad hoc communications, as necessary. The regular communication allowed the project team to update AHRQ constantly on project efforts to respond more effectively to challenges, capture priority requests, implement improvements in the process, and mitigate risks. These formalized communications included monthly progress reports, monthly project meetings, bi-weekly check-in meetings, and meeting minutes.

Monthly Progress Reports were submitted to AHRQ and included the following information: 1) summary of monthly activities, 2) a summation of cost and level of effort expended 3) preliminary or interim results, 4) issues or risks and their mitigation strategies, 5) project schedule status and any proposed adjustments and 6) planned activities for the following month. It also included a list of action items.

Throughout the contract period, the project team held monthly meetings with the POCs to discuss progress as well as raise and discuss any questions or issues. Most of the monthly meetings were held in-person at AHRQ. Meeting agenda and minutes were prepared for each monthly meeting.

Bi-weekly check-In meetings were also scheduled with AHRQ to provide additional and more frequent formalized communication between the project team and the main project stakeholders. Meeting agenda, minutes, and list of actions items were prepared for each of these meetings.

This Project Report is also part of this task area. This report includes a summary of the process, key activities, and outcomes for all of the project task areas.

2.2 Task 2: General Specifications

To get better insight into both the value to be gained and the potential challenges facing the harmonization of these data tools, the Atlas Team conducted an environmental scan to better understand the unique and varied needs of the tools' users and stakeholders. A facet of this environmental scan included gathering and outlining the current technical specifications for each AHRQ data tool. The technical specifications were gathered through a survey that asked for details on the hardware and software components of the data tools, along with the necessary processes to extract, transform and load data into the tools.

The purpose of this task area was to gain a general understanding of the current technical environment and needs of AHRQ's various data tools and to determine the common specifications needed to develop a possible harmonized look and feel across the various data tools used by the agency.

The project's approach to reviewing the technical needs and capability of the AHRQ data tools consisted of collecting, reviewing and matching/comparing all government-provided technical specifications. The scope of the government-provided technical information was focused on general, high-level technical specifications that can be used as basis to conduct, compare, and recommend possible commercial off-the-shelf (COTS) products that are readily available in the industry.

The inputs for the technical specifications were sourced from primary research. The Atlas team used a technical specifications template for each data tool as guidance when interviewing the data tool's POCs. Following the interview, the Atlas team sent a draft version of the template with completed answers when applicable, which the POCs validated and completed.

The government-provided technical specifications that were collected centered around the internal/operational components that make government statistics accessible to the public. These included the databases, servers, and other operational components. The Technical Specifications Report that was submitted provided information on component type, versions, sizes and the other general technical information as provided by the government.

The specifications summarized below are the core technical requirements that were considered when assessing the products that can deliver the required features of the data tools and harmonized interfaces.

In Figure 1 (Summary of Technical Specifications), the technical specifications for each tool are visualized and color-coded. For example, CAHPS (colored brown) has Windows for an Operating System, utilizes ASP.NET as a Programming Language, receives SQL load files for updates, has a SQL server for its database server, and uses a Virtual Machine for its Web Server.

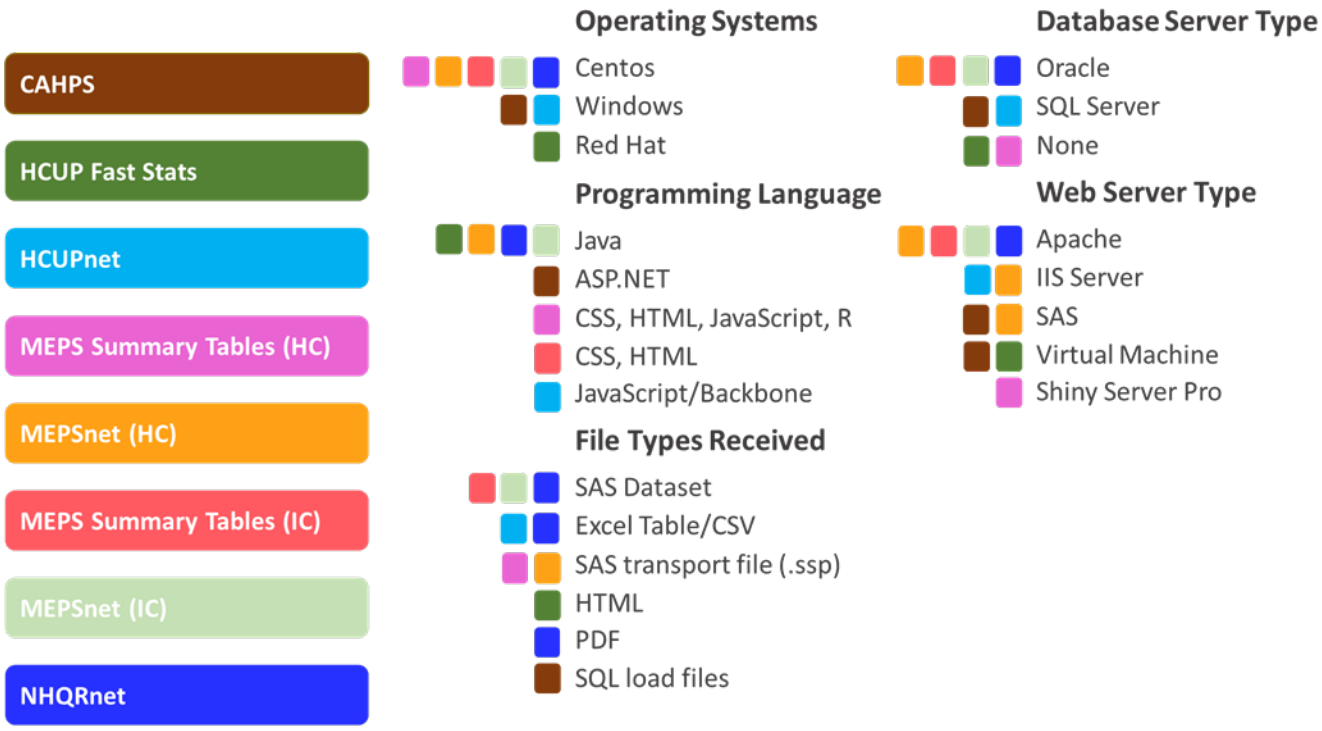


Figure 1. Summary of Technical Specifications

The technical specifications as well as the functional requirements gathered in the Environmental Scan Report (output of Task 3 Environmental Scan phase) were used in assessing products that can deliver the required features of a harmonized data tools interface and help AHRQ determine the data visualization products to build the Alpha-version of the prototypes on.

For more details about this task area, please refer to the report that was submitted to AHRQ, "**AHRQ Data Tool Harmonization Technical Specifications 12.28.2018**".

2.3 Task 3: Environmental Scan

To better understand both the potential for and challenges facing the harmonization of the user interfaces (UI) and basic query functionality of these data tools, the Atlas Team conducted an environmental scan. The purpose of this scan was to understand the unique and varied needs of the tools' users and stakeholders. Between November 2018 and the end of January 2019, the Atlas team interviewed 26 users that were classified as External Users (i.e., those who work outside of AHRQ at other institutions or agencies), Internal Users (i.e., those who work at AHRQ), and Internal Stakeholders (i.e., AHRQ employees who have a direct role in managing or supporting one or more of the data tools that are part of the Data Tool Harmonization Project). These interviews covered users' experiences and thoughts on the current user experience with a data tool, their awareness or use of other data tools, the branding of these tools, and any barriers or opportunities they see from harmonizing these data tools. In addition to discussing these topics, we also observed these users as they completed a typical query in order to witness how they operated the data tool and used the outputs to inform their work. Additionally, we had users conduct a Card Sort which allowed us to understand users' mental models of how information is organized and prioritized.

The results of this qualitative research were analyzed and synthesized into themes of work for each user and user mindset (e.g., persona). Specifically, two user groups, and five user types were identified: External Users Group: Pro, Intermediate, and Novice / Quick Reference; Internal Users Group: Support Users, and Administrators. In addition to identifying insights into the mindsets and needs of key users, we also identified and logged user and system requirements we learned throughout the interviews that was helpful in picking out data visualization partners to review in the Product Assessment Report, inspired our designs in the Style Examples, and ultimately informed which data visualization partners to include in the prototype phase as well as which design elements to prioritize for this project. On January 15th we presented the penultimate findings of the Environmental Scan activities along with a collaborative workshop to help prioritize findings and user requirements (see Figure 2 below).

Figure 2. AHRQ Mid-Phase Presentation and Workshop

For more details about this task area and for the results, please refer to the report that was submitted to AHRQ, "**AHRQ Data Tools Environmental Scan Results Report 02.04.2019**".

2.4 Task 4: Product Assessment

To provide AHRQ a harmonized UI for its multiple data tools, it requires a COTS data visualization product that satisfies the overlapping needs of backend and frontend users.

To determine the appropriate COTS data visualization product, the Atlas Team devised a list of 15 potential solutions to assess. After narrowing the list down to 9, the project further divided these products into 4 High Priority products and 5 Low Priority products for assessment.

Table 2. List of Data Visualization Products

High Priority	Low Priority
MicroStrategy	Birst
Qlik	Domo
Shiny	Looker
Tableau	Microsoft PowerBI
	Sisense

To understand the full capability of each tool, and its ability to satisfy AHRQ's unique needs, the Atlas Team conducted primary and secondary research.

The primary research centered around demonstrations of the Data Visualization Products with company representatives, followed by structured questions from a discussion guide to standardize the topics covered in each product demonstration. These discussions included a request for any Federal examples of the product in use, and in a limited number of circumstances (e.g. Qlik), the Atlas Team secured a demo version of the product for follow up research.

Part of the primary research with the High Priority products also involved fielding estimates for the cost of implementing the products. The Atlas Team standardized its approach in fielding estimates by requesting quotes for embedding the COTS product in a public facing website that sees an average of 75,000 hits per month and has backend access for 10 super users. The Atlas Team used "75,000 users" in our requests for estimates based on the average of the estimated usages provided by the AHRQ team for the data tools.

All potential COTS product teams provided us with estimates and stressed that these were ballpark estimates based on the limited information we provided. Moreover, many of them expressed a willingness to 1) undergo capacity planning with the AHRQ team to determine the optimal amount of infrastructure for the product to work as expected for all uses and 2) accommodate the AHRQ's team budgetary requirements.

The Atlas Team also conducted secondary research involving online research to understand any potential pitfalls for implementing a Data Visualization Product. When possible, the Atlas Team utilized the same source of knowledge to ensure consistency in perspective, such as Gartner's "Critical Capabilities for Analytics and Business Intelligence Platforms".

2.4.1 Summary of Assessments of the High Priority Data Visualization Products

The project team's Product Assessment Report contained assessments for each of the data visualization products. For the high priority products, the project team provided a product overview, hypothetical product implementation, Federal examples of implementation of the product, and final takeaways for each. The project team also provided a summary table containing the strengths, weaknesses, and cost estimates for each.

Table 3. High Priority Data Visualization Products for Assessment

Product	Strengths	Weaknesses	Costs Estimates
MicroStrategy	• High touch implementation team when setting up MicroStrategy. • Advanced visualization engine available to backend users. • Ability to scale across all AHRQ products.	• Approximately 2x more expensive than current competitors' estimates.	If Perpetual License • Base Cost: $350,000 • Annual Maintenance Cost: $58,000 • Average Annual Cost for a 5-year period: $116,400 If Term License • Annual Cost: $150,000 MicroStrategy is willing to work with AHRQ team to accommodate budgets.
Shiny	• Approximately half the cost of Tableau, and 80% less than MicroStrategy. • Entirely customizable by the AHRQ team. • Lowest impact on AHRQ's existing environment.	• Steep learning curve for backend users with no R / Shiny experience. • Architecture impacts ability to easily scale across all AHRQ data tools.	Total Annual Cost: $38,300 • RStudio Shiny Server Pro: $10,000 • RStudio Connect for 20 Users: $15,000 • RStudio Package manager: $10,000 • Shiny Apps Professional: $3,300
Tableau	• Product architecture can scale across all AHRQ products. • Intuitive graphics for creating visualizations.	• Reliance on self-service learning appeals to a limited set of administrators.	Total Annual Cost: $72,000
Qlik	• Experience with Federal clients. • Compelling frontend graphics.	• Less intuitive backend user experience.	Total Annual Cost: $80,300

From among the data visualization products that were assessed, AHRQ selected the two products that will be used to create the Alpha prototypes. These two products were **Qlik** and **Tableau**.

2.4.2 Qlik

Table 4. Summary of Assessment - QLIK

Strengths	Weaknesses
Experienced with Federal clients.	Less intuitive administrative backend.
Compelling frontend graphics	Issues in scaling with large, disparate datasets.

Qlik's current analytics tool, Qlik Sense, is a SaaS Data Visualization Product and has a similar offering as Tableau. It can be deployed either on the cloud or on premises and uses a metadata layer to house data from various sources to display to frontend users. It employs a backend UI that is less intuitive than Tableau and utilizes a SQL-like language to support the data loading process.

Qlik's strongest selling point is Qlik's experience in the Federal space and the compelling graphics it offers for frontend users. Qlik averages approximately 4 releases yearly, which include a mix of product enhancements, such as new visualizations and accessibility additions, and bug fixes.

Once installed, Qlik uses a semantic layer managed through a network of nodes that store the data, run analysis requested in the front facing app, store the data, and schedule uploads of the data. Qlik is able to utilize out-of-the-box connectors and can support custom connectors where out-of-the-box options are not available.

Visualizations are created through dragging and dropping data fields into Qlik's dashboard environment. Additional customizations to the visualizations are provided by a SQL-like language, and most of AHRQ's visualizations will require use of this language, making the learning curve steeper than Tableau. For map visualizations, Qlik requires an in-house application for geospatial data called "Qlik GeoAnalytics", which was introduced only in 2017. Qlik backend users have mentioned the platform's rigidity requires creating custom hacks to provide the frontend user with exactly what they're looking for. For additional functionality that is not provided by Qlik in-house, there are options to purchase an extension to expand Qlik's functionality. These extensions have their developer and support teams and should be treated as a separate product from Qlik.

Qlik offers subscription licenses for Qlik Sense and estimated an annual cost of $83,000 based on our scenario. However, the Qlik team emphasized that a special enterprise license agreement (ELA) or blanket purchase agreement (BPA) might exist for Qlik with HHS or AHRQ. The quoted price is based on Government Services Administration (GSA) pricing.

2.4.3 Tableau

Table 5. Summary of Assessment - Tableau

Strengths	Weaknesses
Established brand and company.	Mild learning curve.
Extensive backend userbase with good resources for self-training.	Some hacks required for certain customizations.

Easy connection between metadata and data extracts.	
Experienced with Federal clients.	
Large visualization library.	

Tableau is a SaaS analytics product whose popularity has dramatically increased recently due to its self-service capabilities and its ability to quickly and easily provide visualizations for frontend users. Tableau benefits from having an extensive userbase while also having experience in the Federal space. The larger userbase provides a network for collaboration or idea sourcing, while the experience in the Federal space secures Tableau's brand as a trusted product.

Nevertheless, backend users of Tableau who are not involved in the granular implementation details will experience a learning curve in getting started, albeit less of a learning curve than Qlik. Additionally, while Tableau has high regard as a quick-to-deploy SaaS solution, some users complain that the product can be rigid in delivering precise and unique customizations. What compounds this issue is the support staff for Tableau is less helpful than other solutions, since it relies on a self-service model. Although the userbase can act as a hive mind to develop solutions for common Tableau problems, it will depend on the backend users to uncover solutions to Tableau problems. Nevertheless, for intuitive visualizations coupled with an easy-to-use intuitive interface, Tableau provides a compelling offering for AHRQ.

Out of all the products assessed, Tableau has the most frequent release schedule, averaging a release once every 3 weeks. These frequent releases include bug fixes, while product enhancements are released approximately once or twice per year.

Tableau offers an Unlimited user model for applications that will be embedded in public facing websites. To accommodate the number of users on a concurrent basis, the Tableau team recommends a Core model, which is hardware-based. The Tableau team estimated the Core model would cost $72,000 per year based on the assumption of roughly 75k monthly users. The Tableau team emphasized that part of the implementation process will involve scoping out a more accurate capacity than the current estimate.

To setup Tableau Core, the Tableau team will work with the AHRQ team to scope out the number of CPUs required to accommodate the usual traffic to AHRQs data tools. Tableau is used to working with large enterprises that display their data for public consumption, and as such have an unlimited user model that support AHRQ's scenario.

Once a Tableau solution is deployed and administrator accounts have been provisioned, administrators determine which data are connected to a Tableau metadata semantic layer, allowing administrators to create visualizations. Administrators, with Tableau Creator licenses, use simple drag-and-drop, or use an easier-than-SQL code, to select fields to visualize. Tableau can accommodate a variety of database servers to connect to this drag-and-drop solution by using out-of-the-box connectors for popular database servers such as Oracle or SQL. For database servers where an out-of-the-box connector is not available, the Tableau team will work to create an ODBC Connector.

Multiple data sources can be visualized simultaneously in this sandbox environment and published for public use. To determine the correct visualization, Tableau automatically recommends visualizations based on the format of the data selected by the administrator. For

further visualization inspiration, Tableau offers best practices on their website and have an extensive userbase from where administrators could source new ideas.

For more details about this task area, please refer to the report that was submitted to AHRQ, "**AHRQ Data Tool Harmonization Product Assessment Report 02.08.2019**".

2.5 Task 5: Style Examples

The Style Examples Final Report was delivered on March 15th, 2019 as an aid to help stakeholders review the Style Examples provided through InVision, a program that allows designs to be interactive, simulating the intended user experience. The Style Examples are color wireframes that provide new flows for the website experience from the AHRQ Data Tool Landing Page to each Data Tool Landing Pages, and finally the Data Tool Query Pages as well as new information architecture with the intent to harmonize the look, feel, and experience of the different data tools while supporting the unique data types and visualizations required by each tool.

The Style Examples designs were developed in three phases: Alpha, Beta, and Final. The Alpha Style Examples were initial concepts for a harmonized data portal, user interface, and functionality of the data tools. The goal for the Alpha version was to establish a common look and feel so users can focus on understanding all the data tools that AHRQ provides and how to find the right data for their research needs. This was also the earliest point of incorporating feedback from stakeholders. AHRQ provided all feedback on the Alpha Style Examples on February 15, 2019.

The Beta Style Examples were revised designs that incorporated the feedback received from AHRQ on the Alpha designs as well as additional designs for the individual tool landing and query pages. From February 25 – March 4, AHRQ POCs and representatives provided feedback during one-hour sessions with the Atlas Team. In addition, the Atlas team held feedback sessions with several external customers of the data tools. Before starting the final style designs, the project team worked with AHRQ leaders to prioritize the pages, requirements, and feedback to include.

The Final designs incorporated all the previous feedback and was intended to serve as the basis for building out the prototypes. The prioritized requirements and page elements decided upon in partnership with the Office of Communications, the AHRQ project lead, and the Atlas team included: implementing the appropriate header and footer and using the prescribed logos and style guidelines (if specified) for AHRQ, the Department of Health & Human Services (HHS), and each individual data tool. In addition, the Atlas Team was provided guidance on the fonts to use for the style examples and prototyping. The Atlas Team was given flexibility by AHRQ project lead to design the site features to best support the creation of prototypes that will enable AHRQ to evaluate the feasibility of a common data visualization product and a harmonized user interface.

While not all the design decisions included in the Final Style Guide Report went forward into the prototype, these designs are still discussed and slated for follow-on projects as AHRQ works to build and launch the harmonized data tool experience. The Final Style Guide Report and related conversations did inform the information architecture and general layout and experience of the Data Tool Query pages.

For more details about this task area and to see the designs, please refer to the report that was submitted to AHRQ, "**AHRQ Data Tool Harmonization Style Examples Report**". The designs that show the full features of the Style Examples are reachable in InVision through this link: (link unavailable at this time)

2.6 Task 6: Prototype Tools

After evaluating the potential for harmonizing AHRQ's statistical data tools through previous tasks (General Specifications, Environmental Scan, Product Assessment, and Style Examples), the project team developed prototypes to demonstrate how a selected COTS data visualization product could deliver the important features of these data tools to the public.

During the Alpha prototype phase, the project team developed a select number of visualizations for a limited number of AHRQ data tools (specifically MEPSnet HC, HCUPnet and CAHPS) for both Tableau and Qlik. The results of the Alpha prototype would inform the selection of a COTS product, and the Beta prototype phase would be a continued build out of all AHRQ data tools with the selected data visualization product. For the Beta and final versions of the prototype, the AHRQ project leads and POCs selected to proceed with **Tableau**.

The project team held iterative work sessions with the data tools' stakeholders through the Beta and final phases of the development of the prototype. From the feedback received during the work sessions, the project team developed each data tool's data visualizations as seen in the current prototype.

Table 6. Summary of Prototype Data Visualizations – CAHPS

Data Tool: CAHPS					
Data Visualization	**Graph & Tabular Data**		**Map**	**Standard Errors Display Option**	**Download/Share**
	Line/Trend	**Bar/ Cross-Section**			
Clinician & Group Survey	N/A	x	N/A	N/A	x
Health Plan Survey	N/A	x	N/A	N/A	x

Table 7. Summary of Prototype Data Visualizations – HCUPnet

Data Tool: HCUPnet					
Data Visualization	**Graph & Tabular Data**		**Map**	**Standard Errors Display Option**	**Download/Share**
	Line/Trend	**Bar/ Cross-Section**			
Inpatient	x	N/A	x	x	x
Emergency Department	x	N/A	x	x	x
Ambulatory Surgery	N/A	N/A	N/A	N/A	N/A
Community	N/A	N/A	N/A	N/A	N/A

Table 8. Summary of Prototype Data Visualizations – HCUP Fast Stats

Data Tool: HCUP Fast Stats					
Data Visualization	**Graph & Tabular Data**		**Map**	**Standard Errors Display Option**	**Download/Share**
	Line/Trend	**Bar/ Cross-Section**			
State Trends in Hospital Use by Payer:					
Inpatient Stay	x	N/A	x	N/A	x
Emergency Department Visit	x	N/A	x	N/A	x
National Hospital Use and Costs:					
Trends in Inpatient Stays	x	x	N/A	N/A	x
Most Common Diagnoses for Inpatient Stays	x	x	N/A	N/A	x
Most Common Operations During Inpatient Stays	x	x	N/A	N/A	x
Opioid-Related Hospital Use:					
Trends in Inpatient Stays and Emergency Department Visits	x	N/A	x	N/A	x

Table 9. Summary of Prototype Data Visualizations – MEPS HC

Data Tool: MEPS HC					
Data Visualization	**Graph & Tabular Data**		**Map**	**Standard Errors Display Option**	**Download/Share**
	Line/Trend	**Bar/ Cross-Section**			
Use, Expenditures, & Population	x	x	N/A	x	x
Health Insurance	x	x	N/A	x	x
Accessibility & Quality of Care	x	x	N/A	x	x
Medical Conditions	x	x	N/A	x	x
Prescribed Drugs	x	N/A	N/A	x	x

Table 10. Summary of Prototype Data Visualizations – MEPS IC

Data Tool: MEPS IC					
Data Visualization	**Graph & Tabular Data**		**Map**	**Standard Errors Display Option**	**Download/Share**
	Line/Trend	**Bar/ Cross-Section**			
Private Sector - National	x	x	N/A	x	x
Private Sector - State	x	x	x	x	x
Public Sector	x	x	N/A	x	x
Civilian Data	N/A	N/A	N/A	N/A	N/A
National Totals	N/A	N/A	N/A	N/A	N/A
Private Sector by Metro Area	N/A	N/A	N/A	N/A	N/A

Table 11. Summary of Prototype Data Visualizations – NHQDR

Data Tool: NHQDR					
Data Visualization	Graph & Tabular Data		Map	Standard Errors Display Option	Download/Share
	Line/Trend	Bar/ Cross-Section			
National	x	x	N/A	x	x
State	x	x	x	x	x

Use this link to go to the Prototype: (link unavailable at this time). This link will take you to the Prototype Introduction Page. Click on the "Portal" link on the upper right-hand side of the screen. Please note that you may need the appropriate login credentials to see the Tableau data visualizations.

Please note that:

- **The Prototype is conceptual and exists to support AHRQ staff in evaluating the feasibility of a harmonized data tool portal using Tableau.**
- **The Prototype is for demonstration purposes only and will not be for public view or use.**
- **As of 7/9/2019, the Prototype has been transferred to AHRQ servers so that it can be viewed and used internally for an unspecified period of time.**
- **The website portal landing page is now simplified to encourage focus on the actual data visualizations that are shown in the prototype.**

For more details about this task area, please refer to the report that was submitted to AHRQ, "**AHRQ Data Tool Harmonization Prototype Report 06.14.2019_Final**".

As requested by the AHRQ team, the following sections are included to provide a high-level overview of the process the project team undertook for creating the dashboard visualizations. In addition to this process overview, the Atlas team was also tasked to write up details of some features that may need to be improved on or implemented during the development of the actual AHRQ data tools portal and visualizations.

2.6.1 Visualizations Process Overview

To replicate the process for creating the visualizations, the project team utilized the following steps to create the visualizations in the current prototype

- **Prepare the data**: Before loading the data into Tableau, the project team used Excel for minor changes to the data tables. For data manipulations that involved more complicated manipulations, the project team used Tableau Prep, which Tableau creators have access to. Tableau prep allows the Creator to merge multiple datasets into one file

through joins or other manipulations in an easy-to-use GUI. Creators can use Tableau desktop for on the fly changes, however it's recommended to keep a consistent single source of data for versioning control. For future implementation, it is recommended to complete the majority of data manipulation prior to importing data into Tableau in order to maintain consistency.

- **Load the data:** Once the data is prepared, the project team loaded the required datasets for each Tableau dashboard. For most visualizations it's a simple direct relationship between the file and which data needs to be loaded. However, Tableau does allow for multiple data files to be loaded in a Tableau Workbook. If a Tableau Workbook requires multiple files, it's recommended that this is planned out before creating visualizations. For future implementation, it is recommended to merge files prior to importing into Tableau, in order to maintain a 1-to-1 relationship between a Tableau workbook and input file.

  Tableau also supports connecting to data that may be stored in a wide variety of places, from a spreadsheet on a computer to a more complex relational database on a server (e.g., a SQL server database). As an alternative to loading the data for the Tableau dashboards, Tableau can also connect to data sources leveraging a live connection via the security protocols set up on the data source.

- **Build out visualizations in worksheets:**

  - **Trend Line Chart**: The trend visualization requires that the x-axis (Columns) contains the Year dimension. The Year dimension should be in date format, however, only displaying the year—Tableau will default to a MM-DD-YYYY format, but this can easily be changed. For the y-axis (Rows), load the Measure that should be graphed (e.g., 'Coef' or 'value'). To achieve the double-slider for a year range, year must be categorized as a number, as the double-slider is only available to numeric filters. This numeric filter will not compromise any functionality in the trend line.

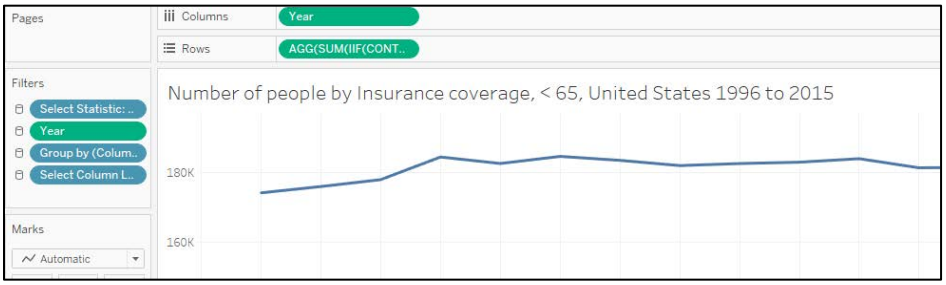


Figure 3. The Data Required to Display Tableau's Trend Line Feature

IMPORTANT: Since most of AHRQ's data visualizations are based on summarized data, filters must be added so that a single series of data is presented in the trend line, otherwise data might be unintentionally aggregated.

  - **Cross Sectional Bar Chart**: Cross Sectional visualizations differ from the trend line chart. First, the user must determine the x-axis (Columns) value of interest (e.g. Coef). The y-axis (Rows) will need to consist of as many groups and

subgroups needed so that one series of data is presented. Additionally, the year filter should only allow for one value selection at a time.

- **Map Visualization**: Map visualizations differ from both trend and cross-sectional visualizations, as the x (Rows) and y-axes (Columns) will be represented by longitude and latitude respectively. To do this, the visualization requires one dimension to be formatted with a geographic role. For example, a dimension such as "State" will need to be assigned the State geographic role, and then Tableau will create Latitude and Longitude data automatically. Tableau automatically graphs out a map with this information. The Creator should then take the Dimension with the geographic role and add it to the "Detail" option in the Marks section. Additionally, the Creator will need to add the value being visualized to the "Color" option in the Marks section. Like cross-sectional and trend visualizations, sufficient filters should be in place to avoid unintentional aggregation.

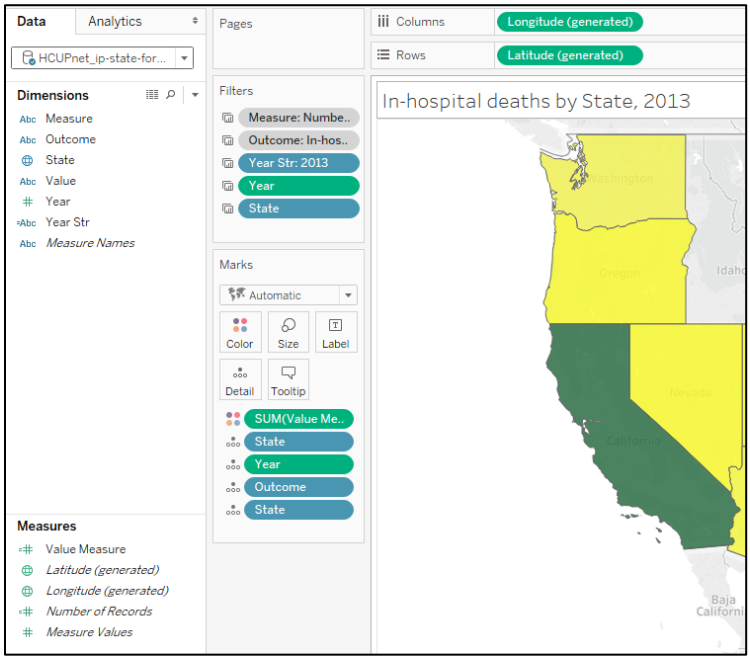


Figure 4. The Data Required to Display Tableau's Map View

- **Standard Error / Confidence Interval Visualization:** To add confidence intervals to vertical bar charts, we duplicated the graphical visualization worksheets. Multiple calculated fields will need to be created for the confidence interval bands (lower band, upper band, and the difference between the two). This can also be accomplished prior to importing the data into Tableau. Add the confidence interval measures as an additional measure in the rows for the Trend Line, or as a reference line for the bar chart. For the trend line, make the second measure "select dual axis" and then "Synchronizes axes". The bar chart's reference line, however, should be a "band" reference line that bands from lower confidence limit ("Standard Errors Lower" in Figure) to the upper confidence limit ("Standard Errors Upper" in Figure). We used the following tutorial to achieve this: https://kb.tableau.com/articles/howto/creating-error-bars.

For horizontal bars, we were able to utilize a service in Tableau called "Reference Line". The reference line allows backend users to add a line, band, distribution, or box plot across the graph area based on predefined measures. To do this, the user must add the measures to the graph area first under "detail" in the Marks area of the graph.

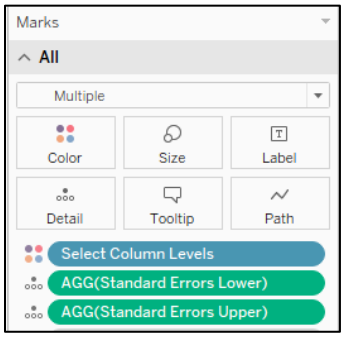


Figure 5. Standard Error Data in Tableau's Marks Section to display Confidence Intervals

After that, the user right-clicks on the x-axis and selects "Add Reference Line". The menu below will display on the screen showing the multiple options available.

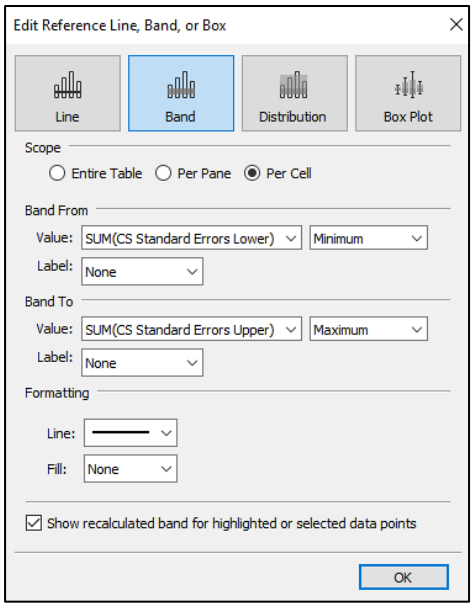


Figure 6. Tableau's Reference Line Options Available

After selecting band, choose the measures that were dragged into the detail option of the Marks area. For formatting, we chose to have no fill, which will not display the band area for the front-end user. The screenshot below provides an example of the result.

Figure 7. An Example Bar Chart with Confidence Interval

We encountered issues when using the band with trendline visualizations. While we hoped to replicate a band that covers the confidence interval area for each year in the trend, it appears instead that Tableau aggregates the measure over all years, so the band ends up covering the minimum confidence limit of all years and the maximum confidence limit of all years. This provides more of a box than a trend line standard errors area. However, Tableau does allow for parameters to be used in the measure, so that might allow future backend users to program logic to display the min/max confidence intervals for every year.

- **Data Tables**: To convert the visualizations into tables, the easiest option available is to duplicate all visualization worksheets and select the table option under Tableau's "Show Me" tab. From there, the creator can reformat as needed.

Figure 8. Tableau's Show Me Feature

- **Create Dashboards**: Once the worksheets are created, the creator can click the ideal number of worksheets to each dashboard. Tableau will automatically add the filters that pertain to each visualization; however, the Creator can add or remove any filter. While the Creator is making a Dashboard, he/she can also add buttons to navigate between the dashboards. For example, the Standard Errors "toggle" is technically a button that navigates a frontend user to a Tableau dashboard with visualizations that include the confidence intervals on the Figures and standard errors in the Table. The default visualization for the button is below.

Figure 9. Tableau's Default Button Format

This can be configured to any image. Moreover, the background color of the button can be configured, which is the route we took to create the ability to navigate to other visualizations.

Figure 10. A Customized Button to Direct Users to a Map View

- **Naming Protocols**: As visualizations are being made, the creator should keep careful track of the name of the worksheets and dashboards. While the names of the worksheets don't display to the end user, the names of the dashboards display to the frontend user in certain circumstances, such as when the frontend user exports the dashboard to PowerPoint. Moreover, the Creator should keep a naming protocol for Worksheets and Dashboards to distinguish between the two. For the worksheets, we recommend including an abbreviation of the data tool's name (e.g., "MEPS HC"), an abbreviation for the subject that the visualization covers (e.g. "cond" for condition), an abbreviation for the visualization (e.g. "trend" or "table") and any other modifiers that are required (e.g. "SE" for Standard Errors, "HI" for Hawaii). This would mean "MEPS HC cond trend" would be the trend visualization for the MEPS HC Conditions data tool. We recommend adding modifiers to the beginning of the name to more easily distinguish them from other files.

  When naming the dashboards, we strongly recommend following the same naming convention for the worksheets while adding a "D" at the beginning of each name. The above example's dashboard counterpart would be "D MEPS HC cond trend". This "D" helps differentiate the dashboard from the worksheet to ensure that only dashboards are published.

- **Publish Dashboards**: Once the dashboards pass internal validation, they should be published to the Tableau server (either available in the cloud or on prem). The Creator should only publish Dashboards, and not worksheets, to Tableau. A dashboard cannot be reviewed by another person until it's published.

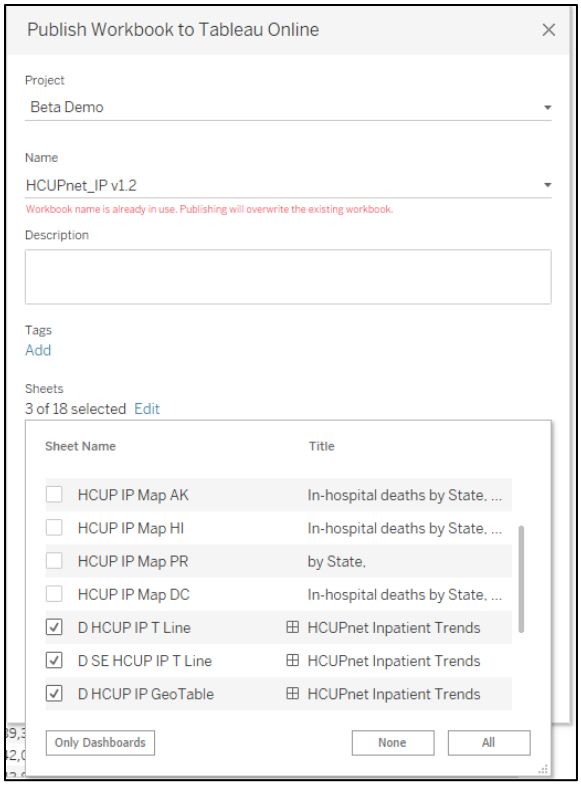


Figure 11. Tableau's Publish Feature

- **Embedding the Tableau in the .HTML File:** Once the files are published to the Tableau server, the backend user will need to take the embed code that's available and include it into the HTML site code. This is available in the "Share" button of the visualization:

Figure 12. Tableau's Share Button

After selecting "Share" the backend user will see the embed code. From there the backend user should select the embed code, make any modifications necessary to customize the view (e.g. place the toolbar at the top, not display certain buttons) and incorporate that embed code into the HTML code where needed.

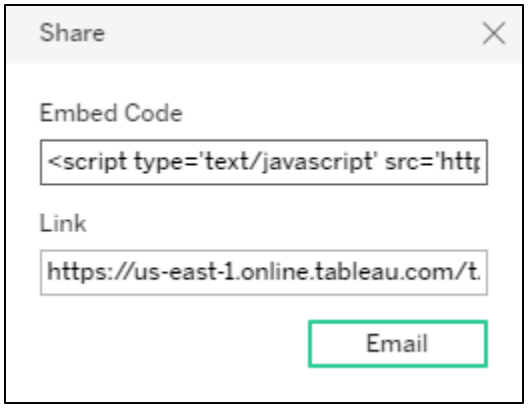


Figure 13. The Tableau Embed Code

2.6.2 Additional Tableau Feature Details

2.6.2.1 Downloading the Data Visualizations

During the demo session, users asked if there's a way to download just the graph part of a data visualization. Tableau provides the following information. This can be done when downloading as a PDF by selecting just the specific sheet in a dashboard.

From Tableau's website (https://onlinehelp.tableau.com/current/pro/desktop/en-us/export.htm):

> **PDF:** Downloads a PDF of the view, specific sheets from a dashboard, or specific sheets from a workbook. Under Include, select the part of the workbook you want to download. Select specific sheets or select all. Select Scaling to control the image's appearance on the PDF. Select Paper Size and Orientation.

2.6.2.2 Cascading Filters

A common issue within the prototyping phase occurred when a user chose a limited selection of options in a lower level filter. The selected options from the lower level filter would carry over when the user switched their selection on a higher-level filter, even if the lower level selections did not correspond with the new higher-level selections. While one solution would be to reconfigure the filters to merge higher and lower-level filter wherever possible, the project team successfully researched ways for alternate solutions that can predetermine lower level filter selections for every high-level filter selection.

A feature in Tableau called "Parameters" allows for making constant data more dynamic for the frontend user and functions in a similar way as a filter. With parameters, it's possible for backend users to configure filter choices based on higher level filter selections. The process involves creating a parameter with a certain value (e.g., "All"). By creating a conditional measure that utilizes the parameter, a backend user can code out every specific iteration for the filter. The new measure is then added to the rows section, and the parameter is added to the filter section. The parameter functions and looks like a regular Tableau quick filter but it has coding on the backend that will predetermine its outcomes.

To incorporate a parameter into the view, you must:

1. Define the parameter's values
2. Create a calculated field that includes all of the parameter's potential outcomes
3. Add the calculated field to the visualization
4. Add the parameter control as a filter for the visualization

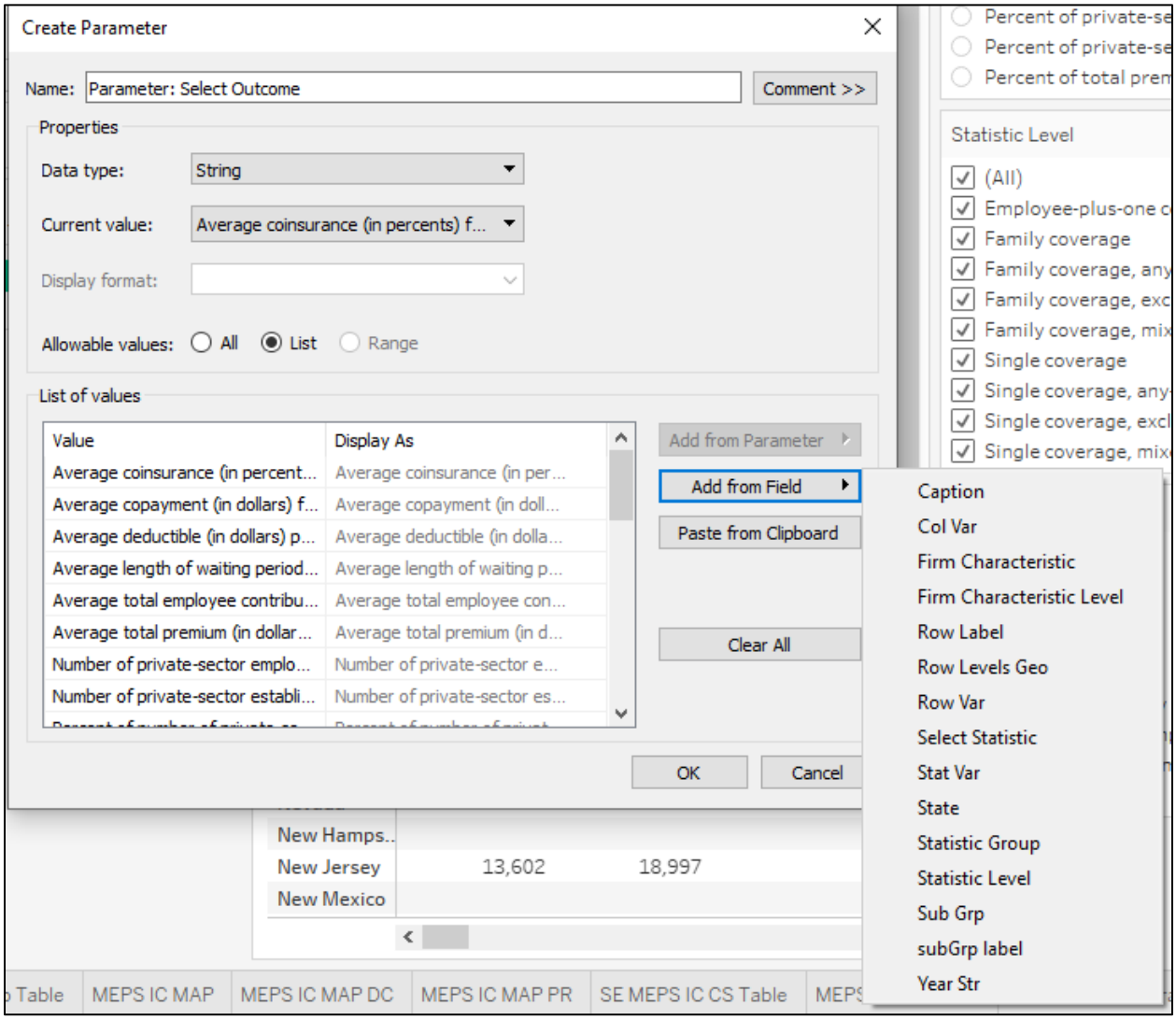


Figure 14. Creating the Parameter

The parameter can take on a number of different data types, such as string, number, float, Boolean, data, date & tie. Once a data type is selected, the user inputs the relevant values for the data type. Luckily, Tableau allows for "Add from Field" which is useful in importing all unique field entries. Once the parameter is created, it won't affect anything in the dashboard until it's

included in a calculated field. It should be noted that there is a “Value” column and a “Display As” column. This allows the backend user to configure the parameter to be a more easily workable value, such as 1 or “a” rather than typing out the entire filter name.

Figure 15. Creating the Calculated Measure

Including the parameter within a calculated field allows the backend user to define exactly what happens when a frontend user selects a choice in the parameter. Since this example uses a string function, the case function is employed to account for all potential outcomes of the filter that the frontend user will encounter. Here, the project team has spelled out all potential options in the “Select Outcome” field to display all “Statistic Level” options, which is a lower level filter.

If the parameter were a number or a float, equalities (e.g. <, <=, =, etc.) should be used to define the range that the frontend user could potentially select. To distinguish this calculated measure as connected to the parameter, the project team has added “Parameter” to the measure’s name. With outcomes defined for a parameter through the calculated measure, it’s now time to incorporate the parameter into the visualization.

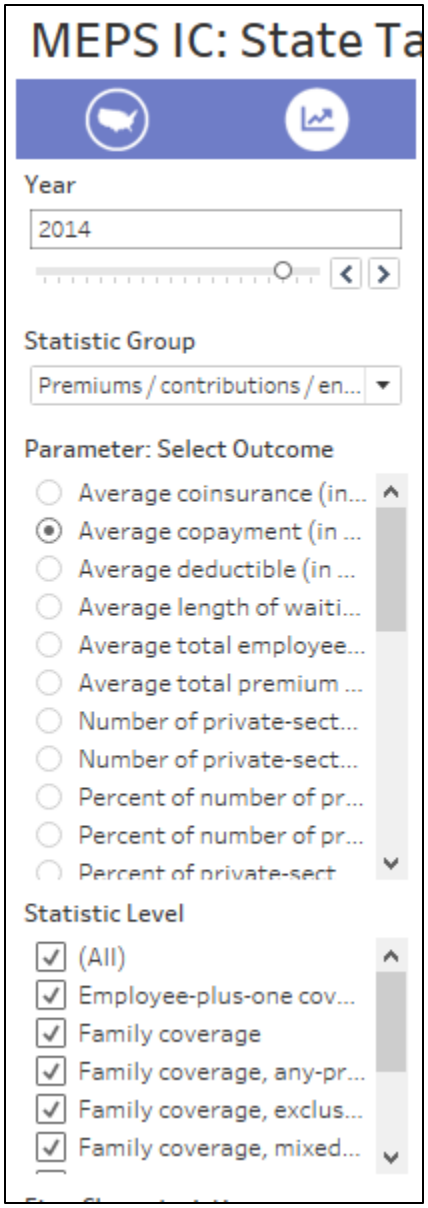


Figure 16. The Parameter Control in the Dashboard View

The parameter control can be thought of as the potential outcomes available to select for a parameter. Here, all potential options for the "Select Outcome" field are available for the frontend user to select. Moreover, no matter what the user selects in either "Select Outcome" or in "Statistic Level" all potential options will display in the visualization. This resolves the potential issues that a frontend user might encounter in backwards cascading filters.

One downside is that the parameter as a filter will need to be added to every visualization. With the Filter, the backend user can choose which visualization this filter applies, which saves a considerable amount of time. Without this connection, the backend user will have to configure an additional parameter to take the inputs of one parameter and apply it to the relevant visualizations. While this will add extra steps to all dashboards, it's a welcomed resolution to the

issue of backwards cascading filters and the project team recommends implementing it as part of a final visualization for the AHRQ Project Team.

2.6.2.3 Tagging of Data

The current tagging functionality available in Tableau is the ability for backend users to tag workbooks. Tagging workbooks allows for backend users to categorize workbooks making them easier to find for backend management purposes. In its current iteration, Tableau does not allow for frontend users to tag workbooks or data.

2.6.2.4 Dynamic Reference Lines

During the demo session, users noticed that the ICD 9 to 10 transition line was displayed for the trend line even when the user selected a point in time that did not involve the transition date. The project team researched whether it's possible to have this line only display when a user is looking at a date involving the transition date. It's possible to have this reference line dynamically appear with a parameter and a calculated measure that uploads the parameter to the trend line only if the year is selected by the user.

To denote the ICD 9 to ICD 10 transition, the project team utilized a reference line on the x-axis for the date of that transition (10/1/15). However, this reference line showed up on all date visualizations even if the user selected a date that did not include the ICD transition date. This issue can be resolved by creating a parameter for the ICD transition date, and then adding a conditional measure for the ICD line. If this resolution is implemented, when a frontend user changes the Year filter to a date that includes the 10/1/15 discontinuity, the ICD line will still appear but when the user changes the filter so that it does not include 10/1/15, the ICD line will disappear.

2.6.2.5 Customizing the Tableau Toolbar Buttons

During the demo session, users noted that the Tableau toolbar included many features that won't be needed by the end users. The project team can confirm that unwanted features can be configured to display on the tool bar via the embed code provided by Tableau (https://onlinehelp.tableau.com/current/pro/desktop/en-us/embed_list.htm). Specifically, these buttons can be configured off:

- Data Refresh & Pause
- View: Original
- Alerts
- Subscriptions
- Share

In the prototype, the project team deactivated the Subscribe and Share in the toolbars for MEPS IC's data visualizations.

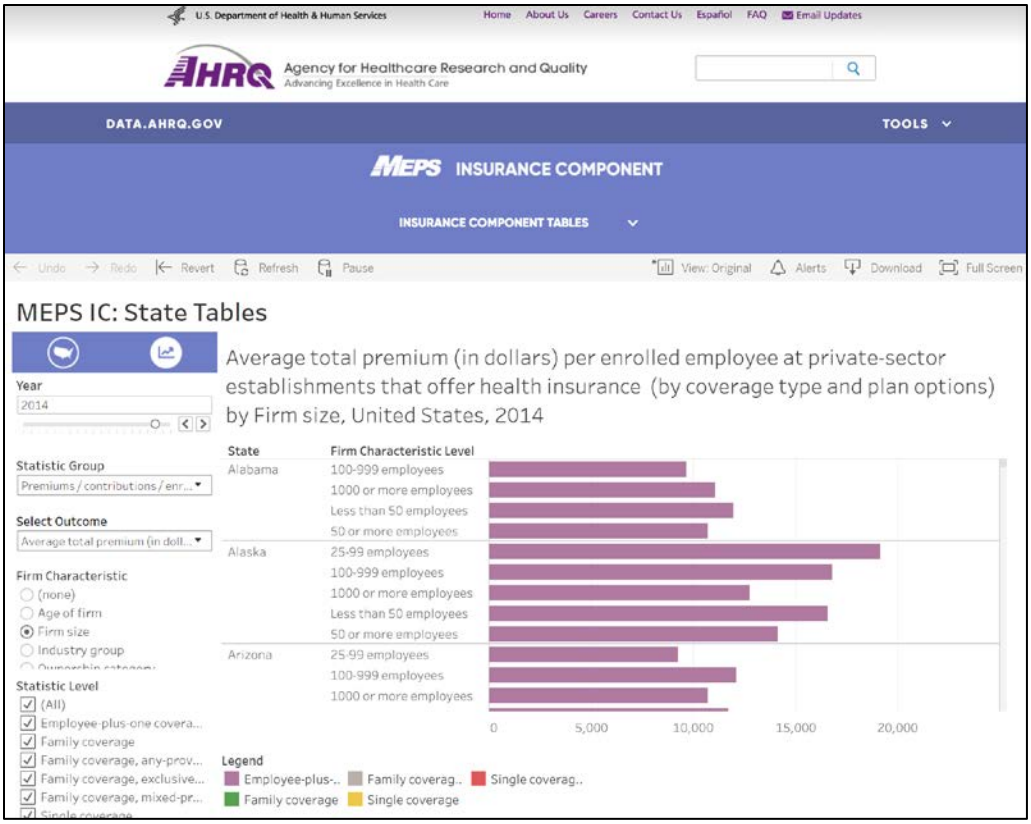


Figure 17. MEPS IC Visualization Without the Share and Subscribe Buttons

2.6.2.6 Z-Test Calculator

Z-Scores can be coded as a calculated field in Tableau based on the AHRQ data loaded into the dashboard. The simplest alternative for a z-test is to create a static Z-test application (see screenshot below), such as the one available in HCUP, in the HTML code of the product. This application performs a Z-test calculation using native client-side JavaScript. A more advanced version of the Z-test could use Tableau's REST APIs to request data from the Tableau visualization to populate a Z-test application. However, this feature was not included in the wireframes or prototype and would have to be considered within a redesign of the query pages.

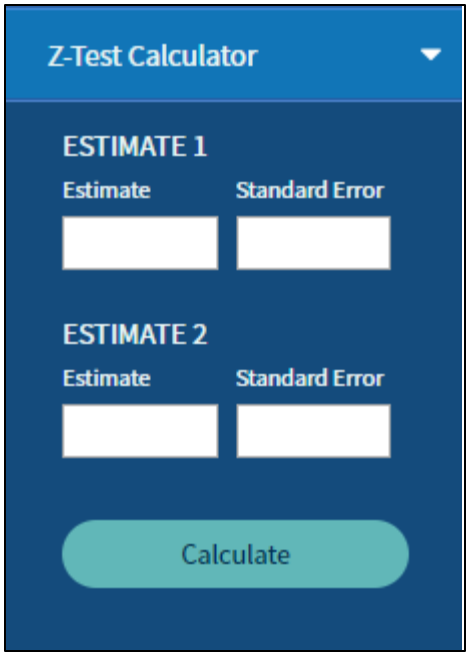


Figure 18. HCUP's Z-Test Calculator

For more details about how to calculate z-scores in Tableau, please refer to this: https://onlinehelp.tableau.com/current/pro/desktop/en-us/calculating_z_scores.htm.

2.6.2.7 Ask Data Feature

An up and coming feature in creating data dashboards is the ability to form a natural language data query in a search bar and receive relevant data insights. In Tableau, this feature is called "Ask Data", and it was released on February 14, 2019 in Tableau version 2019.1 with enhancements released in version 2019.2 released on May 21, 2019. It is available for any pricing level of Tableau.

Based on initial demonstrations of the product by the Atlas Team (see links to the demo below) in version 2019.1 and 2019.2, the service relies on the user providing phrasing for the data that they are looking for. The user has some freedom with phrasing. For example, if the user types in "Add Profit", with Profit being a measure, Tableau will suggest "Sum of Profit". Or if the user types "Profi" Tableau will suggest "Profit" measure). Ask Data has some functions built into it, such as summing measures, or finding measures by certain filters; however, it can't provide advanced calculations such as squaring data. Ask Data is able to build out visualizations relatively quickly, however it doesn't look possible to then modify or customize the data powering the visualization. While the Ask Data tool might save a few steps in creating a visualizations, it seems that if you know your data enough to use this feature, then you likely know how to structure the data through Tableau's easy-to-use drag and drops and build the visualization you really want. Nevertheless, this is a field where Tableau can make some considerable strides in the future where it could prove to be much more valuable than its current state.

If the AHRQ team were interested in implementing Ask Data, some careful consideration would need to be taken. Based on the demo, it looks as if Ask Data is available only in a Worksheet

view, which is the precursor to the Tableau Dashboard. This would mean that the frontend user would be able to see how the names of fields, making it necessary to format the field names in a user-friendly, intuitive way, such as the screenshot below:

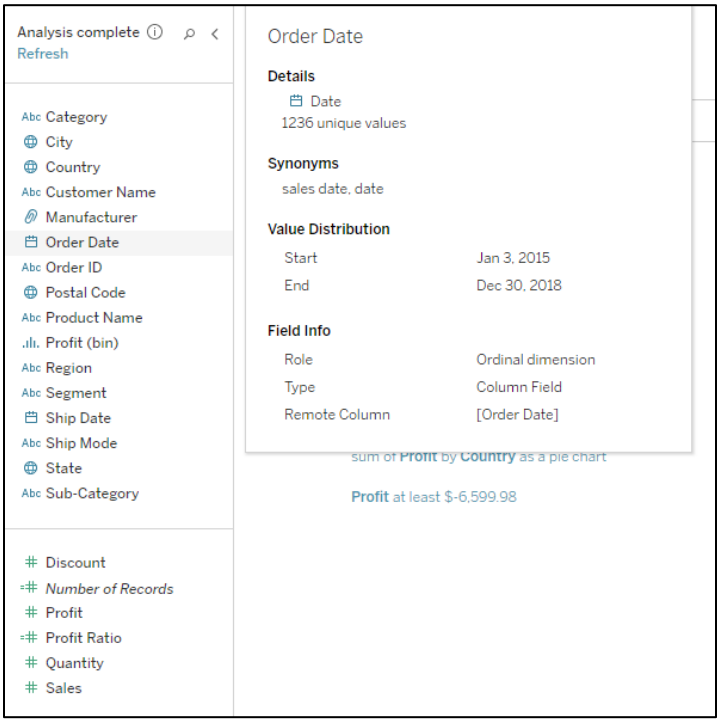


Figure 19. The Data Viewable Through the Ask Data Feature

Moreover, the AHRQ team would need to make sure that only relevant data is available for users to ask. Based on this, it seems that Ask Data might be most relevant to back end users for getting ideas about how to visualize certain data, as it saves a couple of steps in setting up a visualization. However, the Atlas Team does not recommend rolling this feature out for public consumption yet.

For more information regarding Ask Data, Tableau provides the following resources:

- https://www.tableau.com/about/blog/2018/11/ask-data-simplifying-analytics-natural-language-98655
- https://www.tableau.com/products/new-features/ask-data/demo (Note: Requires a Tableau login)
- https://www.tableau.com/products/new-features/ask-data#hero-reveal (Note: Requires a Tableau login)
- https://www.tableau.com/products/new-features#feature-102973 (Ask Data enhancements)

2.6.2.8 Accessibility of Tableau Visualizations (508 compliance)

Tableau supports several controls that can enable visualizations to adhere to the principles of accessibility (e.g., perceivable, operable, understandable). Tableau also provides guidelines

and best practices that creators can follow to conform to Web Content Accessibility guidelines (WCAG 2.0 AA) and comply with US Section 508 requirements. By creating visualizations that utilize and include WCAG-conformant elements, and publishing and embedding these in web pages that also conforms to WCAG 2.0 AA, the AHRQ data portal can ensure that it contains only accessible views.

Some basic examples of how to help ensure 508 compliance includes:

Providing good descriptive text in titles and captions – Rather than using the default title generated by Tableau, edit the title to ensure sufficient description of what information is being shown.

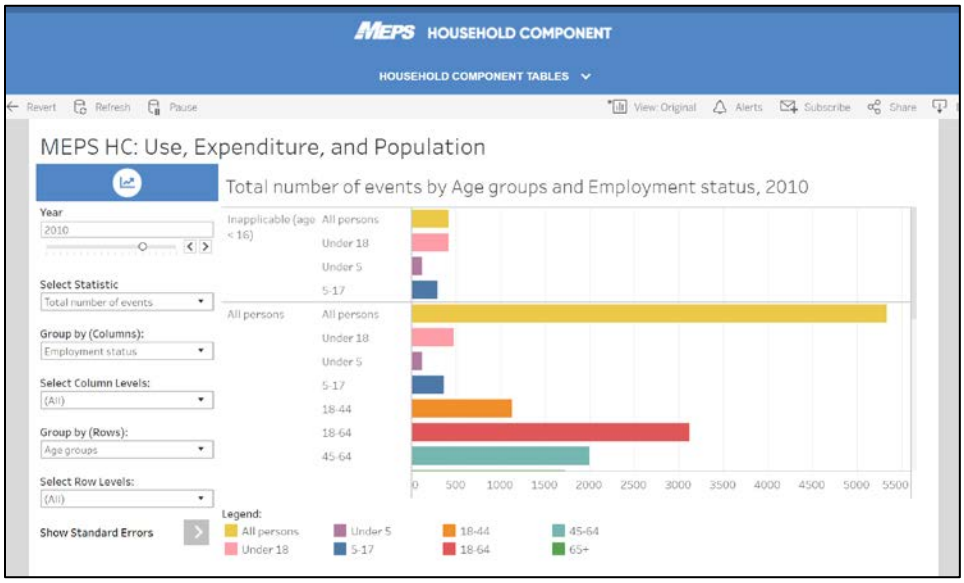


Figure 20. Having Descriptive Title

Providing additional text – Use text beyond titles and captions to provide better context to the readers of the different elements in the visualization.

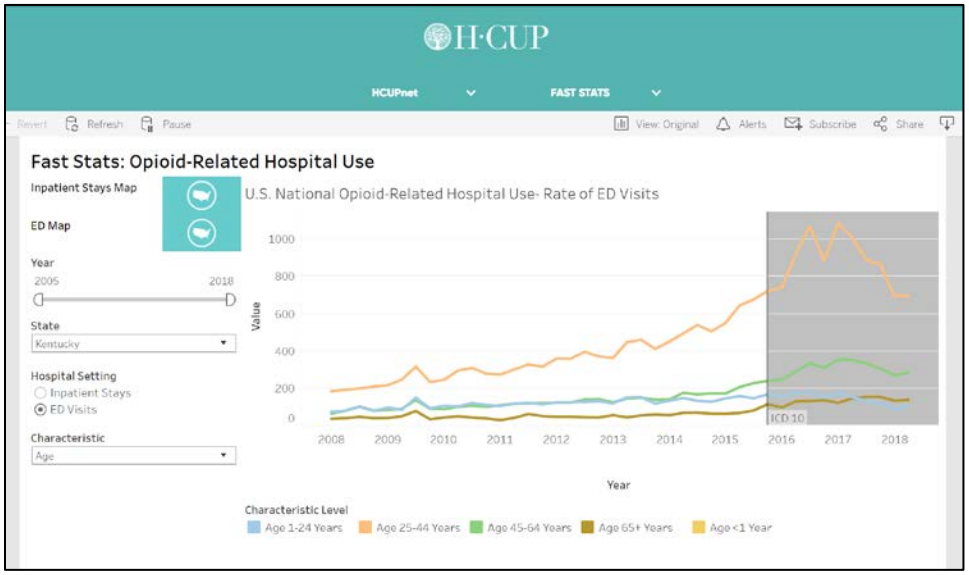


Figure 21. Legend and Descriptive Filter Labels to Provide Better Context

Providing visual cues beyond color – Use shapes, size, and labels to supplement color in distinguishing information.

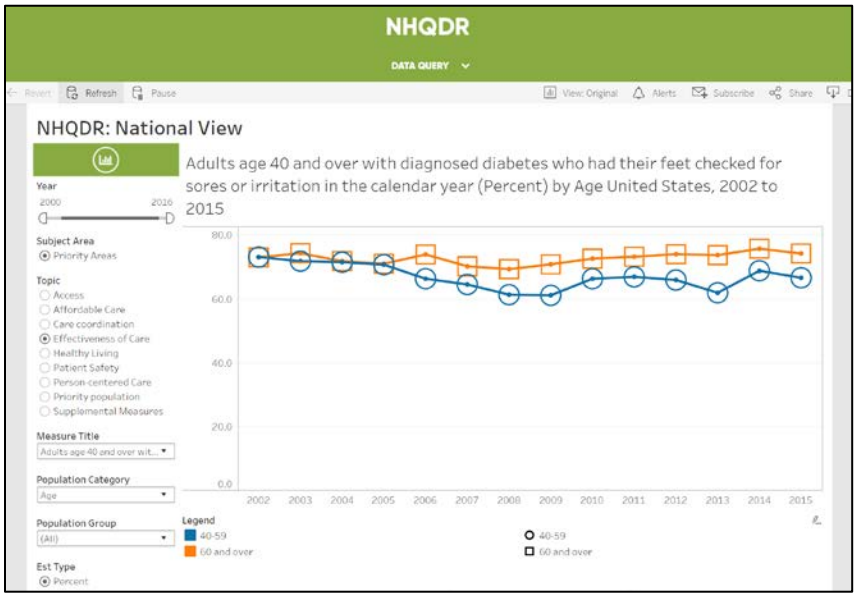


Figure 22. Using Color and Shapes for Easy Distinction

Other examples of things to consider include using Tableau's color-blind palette in creating visualizations and keeping the views as simple as possible while still providing helpful information.

For more information on how to support accessibility and 508 compliance, you may refer to these: https://onlinehelp.tableau.com/current/pro/desktop/en-us/accessibility_overview.htm and https://onlinehelp.tableau.com/current/pro/desktop/en-us/accessibility_dashboards.htm.

In addition, Tableau provides several guidelines to assist creators in building visualizations for accessibility:

- Best Practices for Designing Accessible Views (https://onlinehelp.tableau.com/current/pro/desktop/en-us/accessibility_best_practice.htm)
- Author Views for Accessibility (https://onlinehelp.tableau.com/current/pro/desktop/en-us/accessibility_create_view.htm)

2.6.2.9 Missing Values (Suppression)

Missing values in Tableau can affect the visualization in a number of ways. By default, Tableau tries to reduce the amount of null or missing data displaying in visualizations. This is noteworthy in trend lines, as Tableau will connect one year of data to the next year of data, giving the impression that a trend does not have a break in the data. However, Tableau will display the break in a trend line if the x-axis is categorized as a Date in the year format. After that, the creator must right click on the x-axis and select the option "Show Missing Values".

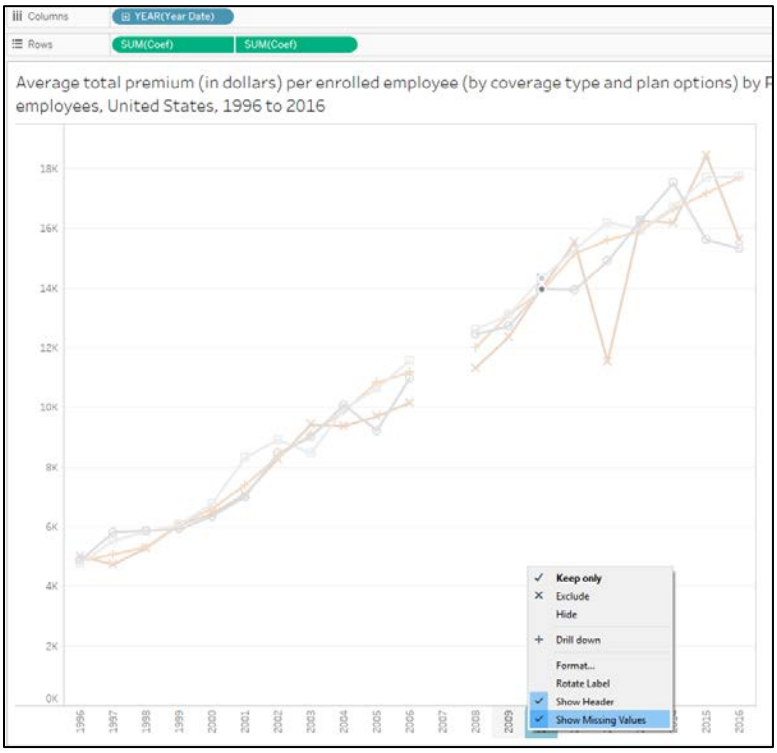


Figure 23. How to Display a Break in Missing Data for Tableau

For values that need to be suppressed, this can be done using parameters. After creating a parameter, the backend user should create a calculated field which contains the logic relevant to suppression (e.g. Sum([coef]) < 11 then “” else [coef]). Once this is included in the visualization’s values, the frontend user should have values suppressed that satisfy the suppression criteria.

2.6.2.10 Tableau Integration with R

Establishing the integration between R and Tableau Desktop is relatively simple, requiring only a few steps to start the connection. Once the integration is established, the relationship between the two programs becomes very powerful, allowing the statistical analysis of R to work with the visualization of Tableau.

Any R script, function, or package (any calculation) can be called from Tableau, so long as the script, function, or package is installed on the Rserve server used for the integration. New scripts, functions, and packages can easily be installed on the server.

One limitation to the integration is that visualizations in R cannot be imported directly into Tableau (i.e., graphs/charts/etc. that are created directly in R cannot be converted to a Tableau integration). On the other hand, the static image file or URL of the R visualization can be used in the Tableau dashboard.

2.6.2.11 Phone Layout

With Tableau Desktop, phone layouts are generated automatically by default. The phone layout automatically synchronizes any changes that the visualization creator implements to the default dashboard as long as the “Use automatic layout” option is in effect.

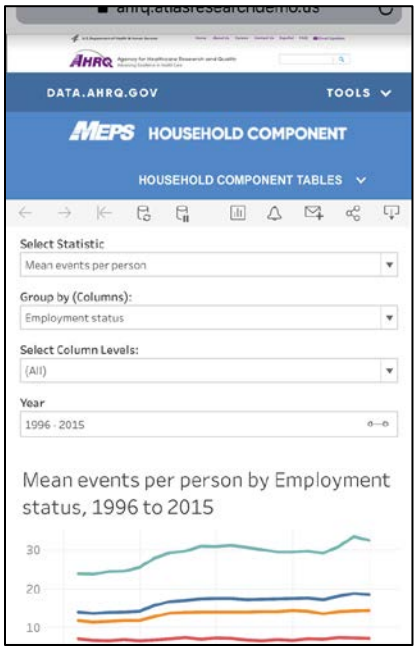


Figure 24. Sample Screenshot of Visualization in the Prototype in Mobile View

With the automatic generation, Tableau will arrange and stack the dashboard contents algorithmically in a phone-friendly manner.

The automatically generated phone layout is a good starting point for creators to further customize to optimize the viewing of the visualizations for small screens. To customize, the creator will need to edit the layout which will make the Phone layout become fully independent (changes made to the default dashboard will need to be manually added to the Phone layout, if applicable).

For more details about creating and optimizing Phone layouts, you may refer to this: https://onlinehelp.tableau.com/current/pro/desktop/en-us/dashboards_dsd_create.htm.

2.6.2.12 Side-by-Side Comparison

HCUP Fast Stats currently has a side-by-side compare option. This allows users to compare two states' statistics in the same screen.

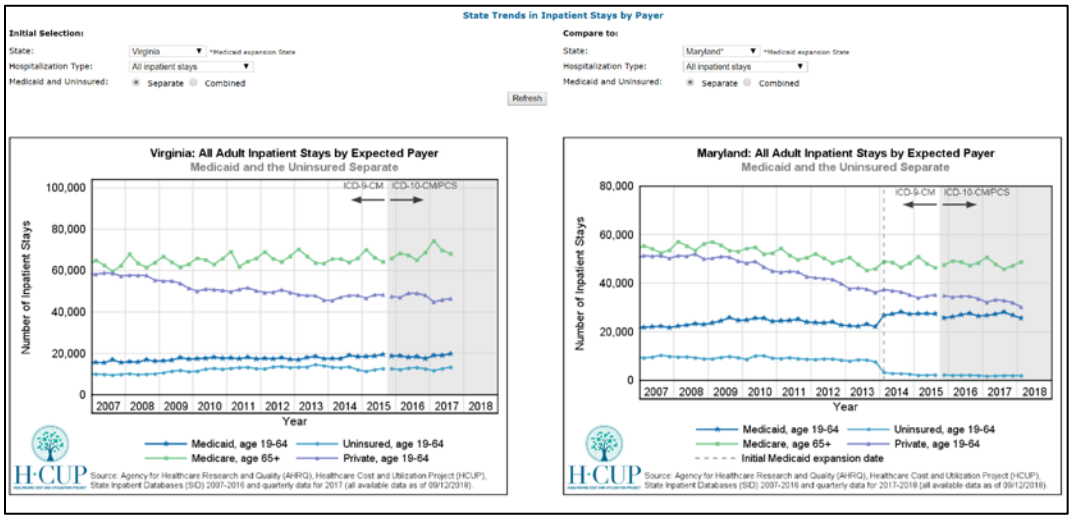


Figure 25. HCUP Fast Stats' Side-by-Side Comparison

This side-by-side compare option is not included in the prototype. With the prototype, a workaround to the side-by-side comparison can be made by opening two windows in the browser and arrange them so that they appear separately and side-by-side. Then, open the visualization on both windows while selecting different states in each to compare.

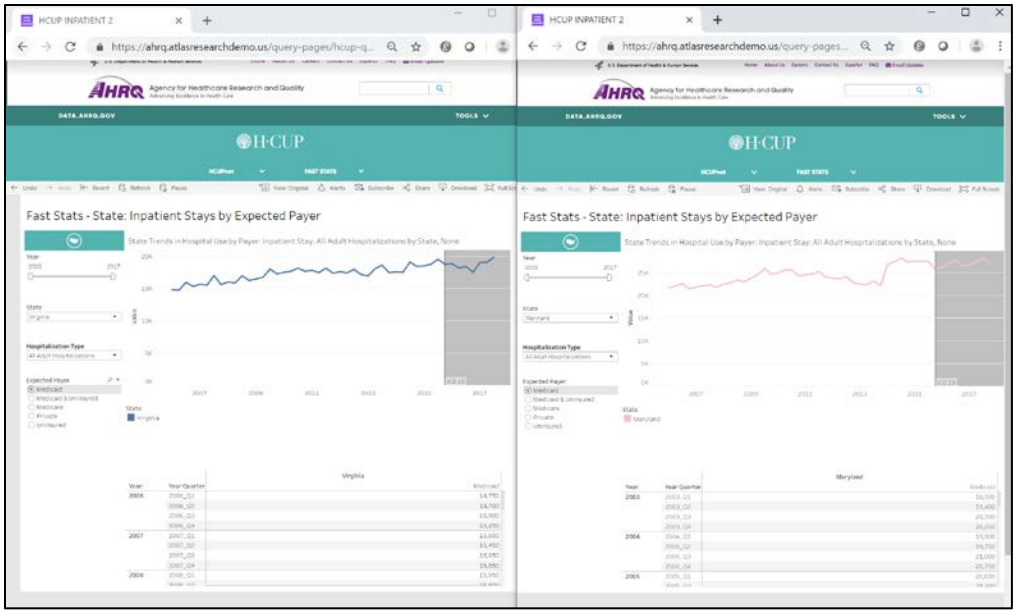
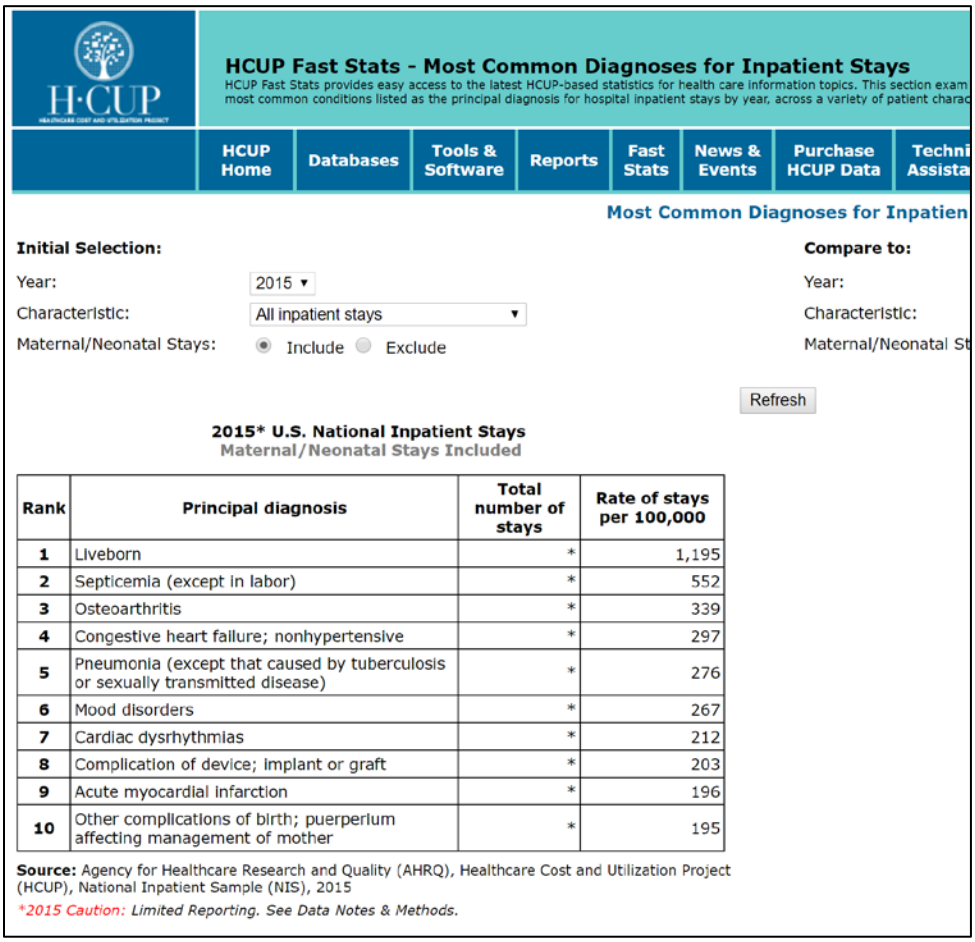


Figure 26. Two Browser Windows to Show Side-by-Side Comparison of Prototype Visualizations

If it is determined during the actual development of the harmonized AHRQ data portal that a side-by-side compare option on one screen is required, one solution would be to have two worksheets in a dashboard with both worksheets using the same data but using their own set of filters. The two worksheets and each of their set of filters will be identical except one set of filters will apply to one visualization while the other set of filters will only apply to the second visualization.

2.6.2.13 Adding Footnotes

Adding footnotes with tableau visualizations may be addressed in a variety of ways. Providing footnotes for specific values in a table is required and important for some of AHRQ's data tools. For example, the current HCUP Fast Stats site may not report statistics that are not based on full year data, which is explained in the notes that the site displays.



HCUP Fast Stats - Most Common Diagnoses for Inpatient Stays

HCUP Fast Stats provides easy access to the latest HCUP-based statistics for health care information topics. This section exam
most common conditions listed as the principal diagnosis for hospital inpatient stays by year, across a variety of patient charac

HCUP Home | Databases | Tools & Software | Reports | Fast Stats | News & Events | Purchase HCUP Data | Techni Assista

Most Common Diagnoses for Inpatien

Initial Selection:

Year: 2015

Characteristic: All inpatient stays

Maternal/Neonatal Stays: Include Exclude

Compare to:

Year:

Characteristic:

Maternal/Neonatal St

Refresh

2015* U.S. National Inpatient Stays
Maternal/Neonatal Stays Included

Rank	Principal diagnosis	Total number of stays	Rate of stays per 100,000
1	Liveborn	*	1,195
2	Septicemia (except in labor)	*	552
3	Osteoarthritis	*	339
4	Congestive heart failure; nonhypertensive	*	297
5	Pneumonia (except that caused by tuberculosis or sexually transmitted disease)	*	276
6	Mood disorders	*	267
7	Cardiac dysrhythmias	*	212
8	Complication of device; implant or graft	*	203
9	Acute myocardial infarction	*	196
10	Other complications of birth; puerperium affecting management of mother	*	195

Source: Agency for Healthcare Research and Quality (AHRQ), Healthcare Cost and Utilization Project (HCUP), National Inpatient Sample (NIS), 2015

**2015 Caution: Limited Reporting. See Data Notes & Methods.*

Figure 27. Fast Stats' Note on Limited Reporting for 2015

A possible way to implement this in the harmonized data portal is to show the footnotes in the caption while adding another column in the table with calculated field that has logical criteria that, when met, will point to the caption for details. For example, the criteria could be when an asterisk or a value for suppression of data is found, the calculated field will state and show “See Notes”.

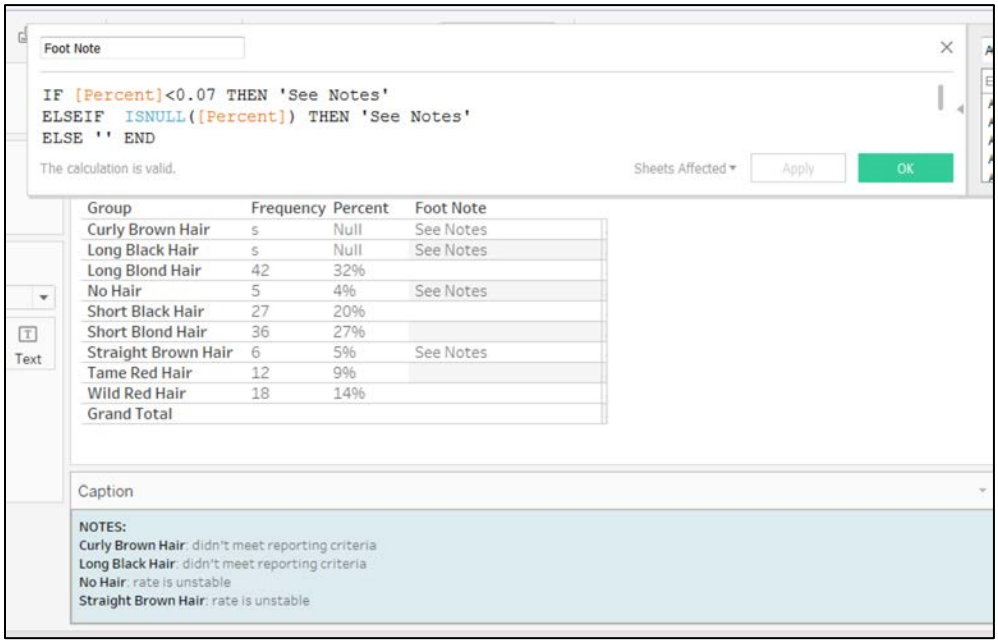


Figure 28. Example of Pointing to the Caption for Footnotes

(source of image: https://community.tableau.com/thread/219147)

2.6.2.14 Notes and Disclaimers

Tableau offers multiple options to displaying notes and disclaimers. Firstly, on the worksheet level, a caption can be added to a visualization. At the worksheet level, the backend user can customize the text to either be static or dynamic (i.e. based on filters selected by the frontend user.

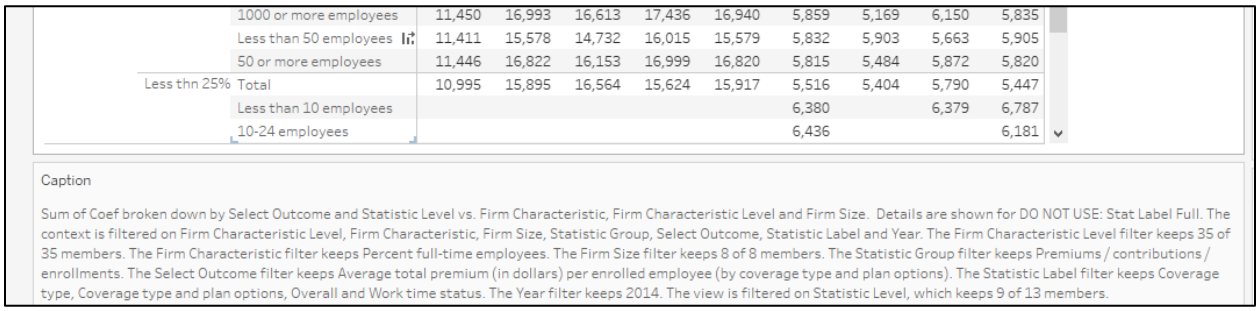


Figure 29. Tableau’s Caption Feature in a Worksheet

This caption can be on display in the dashboard and in fact the text can be modified or formatted by the backend user. However, the backend user cannot directly change the size of the caption box, as the size of the caption box is dependent on the size of the worksheet on the dashboard.

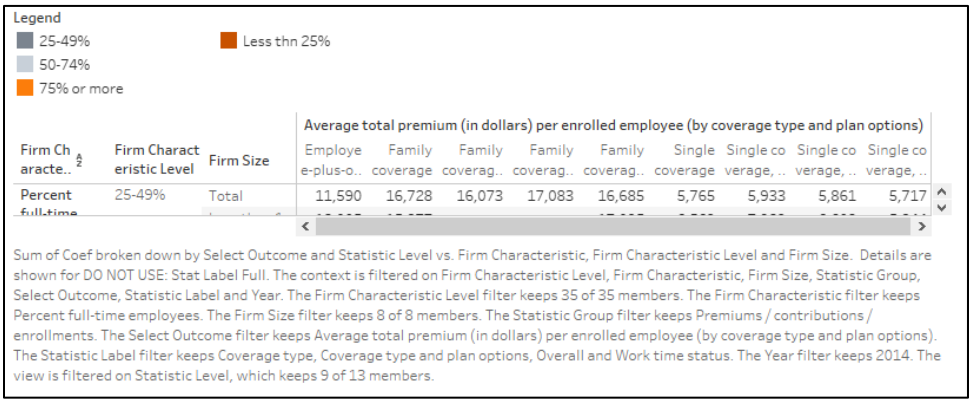


Figure 30. Tableau's Caption Feature in a Dashboard

See this page for more information on captions in Tableau:
https://community.tableau.com/thread/219147

Should a backend user want to more directly configure the size and shape of the caption, the backend user can create a calculated field that contains the text of the caption. This text can be static or dynamic, as the calculated field can be a formula.

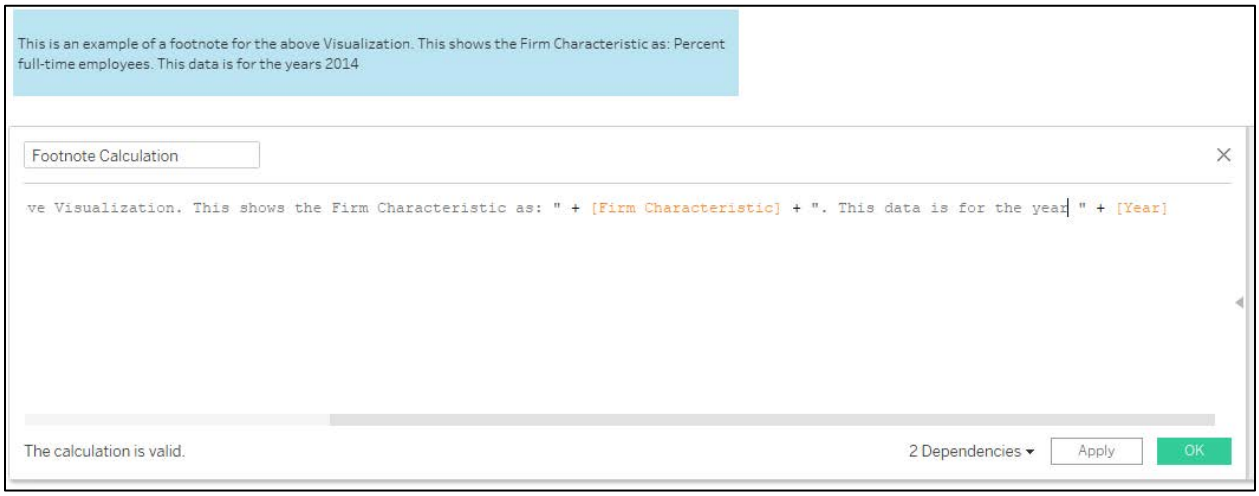


Figure 31. Formula for a Dynamic Footnote

Since this text is a worksheet, it can now be incorporated into the dashboard as either a tiled or floating visualization.

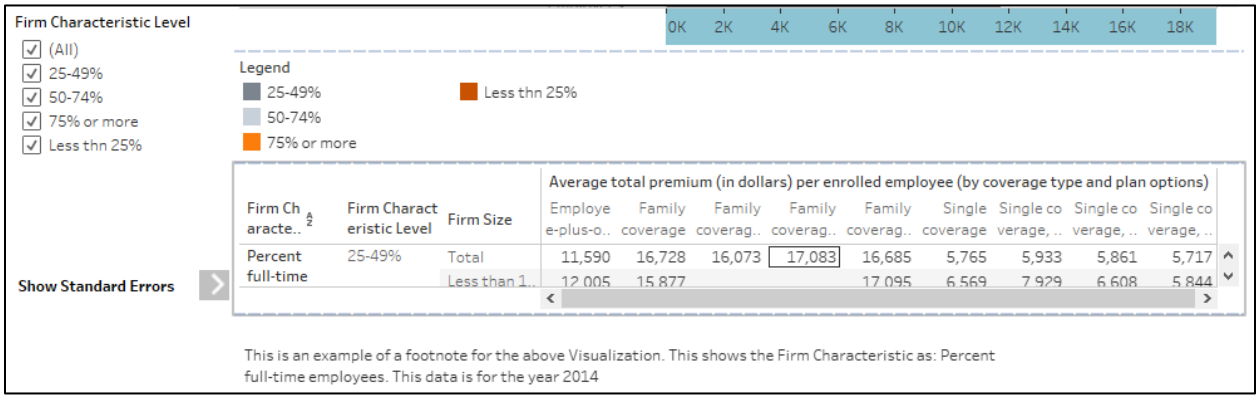


Figure 32. The Dynamic Footnote in a Dashboard

Alternatively, if AHRQ wants to direct users to a collection of notes and disclaimers pertaining to all visualizations, Tableau offers the ability to link to a webpage and display the linked site.

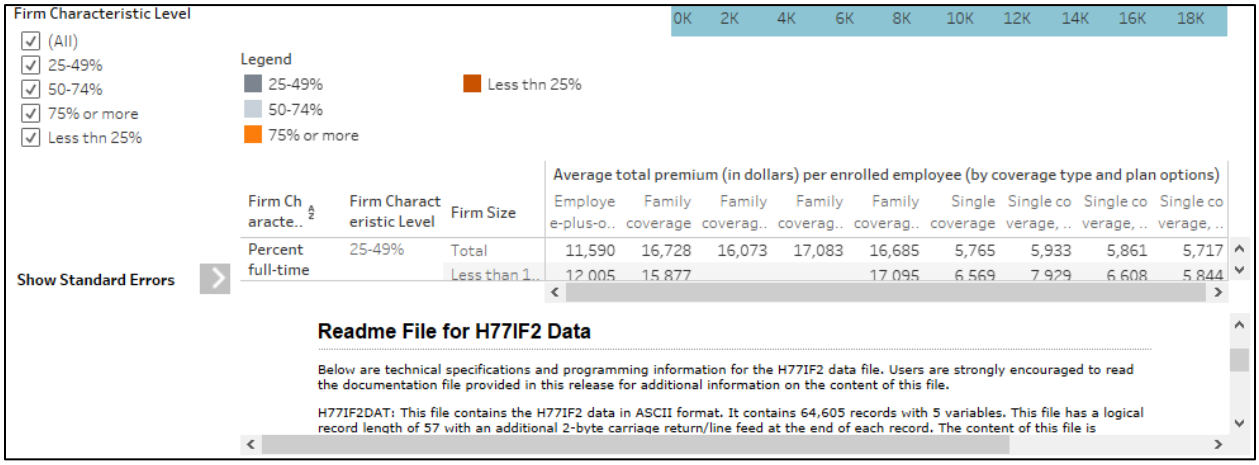


Figure 33. Tableau's Webpage Object linking to a Readme URL

More information on displaying web pages in a dashboard can be found here: https://onlinehelp.tableau.com/current/pro/desktop/en-gb/actions_dashboards.htm.

2.6.2.15 Dynamic Titles and Captions

Titles and captions are created and edited to provide as much descriptive information as possible. Providing the best description or summary of the important information being presented helps the users better understand the visualization.

By default, the worksheet's name is used as the title of the visualization. The title can be edited by selecting the Worksheet > Show Title option. By replacing the default sheet's name in the title with a more descriptive one and including the relevant filters as part of it, users will already be provided with helpful information initially about the visualization just from the title.

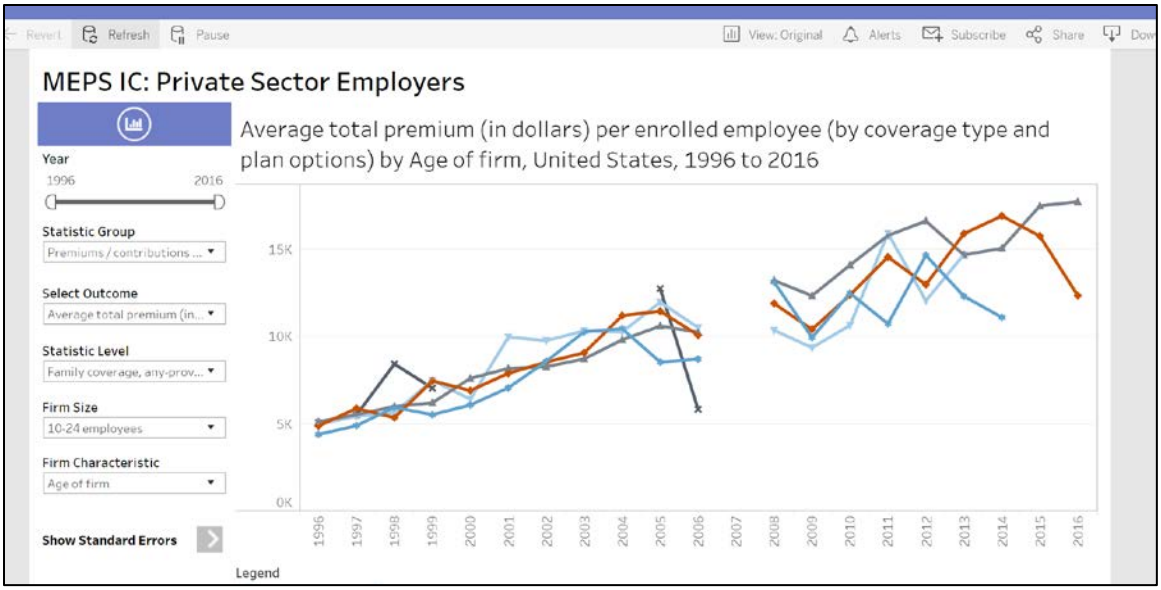


Figure 34. Using Filters in the Title

For more details about how to create dynamic titles using filters, please refer to this: https://kb.tableau.com/articles/howto/creating-dynamic-titles-based-on-filters.

As for captions, Tableau generates them by default and automatically describes the filters in the caption area. However, creators can edit the captions to provide better descriptions and highlight important aspects of the visualizations.

Figure 35. Sample of a Visualization Showing the Caption

(source: https://onlinehelp.tableau.com/current/pro/desktop/en-gb/accessibility_create_view.htm)

For more details about captions (editing, hiding/showing, etc.), you can refer to this: https://onlinehelp.tableau.com/current/pro/desktop/en-gb/viewparts_captions.htm.

Adding as much information as is possible to the title and caption that describe or summarize some of the important points of the visualization helps readers understand the visualization better.

2.6.2.16 Tableau's Stories Feature

All of the visualizations embedded in the prototype were created using a feature called "Dashboards" from Tableau, which allows backend users to compile multiple visualizations into one area. The project team utilized a feature called "Buttons" which allows frontend users to navigate to different dashboards. The project team utilized a feature within filters that allows selections from filters to carry over to relevant filters in selected dashboards, thereby improving the frontend user experience and giving the impression that all visualizations are connected.

Tableau does offer another method for displaying visualizations called "Stories". However, there are losses in functionality when using stories versus dashboards.

Firstly, the navigation buttons for stories offer little room for customization. The stories' buttons can contain text that's configurable, but backend users cannot replace the buttons with a customized image or change the color to match the theme of the query page. With dashboards, buttons can be configured with different images and colors.

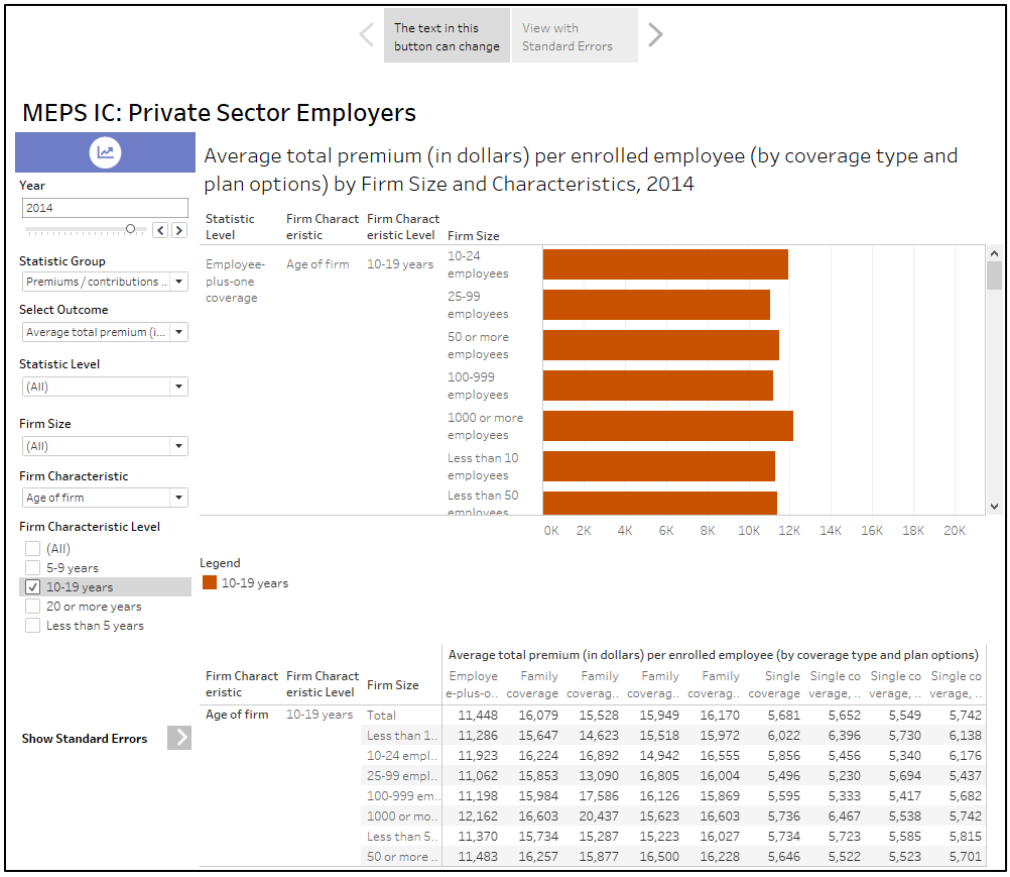


Figure 36. MEPS IC Private Sector Cross Sectional Visualization in Story Format

Secondly, a feature that the project team went to great lengths to enhance the user experience within the prototype is connecting relevant filters to visualizations within a query page. This saves the frontend user time from having to reconfigure filter selections from scratch on every visualization. This functionality is not available within stories. For example, if a user wanted to view the standard errors, they would have to navigate to the standard errors view and start their filter selections from the beginning.

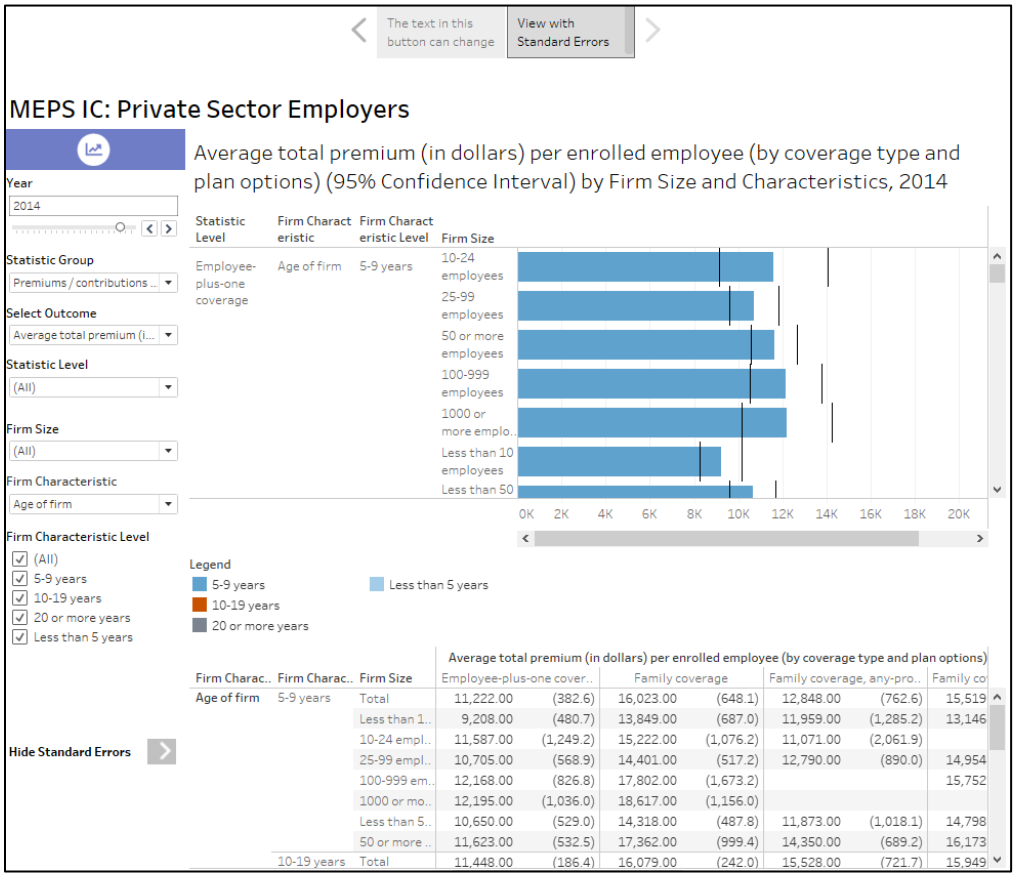


Figure 37. MEPS IC Private Sector Cross Sectional Visualization with Standard Errors in Story Format (Note how the Firm Characteristic Level selections do not carry over from Figure 36)

There are some benefits to the story feature. For each dashboard, the project team added a title to the visualization to note which query page the user is on. This can be localized to one text box on the story feature. Additionally, while only the text in the buttons for the story are configurable, the stories' buttons offer a more transparent way to help users navigate through visualizations, as all potential visualization options are displayed to the frontend user. The Atlas Team recommends attempting to recreate these benefits within the dashboards, as the resulting loss in functionality when using the stories' features will diminish the frontend user's experience.

2.6.2.17 Connecting to Secure Data Sources

Tableau supports connecting to data that is stored in a wide variety of places, from a spreadsheet on a computer to a complex relational database on a server (e.g., a SQL server

database). Upon further consultation with the project team's point-of-contact at Tableau, he confirmed that Tableau can connect to various data sources leveraging a live connection via the security protocols set up on the data source. He further referred us to Tableau's Security landing page for more information (https://www.tableau.com/security).

When building visualizations and connecting Tableau to the desired data, you will be asked to supply different information for each data connection that will be needed. For data connections to a database on a server, this will include includes supplying a server name and valid sign-in credentials.

For more information about connecting to data sources, you may refer to this: https://onlinehelp.tableau.com/current/pro/desktop/en-gb/basicconnectoverview.htm.

Related information, especially relevant when publishing data and setting credentials may be found here: https://onlinehelp.tableau.com/current/pro/desktop/en-gb/publish_overview.htm and https://onlinehelp.tableau.com/current/pro/desktop/en-gb/publishing_sharing_authentication.htm.

2.6.2.18 Adding Insets in Maps

In creating a map visualization in Tableau that spans the whole United States of America, you can present Alaska, Hawaii, and other states as needed, as insets so that they can be displayed closer to the map of the continental United States while removing non-relevant maps (e.g., Canada, if showing only U.S. data) from the focus of the visualization.

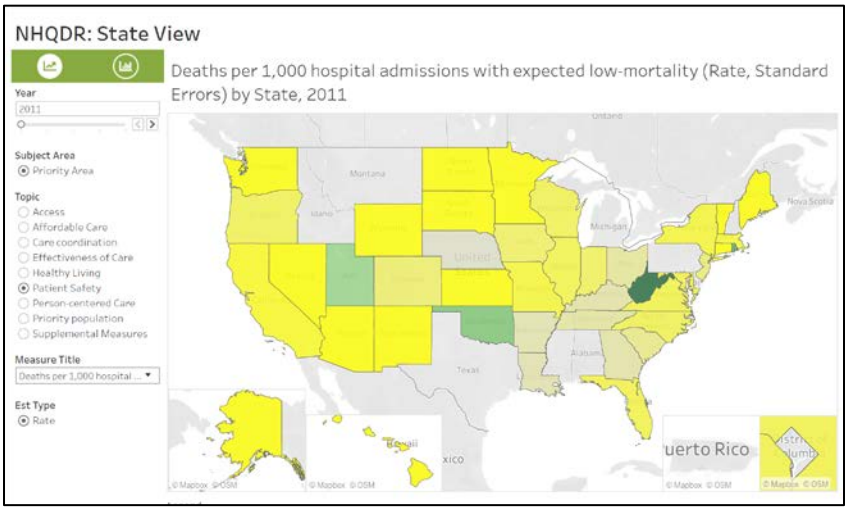


Figure 38. Prototype Showing Insets in the Map View

To do this, separate maps will need to be created for the continental United States, Alaska, Hawaii, and the other states that will be represented as insets. Each map will need to be filtered to only include the states that will be displayed in that map (e.g., for the continental United States' map, all states will be selected except for Alaska and Hawaii; for the map of Hawaii, the state filter will only have Hawaii selected). After this, a dashboard with a floating layout will need to be created to be able to show all maps together.

For more details, you may refer to Tableau's Knowledge Base entry: https://kb.tableau.com/articles/howto/showing-hawaii-and-alaska-near-continental-us.

2.6.2.19 Using Charts as Filters

Using Tableau's Filter Action feature, information may be sent between worksheets. A selected mark or data point from one worksheet (e.g., a map or a line chart) can be sent as a filter to another worksheet (e.g., bar graph) showing related information. The filter action can be configured to run either on hover (i.e., resting the cursor over a mark), select (i.e., clicking on a mark), or menu (i.e., right-clicking first on a mark, then clicking or selecting to proceed).

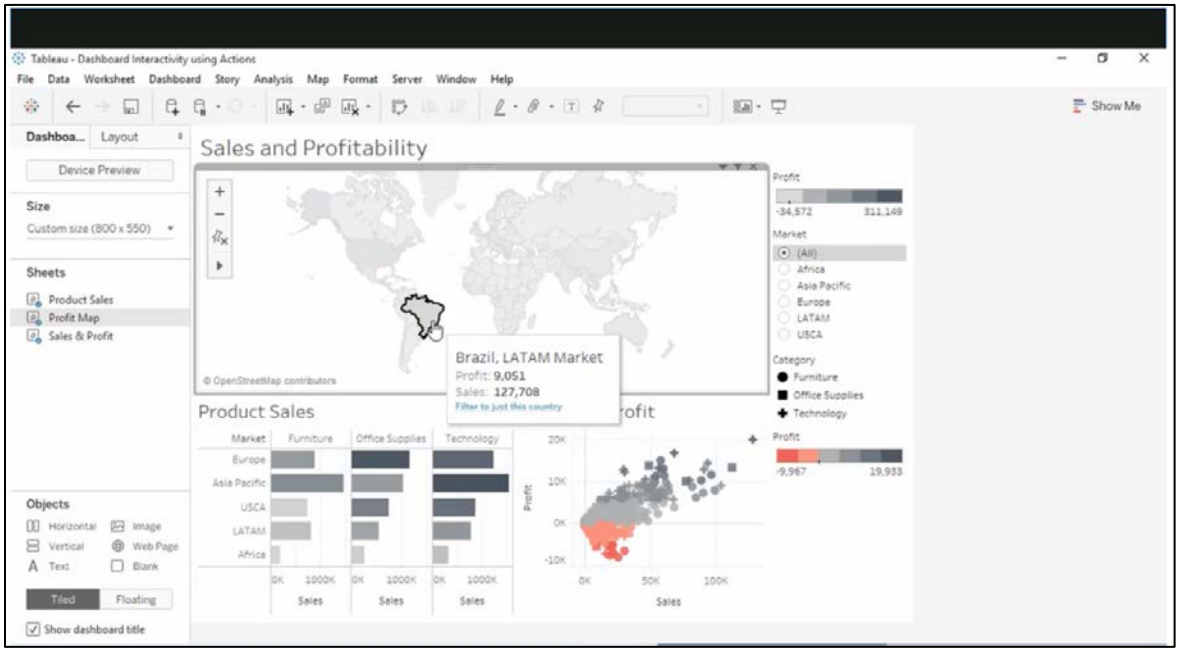


Figure 39. Sample: View Before a Mark is Selected

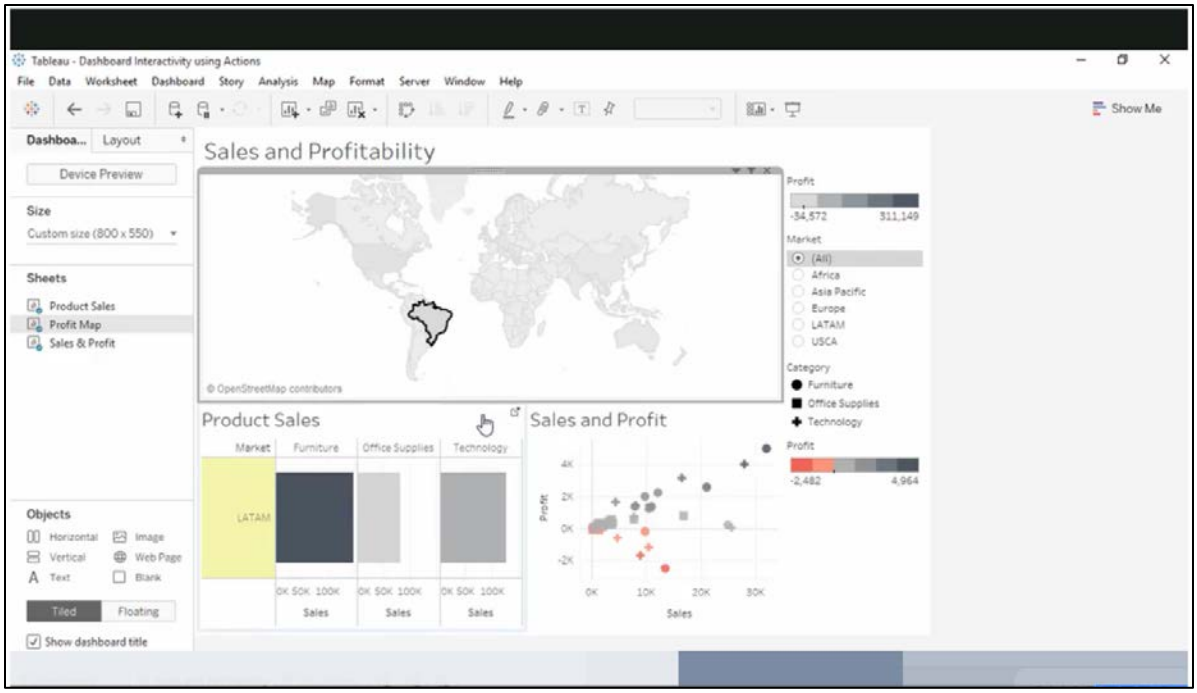


Figure 40. Sample: Information Changes After Brazil LATAM is Selected

For more details (and the source of the two sample figures above), you may refer to this web page on Tableau's Online Help: https://onlinehelp.tableau.com/current/pro/desktop/en-gb/actions_filter.htm

2.6.2.20 Amount of Data and Performance

According to Tableau Knowledge Base (https://kb.tableau.com/articles/howto/maximum-limit-for-rows-or-columns-of-data), Tableau does not have any enforced row or column limits for the amount of data that can be imported. Hardware resources, such as RAM, CPU, and disk space, the capabilities and limitations of the underlying data source, perceived performance and usability are factors that determine the physical and theoretical data size limits.

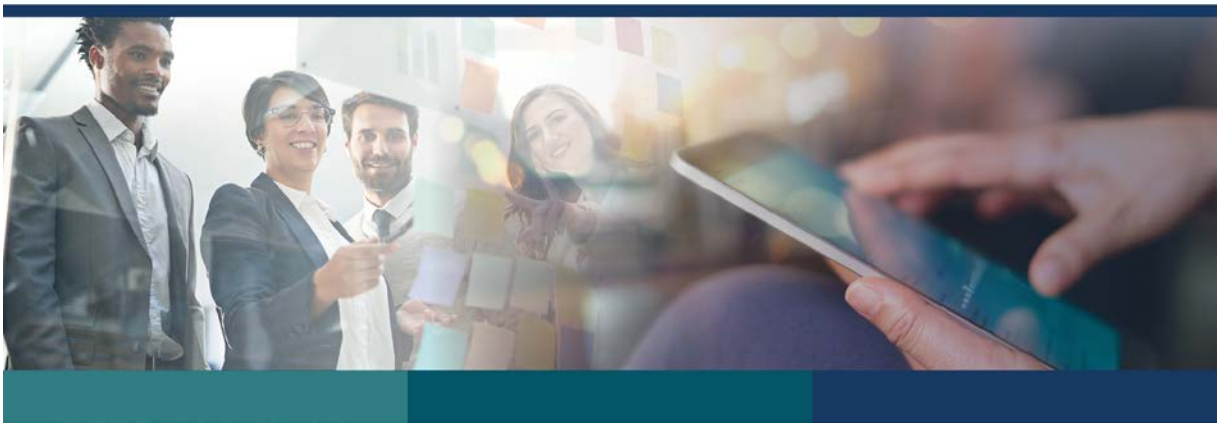
Upon further consultation with the team's POC at Tableau, he provided the additional information that Tableau is a read-only software and allows the end user the flexibility to either connect live to the data source or at a scheduled extract level to improve performance. There is no limit on the amount of data that Tableau can consume. He added that Tableau works with 86,000 organizations including the small handful of the largest companies in the world who work with massive amounts of data.

The following link provides a helpful resource with various discussions on how to optimize performance: https://onlinehelp.tableau.com/current/pro/desktop/en-us/performance_tips.htm. Additional Tableau tips on optimizing data at the database level may be found here: https://onlinehelp.tableau.com/current/pro/desktop/en-us/perf_data.htm.

APPENDIX A: ACRONYMS AND DEFINITIONS

Table 12. List of Acronyms and Definitions

Acronym	Description
AHRQ	Agency for Healthcare Research and Quality
BPA	Blanket Purchase Agreement
CAHPS	Consumer Assessment of Healthcare Providers and Systems
COTS	Commercial off-the-shelf
ELA	Enterprise License Agreement
GSA	Government Services Administration
HC	Household Component
HCUP	Health Cost and Utilization Project
HHS	Department of Health & Human Services
IC	Insurance Component
IE	Internet Explorer
MEPS	Medical Expenditure Panel Survey
NHQDRnet	National Healthcare Quality and Disparities Reports Data Query Tool
POC	Point-of-Contact
POP	Period of Performance
TOO	Task Order Officer
UI	User Interface
UX	User Experience



RFTO #10: Developing Strategies for the Harmonization of AHRQ Data Tools (Data Tools)

Technical Specifications for AHRQ Data Tools

December 28, 2018

CONTENTS

REVISION HISTORY

Date	Version	Description	Author(s)
November 26, 2018	Draft	Initial draft, will be finalized after client comments have been incorporated	Atlas Team
December 3, 2018	AHRQ Review	AHRQ review, corrections, and feedback	AHRQ
December 14, 2018	Final	AHRQ feedback incorporated into the document	Atlas Team
December 21, 2018	AHRQ Review of the Final Report	AHRQ review feedback and updates to the December 14 version of the report	AHRQ
December 28, 2018	Updated Final	AHRQ feedback incorporated into the document	Atlas Team

NOTE: The following document was produced as part of AHRQ's Data Tools Harmonization Project Assessment Phase. The original document contained a link to an embedded spreadsheet. That link was removed and replaced with the following highlighted text:

(removed)

1. EXECUTIVE SUMMARY

The Agency for Healthcare Research and Quality (AHRQ) provides several siloed data tools and datasets that currently lack cohesion of both user experience and branding for AHRQ.

To better understand both the value gained and the potential challenges facing the harmonization of these data tools, the Atlas Team is conducting an environmental scan intended to understand the unique and varied needs of the tools' users and stakeholders. A facet of this environmental scan includes gathering and outlining current technical specifications for each AHRQ data tool. The technical specifications were sourced from a survey that asked for details on the hardware and software components of the data tools, along with the necessary processes to extract, transform and load data into the tools.

The purpose of this report is to gain a general understanding of the current technical environment and needs of AHRQ's various data tools and determine the common specifications needed to develop a possible common look and feel across the various data tools used by the agency. The information gathered in this report will help determine the technical requirements that will be included (along with the functional requirements) in the Environmental Scan Report. The technical and functional requirements in the Environmental Scan Report will be used in assessing products that may deliver the required features of a harmonized data tools interface and final product recommendation.

The specific AHRQ data tools included in the scope of this scan are:

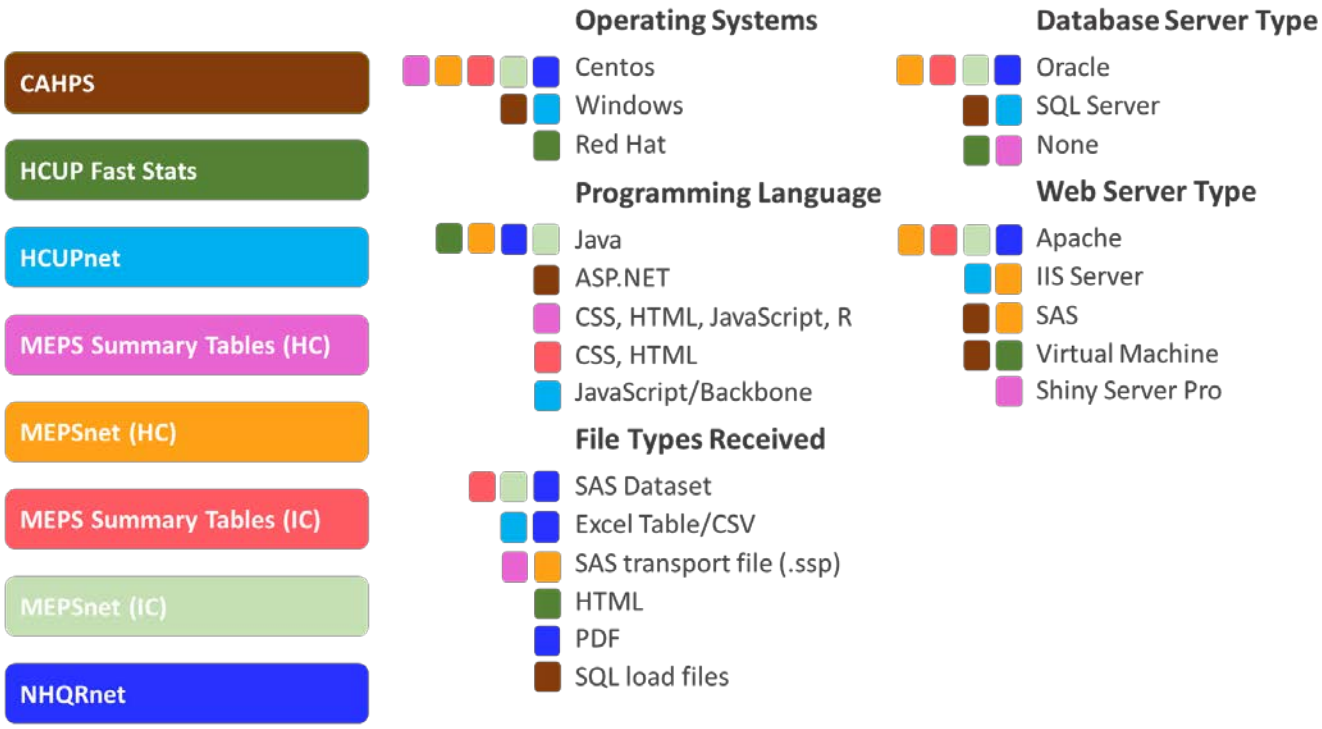
- The Consumer Assessment of Healthcare Providers and Systems (CAHPS) Aggregated databases
- The Health Cost and Utilization Project's Fast Stats
- HCUPnet
- Medical Expenditure Panel Survey (MEPS) Summary Tables: Household Component (HC)
- MEPS Summary Tables: Insurance Component (IC)
- MEPSnet (HC)
- MEPSnet (IC)
- National Healthcare Quality and Disparities Reports (NHQRnet) Data Query Tool

1.1 AHRQ Data Tools Technical Specifications At a Glance

The specifications summarized below are the core technical requirements to be considered when assessing the products that will deliver the required features of the data tools and harmonized interfaces.

In Figure 1 (Summary of Technical Specifications), the technical specifications for each tool are visualized and color-coded. For example, CAHPS (colored brown) has Windows for an Operating System, utilizes ASP.NET as a Programming Language, receives SQL load files for updates, has a SQL server for its database server, and uses a Virtual Machine for its Web Server.

Figure 1. Summary of Technical Specifications



2. TECHNICAL APPROACH

The approach employed to review the technical needs and capability of these data tools consists of collecting, reviewing and matching/comparing all government-provided technical specifications. The scope of the government-provided technical information has been limited to general, high-level technical specifications which will eventually be used to conduct, compare, and recommend possible commercial off-the-shelf (COTS) products that are readily available in the industry.

The government-provided technical specifications collected are centered around the internal/operational components that makes government statistics accessible to the public. The components include databases, servers, and other operational components. This report provides information on component type, versions, sizes and the other general technical information as provided by the government.

The inputs for the technical specifications were sourced from primary research. The Atlas team used a technical specifications template for each data tool as guidance when interviewing the data tool's point of contact (POC). Following the interview, the Atlas team sent a draft version of the template with completed answers when applicable for the POC to validate and complete.

The Atlas team has received technical specifications from the POCs for the following products:

- CAHPS
- HCUPnet
- HCUP Fast Stats
- MEPS Summary Tables (HC)
- MEPS Summary Tables (IC)
- MEPSnet (HC)
- MEPSnet (IC)
- NHQRnet Data Query Tool

3. DATA TOOL TECHNICAL SPECIFICATIONS

The tables for each data tool included in this report contains final information that we have gathered through our AHRQ POCs. Section 4.2 includes a glossary that defines each category captured in the technical specifications table.

3.1 Consumer Assessment of Healthcare Providers and Systems (CAHPS) Online Reporting System

The **CAHPS** Online Reporting System presents results from two surveys:

- **CAHPS** Health Plan Survey (HP-CAHPS)
- **CAHPS** Clinician & Group Survey (CG-CAHPS)

Statistics are available for composite measures, rating measures, and individual survey items, organized according to survey version and field period.

Link: https://cahpsdatabase.ahrq.gov/CAHPSIDB/Public/about.aspx

3.1.1 CAHPS Health Plan Survey (HP-CAHPS)

HP-**CAHPS** Database contains commercial and Medicaid survey results submitted by various sponsors, including public and private employers, State Medicaid agencies, State Children's Health Insurance Programs (CHIP), and individual health plans since 1998.

3.1.2 CAHPS Clinician & Group Survey (CG-CAHPS)

CG-**CAHPS** contains survey data submitted by practice sites and practice groups on patient experiences with doctors and medical groups by region, physician specialty, practice ownership and affiliation, number of visits by practice and survey mode.

Table 1. HP-CAHPS/CG-CAHPS Technical Specifications

Category	Technical Specification
Software Requirements	
Operating System	Windows Server 2008 R2/IIS 7.5
Software Component/Programming Language	ASP.NET v4.6 SQL Server 2014 SAS IntrNet Server
Web Server	
Server Type	WebServer HP Proliant Hardware VMWare ESXI 6.5 SAS IntrNet Server Windows 2008 R2 4 vCPU

Category	Technical Specification
	Memory: 16GB C: 80 GB D: 120GB
Server Storage Size	Web Server: Hard disk: C:80GB D:180GB SAS Server: Hard disk: C:80GB D:120GB
Transaction Volume/Size	Not available
Memory Requirements	Web Server: Memory: 4GB virtual CPU: 4 SAS Server: Memory: 16GB virtual CPU:4
Database Server	
Server Type	SQL Server 2014 standard edition with SP2 OS: Windows 2012 R2
Server Version	2014
Server Needs	N/A
Storage Type	All CAHPS applications are hosted on web database servers located at Westat
Memory Requirements in RAM	Memory: 32GB, vCPU: 4
Disk Space/Storage Requirement	Hard disk: C:80GB D:200GB E:400GB
Data Preservation Requirements (in years)	9 Months after final option year. 9/2023
Dependencies	
System/Technical Dependencies	No
Staff/Manual Dependencies	System requires staff to load cleaned/processed data files from the HP-CAHPS and CG-CAHPS submission systems into the CAHPS Online Reporting system database;
Interfaces	
External	
Input System (Supply Data)	Data from the HP-CAHPS and CG-CAHPS data submission systems are loaded manually into Online Reporting System SQL Server database by executing SQL stored procedures
Output Systems (Extract Data)	No
Internal	
Input System (Supply Data)	No
Output Systems (Extract Data)	No
Data/Deployment	
Data Type	HP and CG Data files are submitted by vendors in CSV file format and are subsequently extracted, combined, edit checked, cleaned and augmented with derived variables and output as SAS datasets. The SAS data are in turn transformed and loaded into the CAHPS Online Reporting System using SQL Server stored procedures.
Runtime (duration)	Data from the HP-CAHPS and CG-CAHPS data submission systems are loaded into Online Reporting System SQL Server database by executing SQL stored procedures
Run Schedule	HP-CAHPS and CG-CAHPS data are loaded into the CAHPS Online Submission system once per year

Category	Technical Specification
File Size requirements	No
Data Sources	SQL Server database; SAS IntrNet Server
File Types Received	SQL Load files
File Types Available to Users	HTML and PDF
Format Displayed to Users	HTML and PDF
System Downtime	Scheduled maintenance every Friday for approx. 2 hours between 4am and 6am
Method of Data Transfer	Data from the HP-CAHPS and CG-CAHPS data submission systems are loaded into Online Reporting System SQL Server database by executing SQL stored procedures
Security Specification	
Can any user pull all available data?	DBA - only
Are there any type of security restrictions on this tool?	SSL/TLS, Role based authentication and authorization

3.2 The Healthcare Cost and Utilization Project's (HCUP) Fast Stats

HCUP Fast Stats provides easy access to the latest HCUP-based visual statistical graphs, maps, and tables for healthcare information topics. **HCUP Fast Stats** uses visual statistical displays in stand-alone graphs, trend figures, maps, or simple tables to convey complex information at a glance. **HCUP Fast Stats** are updated regularly (quarterly or annually, as newer data become available) for timely, topic-specific national and State-level statistics.

Link: https://www.hcup-us.ahrq.gov/Fast Stats/landing.jsp

See table below for government-provided technical specifications for **HCUP Fast Stats**.

Table 2. HCUP Fast Stats Technical Specifications

Category	Technical Specification
Software Requirements	
Operating System	Red Hat Enterprise Linux
Software Component/Programming Language	Java, Apache HTTPD, Tomcat, SimpleMaps, JSP, JSTL, XML, HTML
Web Server	
Server Type	Virtual Machine
Server Storage Size	90 GB
Transaction Volume/Size	N/A - There is no actual transactions or real-time data processing happening on the server. Web server is simply displaying HTML, XML, images, and spreadsheets that were prepared offline and copied on to the Web server for display purposes. Source data is stored separately and analyzed offline (i.e. not connected to the Internet). The installed third-party software (SimpleMaps) utilize aggregated results in a spreadsheet uploaded to the Web server to generate maps.
Memory Requirements	8 GB
Database Server	
Server Type	N/A - There is no actual transactions or real-time data processing happening on the server (see above, Transaction Volume/Size, for more details).
Server Version	N/A
Server Needs	N/A
Storage Type	N/A
Memory Requirements in RAM	N/A
Disk Space/Storage Requirement	N/A
Data Preservation Requirements (in years)	N/A
Dependencies	

Category	Technical Specification
System/Technical Dependencies	None. Web server is air gapped from data source. Web server is hosted in a different physical facility, 1500 miles apart.
Staff/Manual Dependencies	Contractor staff manually transfer files authorized for posting from offline data source to Production Web server for display.
Interfaces	
External	
Input System (Supply Data)	No actual data is inputted from external sources on to the system (Web server).
Output Systems (Extract Data)	Images, spreadsheets, and site content text approved by AHRQ for posting to Website is available for public download. Source data is not stored on the system.
Internal	
Input System (Supply Data)	Images, spreadsheets, and site content text approved by AHRQ for posting to Website is available for public download. Source data is not stored on the system.
Output Systems (Extract Data)	Images, spreadsheets, and site content text approved by AHRQ for posting to Website is available for public download. Source data is not stored on the system.
Data/Deployment	
Data Type	HTML, java script, images, and spreadsheets
Runtime (duration)	Website is running 24x7 except for monthly system maintenance downtime (approx. 1 hour).
Run Schedule	N/A - There is no actual transactions or real-time data processing happening on the server (see above, Transaction Volume/Size, for more details).
File Size requirements	Files sizes are not specified, but typically file sizes do not exceed 5 MB each.
Data Sources	HCUP and State Data Partners
File Types Received	Hypertext markup language - HTML Extensible markup language - XML JavaServer Pages - JSP Image files - PNG Spreadsheets - XLS
File Types Available to Users	Hypertext markup language - HTML Extensible markup language - XML JavaServer Pages - JSP Image files - PNG Spreadsheets - XLS
Format Displayed to Users	HTML and JSP

Category	Technical Specification
System Downtime	No downtime is required to update content to Website.
Method of Data Transfer	SSH upload for posting content. HTTPS/SSL for viewing/downloading content by users.
Can any user pull all available data?	Information posted to Website is publicly available. Underlying source data is not stored on this system (Web server).
Are there any type of security restrictions on this tool?	There is no security restriction on the Fast Stat tool hosted on the HCUP-US Website. It is publicly available. The underlying system (HCUP Web Services) hosting the Website has a FISMA Low ATO.

3.3 HCUPnet

HCUPnet is an online query system based on data from the Healthcare Cost and Utilization Project. **HCUPnet** provides healthcare statistics and information for:

- Hospital Inpatient
- Emergency Department
- Ambulatory Surgery
- County and sub-state region-level statistics

The **HCUPnet** website was recently redesigned in 2015 to simplify the process for querying data while providing more flexibility in changing the content that is displayed to the user.

Link: https://hcupnet.ahrq.gov/#setup

See table below for government-provided technical specifications for **HCUPnet**.

Table 3. HCUPnet Technical Specifications

Category	Technical Specification
Software Requirements	
Operating System	Windows 2012
Software Component/Programming Language	Front-End: Backbone (Marionette) API: Dot net Back-End: SQL Server 2014
Web Server	
Server Type	IIS Server 8.0
Server Storage Size	85GB
Transaction Volume/Size	264,668 hits/month
Memory Requirements	4GB
Database Server	
Server Type	SQL Server
Server Version	2014 SP2
Server Needs	2 CPU
Storage Type	VMWare (Cloud)
Memory Requirements in RAM	16 GB
Disk Space/Storage Requirement	4.5TB
Data Preservation Requirements (in years)	None
Dependencies	
System/Technical Dependencies	Data Prep in SAS is done by Social & Scientific Systems (SSS) before it can begin the ETL process into the HCUPnet database application
Staff/Manual Dependencies	Data Prep in SAS is done by Social & Scientific Systems (SSS) before it can begin the ETL process into the HCUPnet database application
Interfaces	

Category	Technical Specification
External	
Input System (Supply Data)	None
Output Systems (Extract Data)	User driven action, not a system feed: Facebook, Twitter, and Email "share" options on HCUPnet
Internal	
Input System (Supply Data)	AHRQ supplies SSS the data from HCUP partners
Output Systems (Extract Data)	None
Data/Deployment	
Data Type	Aggregated statistics
Runtime (duration)	Avg 3-5 mins. to load a dataset (CSV file) to SQL Server
Run Schedule	7-8 updates/year
File Size requirements	4-5 large updates, up to 40 GB, and 3 smaller updates, less than 5 GB
Data Sources	HCUP NIS (1993-2016), NEDS (2006-2015), NRD (2009-2015), SID (1997-2015), SEDD (2000-2014), and SASD (2012-2014) databases
File Types Received	CSV files
File Types Available to Users	Excel/CSV Exports from public site
Format Displayed to Users	Tables, Graphs
System Downtime	None
Method of Data Transfer	Export function via user interface on hcupnet.ahrq.gov
Security Specification	
Can any user pull all available data?	Yes
Are there any type of security restrictions on this tool?	None

3.4 MEPS Summary Tables (HC), MEPSnet (HC)

3.4.1 MEPS Summary Tables (HC)

MEPS Summary Tables (HC) provides frequently used summary estimates for the U.S. civilian non-institutionalized population on:

- Household medical utilization and expenditures, and population totals
- Health insurance coverage
- Access to care and experience with care
- Events and expenditures associated with select medical conditions
- Prescribed medicine purchases and expenditures

Most tables can be stratified by demographic or socioeconomic characteristics. Features include the ability to view trends over time, automated plot generation, and R and SAS code to replicate the selected tables.

Link: https://meps.ahrq.gov/mepstrends/home/index.html

See table below for government-provided technical specifications for **MEPS Summary Tables (HC)**.

Table 4. MEPS Summary Tables (HC) Technical Specifications

Category	Technical Specification
Software Requirements	
Operating System	CentOS 6 Operating System
Software Component/Programming Language	CSS, HTML, JavaScript
Web Server	
Server Type	Shiny Server Pro
Server Storage Size	Storage 40 GB Memory 4 GB
Transaction Volume/Size	35,774 hits/month
Memory Requirements	None
Database Server	
Server Type	None
Server Version	N/A
Server Needs	N/A
Storage Type	N/A
Memory Requirements in RAM	N/A
Disk Space/Storage Requirement	N/A
Data Preservation Requirements (in years)	No
Dependencies	
System/Technical Dependencies	Summary tables are created from AHRQ public use files. Creation of summary tables is done using 'R' (open-source programing language) to produce .csv files

Category	Technical Specification
	(intermediate) and then JSON files (summary tables).
Staff/Manual Dependencies	Data needs to be downloaded, packaged and sent to SSS
Interfaces	
External	
Input System (Supply Data)	None
Output Systems (Extract Data)	None
Internal	
Input System (Supply Data)	None
Output Systems (Extract Data)	None
Data/Deployment	
Data Type	HC summary tables are updated on the website using the JSON files
Runtime (duration)	4 hours to produce new tables
Run Schedule	Once per year
File Size Requirements	150 MB
Data Sources	AHRQ Public Use Files
File Types Received	SAS transport file (.ssp) from MEPS public use files, converted into JSON file
File Types Available to Users	.csv files, .png plot, R code, SAS code
Format Displayed to Users	web display
System Downtime	None
Method of Data Transfer	Secure FTP
Security Specification	
Can any user pull all available data?	Yes
Are there any type of security restrictions on this tool?	FISMA Compliance (Low)

3.4.2 MEPSnet (HC)

MEPSnet (HC) provides easy access to nationally representative statistics of health care use, expenditures, sources of payment, and insurance coverage for the U.S. civilian noninstitutionalized population. **MEPSnet (HC)** allows you to generate statistics using MEPS-HC public use files.

Link: https://meps.ahrq.gov/mepstrends/home/index.html
https://meps.ahrq.gov/mepsweb/data_stats/MEPSnetHC/startup

See table below for government-provided technical specifications for **MEPSnet (HC)**.

Table 5. MEPSnet (HC) Technical Specifications

Category	Technical Specification
Software Requirements	
Operating System	CentOS 6/Windows 12

Category	Technical Specification
Software Component/Programming Language	Java, SAS (SAS/IntrNet to create estimates in real time)
Web Server	
Server Type	IIS Server, SAS, Apache web server, Tomcat servlet container Apache manages the static content and Tomcat manages the dynamic content. The IIS server uses SAS/IntrNet to run SAS. Note: The Apache and Tomcat servers are used by MEPSnet – HC, MEPS Summary Tables (IC), and MEPSnet – IC as well as the remainder of the MEPS website.
Server Storage Size	Apache Web Server: Storage: 90 GB Memory: 7 GB Tomcat Server: Storage: 140 GB Memory: 9 GB IIS Server (SAS Server): Storage: 120 GB Memory: 4 GB
Transaction Volume/Size	5,678 hits/month
Memory Requirements	None
Database Server	
Server Type	Oracle Note: One server is used by MEPSnet – HC, MEPS Summary Tables (IC), and MEPSnet – IC, as well as the remainder of the MEPS website, NHQRnet, and PSO
Server Version	12c Database
Server Needs	None
Storage Type	VMWare (Cloud)
Memory Requirements in RAM	13 GB
Disk Space/Storage Requirement	550 GB
Data Preservation Requirements (in years)	No
Dependencies	
System/Technical Dependencies	Queries FYC micro-data file directly, calculates estimates in SAS, displays SAS output

Category	Technical Specification
Staff/Manual Dependencies	SSS produces a subset of the FYC file (single micro-data table for each data year) as a SAS dataset
Interfaces	
External	
Input System (Supply Data)	None
Output Systems (Extract Data)	None
Internal	
Input System (Supply Data)	None
Output Systems (Extract Data)	None
Data/Deployment	
Data Type	SAS is run to produce the SAS datasets needed from a new FYC file. Metadata are created using SAS and uploaded into the Oracle database.
Runtime (duration)	30 min
Run Schedule	Once per year
File Size requirements	1 GB
Data Sources	AHRQ Public Use Files (FYC)
File Types Received	SAS transport file (.ssp)
File Types Available to Users	None
Format Displayed to Users	Web display
System Downtime	None
Method of Data Transfer	SAS files are uploaded to the SAS server using Secure FTP. Metadata are uploaded using SAS/Access
Security Specification	
Can any user pull all available data?	Yes
Are there any type of security restrictions on this tool?	FISMA Compliance (Low)

3.5 MEPS Summary Tables (IC), MEPSnet (IC)

3.5.1 MEPS Summary Tables (IC)

The MEPS IC summary tables provide data on health insurance gathered from a survey of establishments and governments. Data include private sector data by firm size, industry group, census region and other characteristics; public sector data by state and local government type, government size, and census division; civilian estimates that incorporate both the private and state and local government sectors; national totals for enrollees and cost of health insurance coverage for both the private and public sectors.

Link:
https://meps.ahrq.gov/mepsweb/data_stats/quick_tables_search.jsp?component=2&subcomponent=1

See table below for government-provided technical specifications for **MEPS Summary Tables (IC)**.

Table 6. MEPS Summary Tables (IC) Technical Specifications

Category	Technical Specification
Software Requirements	
Operating System	CentOS 6 Operating System
Software Component/Programming Language	CSS, HTML, Static HTML and PDF files
Web Server	
Server Type	Apache web server, Tomcat servlet container
Server Storage Size	Apache Web Server: Storage: 90 GB Memory: 7 GB Tomcat Server: Storage: 140 GB Memory: 9 GB
Transaction Volume/Size	20,926 hits/month
Memory Requirements	None
Database Server	
Server Type	Oracle
Server Version	12c Database
Server Needs	None
Storage Type	VMWare (Cloud)
Memory Requirements in RAM	13 GB
Disk Space/Storage Requirement	550 GB
Data Preservation Requirements (in years)	No
Dependencies	
System/Technical Dependencies	None
Staff/Manual Dependencies	Census conducts survey and compiles data into Microdata (not available to AHRQ) and

Category	Technical Specification
	creates the Summary Tables (for public) as SAS datasets. SAS is used to generate HTML files and Apache FOP is used to generate PDF files. Staff run queries to add information into Oracle tables about new IC summary tables
Interfaces	
External	
Input System (Supply Data)	None
Output Systems (Extract Data)	None
Internal	
Input System (Supply Data)	None
Output Systems (Extract Data)	None
Data/Deployment	
Data Type	SAS datasets – summary tables
Runtime (duration)	15 min
Run Schedule	2x per year (Jan, Jun)
File Size Requirements	60 MB
Data Sources	Census: Summary Tables
File Types Received	SAS dataset
File Types Available to Users	Excel, HTML, and PDF, Zip files
Format Displayed to Users	Web display
System Downtime	None
Method of Data Transfer	Secure FTP
Security Specification	
Can any user pull all available data?	Yes
Are there any type of security restrictions on this tool?	FISMA Compliance (Low)

3.5.2 MEPSnet (IC)

MEPSnet (IC) provides easy access to national and state level statistics and trends about health insurance offered by private establishments and state and local governments. **MEPSnet (IC)** guides you step-by-step in locating statistics of interest across all available years using data from the MEPS Summary Tables (IC).

Link: https://meps.ahrq.gov/mepsweb/data_stats/MEPSnetIC/startup

See table below for government-provided technical specifications for **MEPSnet (IC)**.

Table 7. MEPSnet (IC) Technical Specifications

Category	Technical Specification
Software Requirements	
Operating System	CentOS 6 Operating System
Software Component/Programming Language	Java
Web Server	

Category	Technical Specification
Server Type	Apache web server, Tomcat servlet container
Server Storage Size	Apache Web Server: Storage: 90 GB Memory: 7 GB Tomcat Server: Storage: 140 GB Memory: 9 GB
Transaction Volume/Size	4,625 hits/month
Memory Requirements	None
Database Server	
Server Type	Oracle
Server Version	12c Database
Server Needs	None
Storage Type	VMWare (Cloud)
Memory Requirements in RAM	13GB
Disk Space/Storage Requirement	550 GB
Data Preservation Requirements (in years)	No
Dependencies	
System/Technical Dependencies	Queries, Oracle Tables
Staff/Manual Dependencies	Census conducts survey and compiles data into Microdata (not available to AHRQ) and Summary Tables (for public) as SAS datasets; SSS then standardizes the Census tables (Master SAS dataset)
Interfaces	
External	
Input System (Supply Data)	None
Output Systems (Extract Data)	None
Internal	
Input System (Supply Data)	None
Output Systems (Extract Data)	None
Data/Deployment	
Data Type	SAS datasets from summary tables are standardized by SSS into Master SAS dataset, which is uploaded into an Oracle database
Runtime (duration)	15 min
Run Schedule	2x per year (Jan, Jun)
File Size Requirements	80 MB
Data Sources	Census: Summary Tables
File Types Received	SAS dataset
File Types Available to Users	PDF
Format Displayed to Users	Plot graphs with statistical Z-tests; web display, including view over time

Category	Technical Specification
System Downtime	None
Method of Data Transfer	Tables are uploaded using SAS/Access
Security Specification	
Can any user pull all available data?	Yes
Are there any type of security restrictions on this tool?	FISMA Compliance (Low)

3.6 National Healthcare Quality and Disparities Reports (NHQRnet) Data Query Tool

This integrated website provides a unified data query tool for investigating information presented in the National Healthcare Quality and Disparities Reports (**NHQRnet**). It allows users to drill down from the broadest picture of healthcare quality and disparities on the national and state levels.

Link: https://nhqrnet.ahrq.gov/inhqrdr/

See table below for government-provided technical specifications for **NHQRnet**.

Table 8. NHQRnet Technical Specifications

Category	Technical Specification
Software Requirements	
Operating System	CentOS 6 Operating System
Software Component/Programming Language	Java: J2EE Technologies (Java, Groovy, Groovy Server Pages, Grails, Spring, Hibernate, JQuery), ForeSee survey
Web Server	
Server Type	Apache web server, Tomcat servlet container
Server Storage Size	Web application uses two different servers. Apache manages the static content and Tomcat manages the dynamic content. Here are the storage and memory requirements for the two servers. Apache Web Server: Storage: 40 GB Memory: 3 GB Tomcat Server: Storage: 200 GB Memory: 9 GB
Transaction Volume/Size	120,000 hits/month
Memory Requirements	None
Database Server	
Server Type	Oracle (Note that one server is used by three web sites. MEPS, NHQRnet and PSO)
Server Version	12c Database
Server Needs	None
Storage Type	VMWare (Cloud)
Memory Requirements in RAM	13 GB
Disk Space/Storage Requirement	550 GB
Data Preservation Requirements (in years)	None
Dependencies	

Category	Technical Specification
System/Technical Dependencies	IQDnet - this tool depends on IQDnet as the main repository for its data, which stores the about 40 microdata from national healthcare surveys administered by AHRQ, CDC, HRSA, IHS, NIH, SAMHSA, AHA, NHPCO and FEHCS SAS-DDE macro programs formats tables, process and builds the data into Oracle data
Staff/Manual Dependencies	Unstructured/scattered data needs formatting, filtering and mapping to become usable to users, Social & Scientific Systems (SSS) complete these steps after receiving the data
Interfaces	
External	
Input System (Supply Data)	None
Output Systems (Extract Data)	None
Internal	
Input System (Supply Data)	IQDnet, which is sourced from AHRQ, CDC, HRSA, IHS, NIH, SAMHSA, AHA, NHPCO and FEHCS
Output Systems (Extract Data)	None
Data/Deployment	
Data Type	Raw aggregated statistics from diverse microdata
Runtime (duration)	15 min to load SAS datasets to Oracle
Run Schedule	Once a year
File Size Requirements	Maximum size for data source files is 100 MB
Data Sources	External: CDC, NIH, HRSA, IHS, SAMHSA, AHA, HHS, NHPCO, FEHCS (unstructured and scattered data) Internal: IQDnet (main repository, updates annually)
File Types Received	SAS dataset and Excel, Word, PDF, Web tools
File Types Available to Users	XLS, HTML
Format Displayed to Users	Excel table, Plot graphs, Web Tables
System Downtime	No
Method of Data Transfer	Tables are uploaded using SAS/Access Static content is uploaded using Secure FTP
Security Specification	
Can any user pull all available data?	Yes
Are there any type of security restrictions on this tool?	FISMA Compliance (Low)

4. APPENDICES

4.1 Technical Specifications Worksheet

To see the technical specifications of all data tools together, please double-click on this embedded spreadsheet below.

(removed)

4.2 Technical Specifications Glossary

Table 9. Technical Specifications Glossary

Category	Definition
System Requirements	
Operating System	Current operating system for the tool
Software Component/Programming Language	Language that the current tool operates in for both front and back end (e.g., CGI, ASP, Java, Connection Wizard)
Web Server	
Server Type	Server type that's required to host the tool
Server Storage Size	Current memory capacity for the tool
Transaction Volume/Size	Estimated number of transactions the tool has on average
Memory Requirements	Any additional requirements for memory storage or RAM that are needed for the web app
Database Server	
Server Type	Server type that's required to store the tool's data
Server Version	Version of server if available
Server Needs	Any requirements needed to store the web apps memory
Storage Type	Local or cloud storage
Memory Requirements in RAM	Required memory space
Disk Space/Storage Requirement	Required disk space
Data Preservation Requirements (in years)	Any purging requirements for storage
Dependencies	
System/Technical Dependencies	The technical steps required to extract, transform, load (ETL) external data into the database
Staff/Manual Dependencies	The manual steps required to ETL external data into the database
Interfaces	
External	
Input System (Supply Data)	Interfaces outside of AHRQ that feed data into this tool

Category	Definition
Output Systems (Extract Data)	Interfaces outside of AHRQ that this tool feeds into
Internal	
Input System (Supply Data)	Interfaces inside of AHRQ that feed data into this tool
Output Systems (Extract Data)	interfaces inside of AHRQ that this tool feeds data into
Data/Deployment	
Data Type	Description of the data (e.g., aggregated, microdata)
Runtime (duration)	Length of time to upload the data
Run Schedule	Frequency of uploading data
File Size Requirements	Maximum size allowed for the data source file
Data Sources	Name of source organizations or entities where data for this data tool comes from
File Types Received	In what format does the source organization send the data to AHRQ
File Types Available to Users	In what format can the user download the data (e.g., .csv)
Format Displayed to Users	In what format is the information displayed to the user (e.g., web display)
System Downtime	Any downtime needed to load data into the data tool
Method of Data Transfer	How does AHRQ receive the data from the data source (e.g., email, direct interface)
Security Specification	
Can any user pull all available data?	N/A
Are there any type of security restrictions on this tool?	N/A

4.3 Acronyms and Definitions

Table 10. List of Acronyms and Definitions

Acronym	Description
AHA	American Hospital Association
AHRQ	Agency for Healthcare Research and Quality
API	Application Program Interface
ASP	Active Server Pages
ASP.NET	Newest version of Active Server Pages
CAHPS	Consumer Assessment of Healthcare Providers and Systems
CDC	Centers for Disease Control and Prevention
CG-CAHPS	Clinician and Group Survey for the Consumer Assessment of Healthcare Providers and Systems
CGI	Common Gateway Interface
COTS	Commercial off the-Shelf
CPU	Central Processing Unit
CSS	Cascading Style Sheets
CSV	Comma-Separated Values (File Extension)
DBA	Data Base Administrator
ETL	Extract, Transform, and Load
FEHCS	Family Evaluation of Hospice Care Survey
FISMA	Federal Information Security Management Act of 2002
FOP	Formatting Objects Processor
FTP	File Transfer Protocol
FYC	Forward Yield Curve (File extension)
GB	Gigabyte
HC	Household Component
HCUP NIS (1993-2016)	Healthcare Cost and Utilization Project's National (Nationwide) Inpatient Sample
HHS	U.S. Department of Health and Human Services
HP	Hewlett-Packard
HP-CAHPS	Health Plan Survey for the Consumer Assessment of Healthcare Providers and Systems
HRSA	Health Resources and Services Administration
HTML	Hyper Text Markup Language
HTTP	Hypertext Transfer Protocol
HTTPD	Hypertext Transfer Protocol Daemon
HTTPS	Hypertext Transfer Protocol Secure
IC	Insurance Component
IIS	Internet Information Services
IHS	Indian Health Service
IQDnet	Internal Quality/Disparities Network
J2EE	Java 2 Platform Enterprise Edition
JSON	JavaScript Object Notation

Acronym	Description
JSP	JavaServer Pages
JSTL	JavaServer Pages Standard Tag Library
MB	Megabyte
MEPS	Medical Expenditure Panel Survey
MHPCO	National Hospice and Palliative Care Organization
N/A	Not applicable
NEDS (2006-2015)	Nationwide Emergency Department Sample
NHPCO	National Hospice and Palliative Care Organization
NHQRnet	National Healthcare Quality and Disparities Reports Data Query Tool
NIH	National Institutes of Health
NRD (2009-2015)	Nationwide Readmissions Database
OS	Operating System
PDF	Portable Document Format (File Extension)
PNG	Portable Network Graphics (File Extension)
POC	Point of Contact
PSO	Patient Safety Organization
R	R Programming Language
R2/IIS 7.5	Windows Server 2008, Internet Information Services (Server OS)
RAM	Random-access Memory
SAMHSA	Substance Abuse and Mental Health Services Administration
SAS	Statistical Analysis Software®
SASD (2012-2014)	State Ambulatory Surgery and Services Database
SAS-DDE	Statistical Analysis Software® Dynamic Data Exchange
SEDD (2000-2014),	State Emergency Department Database
SID (1997-2015),	State Inpatient Databases
SP2	Windows Server 2008 Service Pack 2
SQL	Structured Query Language
SSH	Secure Shell
SSL/TLS	Secure Sockets Layer/Transport Layer Security
SSP	Professional Screen Saver Producer (File Extension)
SSS	Social and Scientific Systems
TB	Terabyte
XLS	Microsoft Excel (File Extension)
XML	Extensible Markup Language

RFTO #10: Developing Strategies for the Harmonization of AHRQ Data Tools (Data Tools)

Environmental Scan Results Report

February 4, 2019

CONTENTS

REVISION HISTORY

Date	Version	Description	Author(s)
January 11, 2019	Draft	Initial draft, will be finalized after client comments have been incorporated	Atlas Team
January 18, 2019	AHRQ Review	AHRQ review, corrections, and feedback	Agency for Healthcare Research & Quality
January 25, 2019	Final	AHRQ feedback incorporated into the document	Atlas Team
February 4, 2019	Final Review and Updates	Updated report based on AHRQ's final feedback	AHRQ and Atlas Team

EXECUTIVE SUMMARY

The Agency for Healthcare Research and Quality (AHRQ) provides several data tools and summary data tables that have dedicated userbases but are siloed and lack cohesion of user experience (UX) and branding among them.

To better understand both the potential for and challenges facing the harmonization of the user interfaces (UI) and basic query functionality of these data tools, the Atlas Team has conducted an environmental scan intended to understand the unique and varied needs of the tools' users and stakeholders.

This scan has targeted three research sources:

- External Users
- Internal Users & Stakeholders
- Comparable Agencies

Three subject areas were investigated for each of these research sources:

- UX
- Branding
- The barriers to and opportunities for data tool harmonization

The specific AHRQ data tools included in the scope of this scan are:

- The Consumer Assessment of Healthcare Providers and Systems (CAHPS) Aggregated databases
- The Health Cost and Utilization Project's (HCUP) Fast Stats
- HCUPnet
- Medical Expenditure Panel Survey (MEPS) Summary Tables: Household Component (HC)
- MEPS Summary Tables: Insurance Component (IC)
- MEPSnet (HC)
- MEPSnet (IC)
- National Healthcare Quality and Disparities Reports Data Query Tool (NHQDRnet)

The environmental scan has incorporated Human-Centered Design (HCD) methods of user research and analysis to best align project outcomes with the needs of both the users and the business owners. The environmental scan has consisted of customer identification, research planning, data gathering, synthesis, actionable interpretation of the findings, and a Design Thinking workshop with the data tools points of contact (POCs). The Atlas Team interviewed 26 users and proxy users, resulting in 1,026 data points and 47 recommended user requirements for harmonized UI and a commercial off-the-shelf (COTS) data visualization and analysis product.

Findings

Using an HCD approach to conduct qualitative research of AHRQ data tool users and synthesize the findings, the Atlas Team determined that there are significant opportunities to build a harmonized UI based on the overlapping needs of the users of the various tools, as well

as general opportunities to improve usability. Specifically, two user groups and five user types were identified:

- **External Users:**
    - Pro
    - Intermediate
    - Novice / Quick Reference
- **Internal Users:**
    - Support
    - Administrator

Section 4.2 provides a description of each user type.

The following themes of user behavior were identified through these HCD techniques:

- **External User Themes:**
    - Navigating the Data Tools with Ease
    - Knowing Where to Start My Query
    - Understanding Methods & Variables
    - Accessing Data Files
    - Formulating & Refining a Data Tool Query
    - Reviewing the Data
    - Seeking Support
    - Completing My Query Task
- **Internal User Themes:**
    - Navigating the Data Tools with Ease
    - Receiving & Resolving User Queries
    - General Support
    - Customer Outreach
    - Data Tool Management
    - Ad Hoc Use of Tools
    - Quality Control

The Atlas Team will incorporate these themes into further analysis of stakeholders to optimize the design process for two of this project's subsequent deliverables of Style Examples and the Prototype.

Analysis

The Atlas Team's Environmental Scan analysis comprised the following phases:

1. Qualitative Analysis of Interviews (Affinity Diagramming)
2. User Requirements Development
3. Quantitative Analysis of Card Sort Activities

1. INTRODUCTION

AHRQ requires a harmonized UI and data visualization and analysis COTS product that meets the overlapping needs of the users of its various data query tools. To this end, the Atlas Team conducted an Environmental Scan to evaluate the challenges, behaviors, and tools of the users of AHRQ's various web-based data tools and summary data tables. Using an HCD approach to conduct qualitative research and synthesize the findings, the Atlas Team engaged 26 users and stakeholders by conducting interviews, data tool observation, and card sort activities. This yielded 1,026 data points, which were synthesized into 47 User Stories.

2. PROJECT OBJECTIVES

The purpose of this Environmental Scan was to gain an understanding of the behaviors, resources, and challenges of a broad-based sample of AHRQ data tool users as it pertains to the UI design and functionality of the data query tools and summary data tables.

The original scope of work proposed customer identification, organizing of interviews, interview protocol development, data gathering, analysis, and interpretation. These goals were achieved.

3. PROJECT APPROACH

The Atlas Team used an HCD approach to conduct qualitative research and synthesize the findings. The goal of this iterative design research methodology is to learn directly from the people experiencing a given issue, create a broad variety of creative solutions, and then test and design solutions based on what's most desirable, feasible, and viable for those stakeholders (see Figure 1).

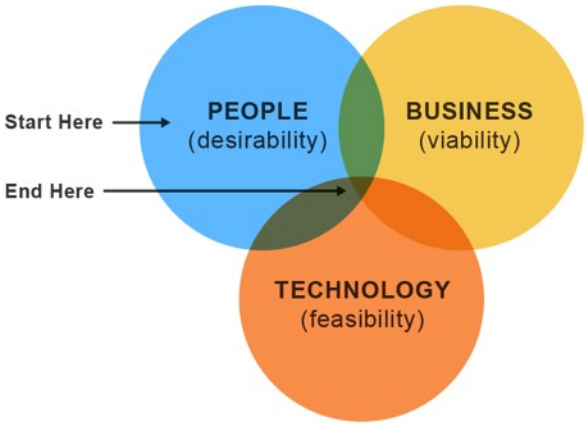


Figure 1. Human Centered Design Overview

HCD employs a research and design process called Design Thinking which begins with conducting qualitative research of the behaviors, resources, and settings of relevant stakeholders. That research is then synthesized to identify key themes, insights and opportunities. After this, a team collaboratively brainstorms and rapidly prototypes ideas, concepts, potential solutions and accompanying experiments where those solutions can be tested by real "users". With each new cycle of testing, the Design Thinking team comes closer to a solution more optimized to user needs (Figure 2). Importantly, the inclusion of relevant stakeholders, including the customer, throughout this process is essential to user-centric outcomes for any proposed product or service.

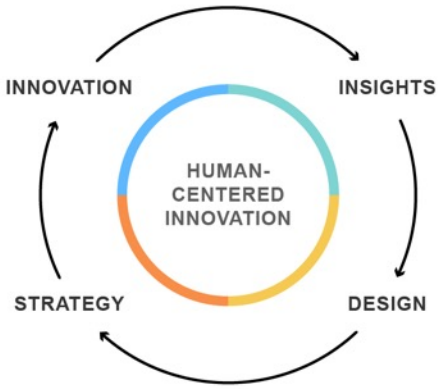


Figure 2. Product Development informed by Design Thinking

3.1. Interviews

The Atlas Team, with the help of AHRQ data tools POCs, set out to interview a broad-based sample of the user representatives of all data tools and summary data tables. The Atlas Team created a prioritized roster of key stakeholders in collaboration with the AHRQ POCs for each of the data tools. The Atlas Team developed a standardized discussion guide for each user type (sample questions from these discussion guides are in Appendix B.1).

Interviews commenced on November 30th, 2018 and will conclude on January 17th, 2019. During that time, the Atlas Team reached out to 38 individuals (with a response rate of approximately 68%) and interviewed 26 individuals including 17 internal AHRQ users and proxy users, 5 non-AHRQ federal employees, and 4 non-federal employees.[1]* The following is a breakdown of interviewees by tool:

- CAHPS: 2
- HCUP Fast Stats: 2
- HCUPnet: 4
- MEPS summary tables (HC) / MEPSnet (HC): 7
- MEPS summary tables (IC) / MEPSnet (IC): 7
- NHQDRnet: 4

The primary interviewers were Cary Euwer and Dennis Legislador. Emily Mitchell and Kevin Heslin of AHRQ were the primary interview recruiting contacts at AHRQ. AHRQ and the Atlas Team agreed at the outset of interview phase to keep the participants' identification anonymous in deliverables.

3.2. Observation

In addition to targeted questions about the UX, the Atlas Team observed simulated queries in each of the data tools and summary data tables (as well as some analogous data tools). Attendees were observed using their work computer either in-person at AHRQ or via a remote video conference. Interviewers asked participants to walk through use cases of each of the data tools with which they were familiar while the interviewer made note of user behavior, work-arounds, pain points, and opportunities for improvement. Participants were asked to narrate their task from impetus to output, noting any points of confusion. At the end, the participants were asked if there were any features they'd prefer to be included or omitted in this particular tool, and, if relevant, to guide the interviewer through a similar product with an optimal user experience from the perspective of their use case(s).

3.3. Card Sorts

A Card Sort is a design research method for determining users' mental models of how information is organized and prioritized. Card Sorts are typically used to help inform the development of site maps, as they support development teams in deciding which features to group and promote within the site based on real users' feedback. The Atlas Team used Card Sorting as a means to identify which functions of the data tools certain user groups are drawn to and to understand their level of importance. The Atlas Team pulled 85 high-level terms that are

[1] Please note that extensive effort was taken to contact all user groups; however, given the proximity of this project to other recent research and the holidays, as well as end of year tasks, the responses to both interview recruiting, and interviews scheduling efforts was low despite repeated, effortful attempts

representative of navigation and functions from all the data tools and summary data tables. With these terms, the Atlas Team facilitated 24 in-person and remote card sort activities and subsequent analysis.

4. PROJECT OUTCOMES

The insights of all 26 stakeholders interviewed were consolidated into 2 user groups (External Users and Internal Users) and 5 user types (Pro, Intermediate, Novice/Quick Reference, Support, and Administrator). While each user might use the platform differently, functionality will overlap for a majority of the users.

Each user group includes a brief description of their role and key challenges, themes of user behavior, and a sample of high-level, "priority" User Stories with descriptions. User Stories are used in HCD and software development to be specific for the types of functionality users require, with a justification, and to determine which users require which functionalities. User stories are typically structured in this way "As a **USER**, I need **Insert Functionality Here**, in order to **Insert Justification Here**." As an example: "As a Pro User, I need to be able to easily download data, in order to be able to conduct my own independent data wrangling." The full list of User Stories follows at the end of each user group section. Based on these User Stories, user requirements can be distilled into easily comprehensible functionality and desires from users.

4.1. Structure of the Environmental Scan Review

- **User Findings Summary:** Section 4.2 outlines and describes the identified user types and their consolidated themes of activity as they relate to AHRQ's data tools and summary data tables.
- **User Group:** A high-level description of the user groups (identified in Section 4.2)
    - **User Type:** Subsequent sections (Section 4.3 and Section 4.4) contain a description of the user types within each group, as well as descriptions of their role, relevance, and key challenges.
        - **Theme:** Within each of the User Type sections are sub-sections by theme of user behavior of that type. These high-level themes characterize this user group's usage of AHRQ's web-based data tools and summary data tables.
            - **Featured User Story:** A select number of user stories are used to highlight fundamental user requirements, followed by a description of that requirements and its rationale.
                - **Agency Comparison:** Relevant design lesson from analogous products to the data tools in question.
- **Consolidated User Stories:** Section 4.5 contains the full list of User Stories for all User Groups and Types.
    - **Theme:** All user requirements are organized by theme.
        - **Requirements List:** All user requirements are cross-referenced with the user groups they satisfy.
- **Card Sort Analysis:** Section 4.6 features the accumulated results of the card-sort activity for all user groups and their implications for information hierarchy and site mapping for a harmonized UI.

- **General Usability Best Practices:** Section 4.7 outlines recommended usability design best practices for the creation of a harmonized UI.
- **Appendix Section:** For further reference, the appendices include a list of acronyms featured throughout the document, sample discussion questions from our interviews, and sample data points used in our Design Thinking-based analysis (Affinity Diagrams).

4.2. User Findings Summary

Based on conclusions drawn from interviews, the Atlas Team has defined five types of users of AHRQ's Summary Data Tables & web-based data query tools across two broad user groups, which are:

- **External Users:**
    - **Pro:** Seasoned researchers, including (but not limited to) graduate students, academics, researchers (often contributing to publications or journals), economists, and statisticians. Characterized by a cursory use of the data query tools and summary data tables, they are fluent in manipulating raw data and often conduct analysis of these data files using 3rd party applications (e.g. R, SAS, Stata) to conduct analyses and, in some cases, reach publication-worthy conclusions.
    - **Intermediate:** Researchers including (but not limited to) federal government agencies (including AHRQ), state government agencies, non-profit and for-profit healthcare providers, advocacy groups, policymakers, and some researchers. Characterized by a limited fluency with manipulating and conducting analyses of raw data, **they are the most heavily reliant on the data query tools provided by AHRQ of all the user types**.
    - **Novice / Quick Reference:** Low-intensity users including (but not limited to) the media, undergraduate students, economists, congressional staffers and some federal agencies, insurance companies, and the general public. Characterized by their reliance on AHRQ technical support requests to help find relevant data; this reliance is due either to poor usability of some of AHRQ's data tools, or time constraints, or the need for custom data queries that the digital tools are unable to provide. Those that do not engage with support tend to seek quick, straightforward answers to queries via user-friendly data tools and summary data tables.

- **Internal Users:**
    - **Support:** Customer support users include (but are not limited to) statisticians, researchers, media and policy relations staff, and database contractors who help field Technical Assistance Requests, special runs, and general support and outreach as it relates to the data query tools and summary data tables. As a result, they have important insight into the needs and challenges of External Users, much of which has been incorporated into that section of this document. For the purpose of providing customer support, this user type is most similar to Intermediate Users in their reliance on and fluency with the query tools for quick reference in response to these requests.
    - **Administrator:** This user type is characterized by all use cases of the data tools that are not directly related to customer support. Stakeholders included in this user type include (but are not limited to): statisticians, researchers / research scientists, research division directors, project managers, writer-editors, and database operations contractors. Their use of the data tools is most similar to Pro

Users in that they largely forgo the tools and summary tables in favor of running their own data. However, the themes of user behavior as it pertains to the data tools and summary data tables encompass general usability needs, the management of the databases and corresponding tools, ad hoc or miscellaneous usage of the tools, and quality control of content including stat briefs, reports, and data tool content.

The collective behavior of External Users falls into the following themes of activity:

- **Navigating the Data Tools with Ease:** entails all basic usability needs for users of each AHRQ product and for a harmonized UI across all AHRQ products.
- **Knowing Where to Start My Query:** understanding the process for choosing to engage with a particular data tool or summary data table: starting with the initial motivation, the decision-making process for choosing a particular tool, how to navigate to that tool, and determining which features to engage with to complete a task.
- **Understanding Methods & Variables:** all tasks related to understanding the dataset in question, including methods of data gathering and how the resulting variables are stored.
- **Accessing Data Files:** the steps needed to download raw data files.
- **Formulating & Refining a Data Tool Query:** the process of engaging with the primary UI for conducting a query, including reviewing and selecting from a list of results.
- **Reviewing the Data:** behavior related to the review of the selected data, preferred formats and analysis.
- **Seeking Support:** the nature of support varies widely among the user types, ranging from requesting a special run of a query for specific data results, to submitting a technical assistance request.
- **Completing My Query Task:** final tasks to complete a particular use case of the data tool or summary data table.

The collective behavior of Internal Users falls into the following themes:

- **Navigating the Data Tools with Ease:** entails all basic usability needs for users of each AHRQ product and for a harmonized UI across all AHRQ products.
- **Receiving & Resolving User Queries:** all passive behavior relating to receiving Technical Assistance Requests, special runs, or general inquiries from customers, and resolving them.
- **General Support:** broader customer support tasks unrelated to direct communications regarding data tool queries.
- **Customer Outreach:** active measures by AHRQ staff to push data content related to their specific data tool to users.
- **Data Tool Management:** all activities relating to the relevant backend management needs of the data tool in question.
- **Ad Hoc Use of Tools:** behavior related to miscellaneous, impromptu use of the data tools.
- **Quality Control:** activities in which the data tools are used for the purpose of quality control of data, reports, statistical briefs, and other products.

While there is some overlap in their use cases and behavior, some users engage in certain themes of behavior more than others. In the following sections, these themes serve to organize the observed behaviors, challenges, and opportunities presented by each user type.

4.3. External Users

4.3.1. Pro Users

This group includes a variety of use cases for seasoned researchers, including (but not limited to) graduate students, academics, researchers contributing to publications or journals, and statisticians. The group is mainly characterized by a cursory use of the data query tools and summary data tables for the purpose of validation or discovery (for instance, at the outset of a research initiative), with heavy usage of the raw data files to conduct analyses. Any use of the data query tools is focused and brief for the purposes of quick reference and validation often accompanied by the initial discovery phases of research.

Our Pro Users were researchers characterized by their wide-ranging use cases for the various tools. These researchers included a PhD student who performed basic queries to verify estimates and gain contextual understanding at crucial points of research for her dissertation, while performing analysis on the raw data files in Stata. It also included a senior scientist who outsourced analysis to contractors, preferring raw data files over data tools/summary tables, while using the search functionality for high-level, targeted queries. Unlike other External Users (i.e., Intermediate or Novice Users), Pro Users are less likely to submit technical assistance requests to AHRQ.

4.3.1.1. Navigating the Data Tools with Ease

For Pro Users, the foremost usability principle for their workflow is versatility. This principle is reinforced by the rich datasets these users rely upon, the relatively broad array of topics in the purview of their research, their methods of engaging with the data, from impromptu quick-reference queries to granular analysis of raw data files. And while not unique to this user group, the user group often relied on using Google to uncover the datasets and tools provided by AHRQ.

“I need a user-friendly interface with highly streamlined & visualized content – preferably that can be viewed with minimal scrolling and scanning of extraneous text – in order to reduce the cognitive burden of browsing complex information.”

While the functionality of MEPSnet was found to be the most complimentary data query tool to Pro Users workflow, there was much praise for the intuitive and topic-oriented interface of the (relatively infrequently-used) MEPS Summary Tables. As will be discussed in later “General Usability Principles” section, the use of visual indicators and effective information hierarchies is proven to lessen the cognitive burden of navigating and using a graphical UI and generally improve the UX. As will be indicated in other user types, this usability best practice can be implemented to great benefit to users engaging with complex data or working under tight timelines while not compromising necessary text. For this user type in particular, we found that due to time constraints and a level of familiarity with their typical workflow, they often won’t take the time to explore the full breadth of functionality on a data tool, sticking instead to what they know.

"I need access to the archived version of original site in order to reference where I know information was while transitioning to a new interface."

"I need a similar site map and terminology to the original site in order to continue my use of the site relatively unabated."

While streamlining and visualizing content is a widely-requested suggestion across all users interviewed, the data tools in question have well-established audiences who have grown accustomed to aspects of the established content layout and terminology. However, this transition to a more usable interface has precedent at AHRQ: when transitioning to an improved UI, HCUPnet retained an archived version of the query tool on their site to help ease tried-and-true users into the new interface. If AHRQ decides to pursue streamlined content as part of its efforts to harmonize data tools' UIs, an alternative option is merely to conform any updated content as closely as possible to the terminology of the original site.

4.3.1.2. Knowing Where to Start My Query

Pleasant interactions with digital tools involve both clear starting points combined with specific calls to action and an understanding of where to go next. For example, if you're booking a flight on a travel site, knowing that you need to provide when and where you're traveling is an idealized version of "knowing when to start". When first entering a new data tool, the Atlas Team observed certain behaviors, challenges, opportunities, and "best practice" resources. Pro Users, exhibiting use cases for engaging with both the query tools and the data files, are looking for certain characteristics of both that will help determine whether they will engage.

The behaviors and needs identified for "Knowing Where to Start My Query" have significant implications for a unified data tools portal. These implications could potentially address some branding issues identified over the course of the Environmental Scan, particularly a perception among a few users that AHRQ's focus spread thin across disparate data efforts; as such, there is tremendous potential to use a user-friendly portal and UIs to reinforce a harmonized brand and upend this perception.

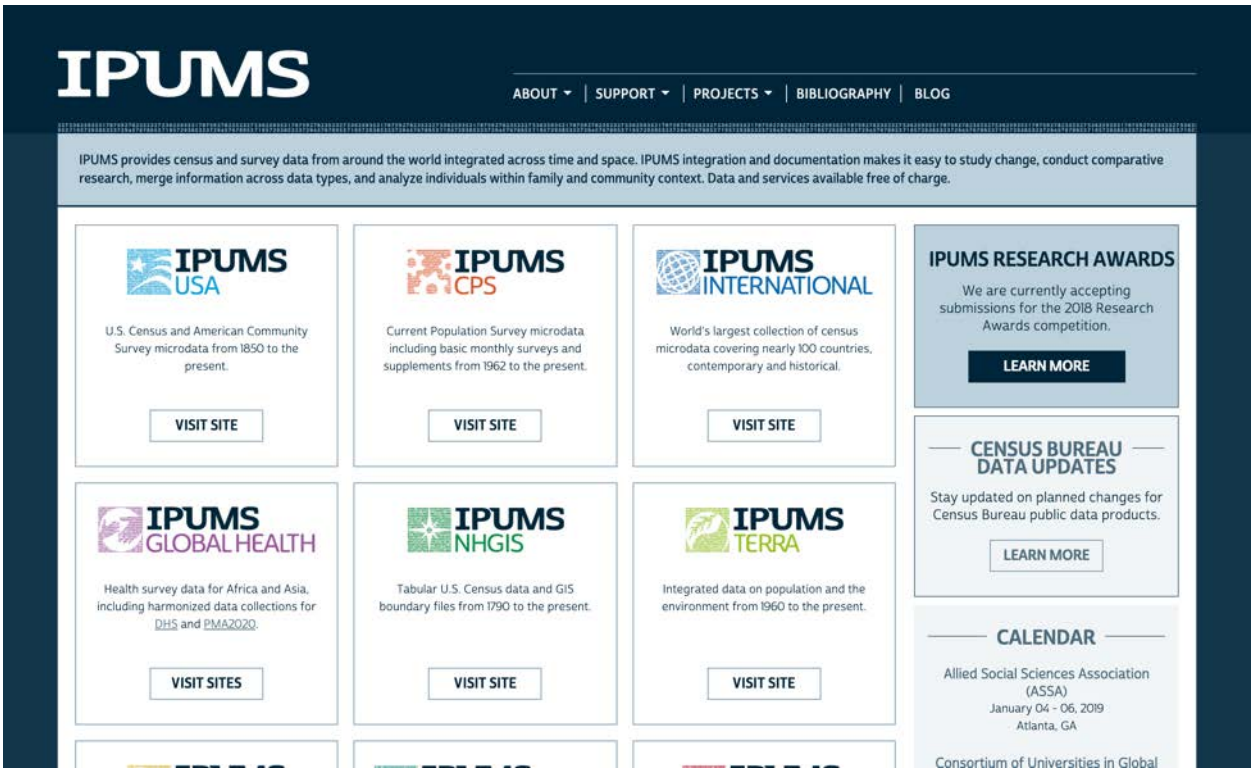
"I need a simple, unified data portal from which to access all of AHRQ's data tools in order to seamlessly integrate multiple data sources in my research workflow."

Almost all users interviewed start with Google as their primary portal to AHRQ's data tools and summary data tables. It's ideal for this purpose, providing a clean, simple interface oriented around the ability to perform simple queries across multiple data sources. There exists an opportunity to emulate this simple interface in an AHRQ data tools portal (which is further discussed in the next user story).

It's possible to create a simple UI for this portal while addressing all three user types. Pro and Intermediate Users, being familiar with their tool of choice, will often seek out the name and brand of their preferred tool. Meanwhile, Novice Users will seek out high-level descriptions of the intent and application of each of the tools (e.g., "click here to know more about what people pay for their Healthcare" for MEPS (HC), and "click here for information on hospitals" for HCUPnet). It's important to take these needs for each user type into consideration when designing a portal that appeals to all of them.

The user requirements in this and the following sections will touch upon various features that might be included in such a portal, allowing users representing varied backgrounds, skill levels, and use cases to centrally and effectively engage with all of AHRQ's web-based data tools and summary data tables.

Agency Comparison: Two unique pathways for creating a unified portal of multiple data tools that AHRQ could consider are the data.HRSA.gov website of the Health Resources & Services Administration (HRSA) and the Integrated Public Use Microdata Series (IPUMS). The data.HRSA.gov has organized its portal homepage into, consecutively: a simple call to action, a list of color-coded tools categorized by function and phrased as actions (e.g., "explore maps," "query data," "download data"), data snapshots, an interactive map, a list of initiatives accompanied by icons for each, and a high-level "about" and miscellaneous information section. While much can be learned by emulating this user-centered interface, the function-and-topic-oriented nature of this interface, however, isn't entirely applicable to AHRQ's various and siloed tools, whose delineations aren't as clear-cut and whose individual brands need to be retained. The IPUMS interface, on the other hand, is unique in that it retains the individuality of each of the websites under its mantle: upon visiting the IPUMS homepage, you see a grid of individual websites, each with its own brand that's nonetheless harmonized with the others and featuring a one-sentence description. The individuality of each is retained, while branding is harmonized and feels uniform.


IPUMS
ABOUT | SUPPORT | PROJECTS | BIBLIOGRAPHY | BLOG
IPUMS provides census and survey data from around the world integrated across time and space. IPUMS integration and documentation makes it easy to study change, conduct comparative research, merge information across data types, and analyze individuals within family and community context. Data and services available free of charge.
IPUMS USA
U.S. Census and American Community Survey microdata from 1850 to the present.
VISIT SITE
IPUMS CPS
Current Population Survey microdata including basic monthly surveys and supplements from 1962 to the present.
VISIT SITE
IPUMS INTERNATIONAL
World's largest collection of census microdata covering nearly 100 countries, contemporary and historical.
VISIT SITE
IPUMS RESEARCH AWARDS
We are currently accepting submissions for the 2018 Research Awards competition.
LEARN MORE
IPUMS GLOBAL HEALTH
Health survey data for Africa and Asia, including harmonized data collections for DHS and PMA2020.
VISIT SITES
IPUMS NHGIS
Tabular U.S. Census data and GIS boundary files from 1790 to the present.
VISIT SITE
IPUMS TERRA
Integrated data on population and the environment from 1960 to the present.
VISIT SITE
CENSUS BUREAU DATA UPDATES
Stay updated on planned changes for Census Bureau public data products.
LEARN MORE
CALENDAR
Allied Social Sciences Association (ASSA)
January 04 - 06, 2019
Atlanta, GA
Consortium of Universities in Global

"I need a simple, predictive search engine in order to reference everything from estimates to survey questions."

As mentioned in the previous requirement, Google was the de-facto portal for nearly every user interviewed. Most attempts at discovering the data tools through AHRQ's website ended in frustration by users. Even Pro Users are no exception to this trend, despite their savviness in using the products. Moreover, Pro Users often exhibit best practices in wielding search engines to discover relevant information, in many cases having developed work-arounds for finding relevant information, while AHRQ-related search engine results lag in optimization for providing such results.

Improved search functionality emerged as one of the leading features across all interviews. Some users reported being confused by the results of their queries in AHRQ's main website and the data tools themselves, with incongruous results lacking customization options and sorted by relevance (when most Pro Users sought to sort by date for the latest data).

There are a few options for implementing a search functionality as a central design feature of a harmonized UI across the data tools and summary data tables, some more geared toward the needs of Pro Users and others toward Novice and Quick Reference users. Universally, however, is the ability to perform a search of a topic area (e.g., inpatient hospitalization) and receive a list of all the data tools that touch upon it. More specifically to Pro Users is the improved capability to refine searches – including by date and document type – in order to support their more granular research needs.

It's important to note that in addition or even alternative to the implementation of an AHRQ-specific predictive search function, it would also behoove all user types for all tools to be search engine optimized.[2]

"I need to view what tools are available by topic area in order to determine which datasets can answer my research question."

Like other external users, Pro Users will often approach AHRQ's data tools not necessarily having a particular tool in mind, but with a research question that needs answering. As such, the ability to either search by topics or browse a list that would yield applicable data tools would be a boon to all 3 user groups. This feature can be incorporated into both a unified portal for all of AHRQ's tools as well as in the homepage of each tool.

[2] Additional user stories related to searching:
"I need to view what tools are available by topic area in order to determine which datasets can answer my research question."
"I need a customizable search engine that enables me to sort/refine results – including by topic, data tool, date, and document type – in order to find potential answers to my research questions across multiple data tools and summary data tables"
"I need to improved searching of the query tools' data dictionaries in order to perform granular searches related to my research question."
"I need the data tools I use to be optimized for popular search engines so I can continue to engage with the datasets via my search engine of choice."
"I need simple, intuitive URLs to access the data tools collectively or individually in order to enable quick navigation."

"I need to be able to quickly determine the ease of access to data files for a particular data source in order to determine the level of effort (LOE) and feasibility of its use for my research."

Along with the usability principle of versatility, when Pro Users are determining which data resources to use, ease of access is a primary concern. While some users of this type (such as Senior Researchers) will have the financial and/or staffing resources to either purchase data or to engage in the often-arduous process of acquiring micro-data, many do not. These lesser-resourced Pro Users, such as graduate students, will gravitate toward rich, public-use data files such as MEPS-HC (becoming a gold standard for "starter" research projects), and otherwise need to understand both the financial investment and level of effort involved in engaging with a data tool. Having access to these public-facing datasets help tremendously throughout the research process, from conducting initial analysis to understand whether the dataset can provide an answer to their research question, all the way through the preliminary analysis and beyond. As such, this user story suggests prioritizing the display of high-level information on dataset access restrictions (specifically, whether a dataset is public-use or at-cost); the display of this information can be integrated in both the unified portal for these data tools as well as in the homepage of a data tool itself. Additionally, access points and, when applicable, any instructions for accessing data files should be prominent on any data tool homepage.

"I need to quickly assess the characteristics of a data source – particularly if it provides nationally-representative, person-level data – in order to determine the LOE of, and applicability of that data, to my research."

Some Pro Users noted their preferences for certain datasets over others. Within the scope of our research, PhD students will, for instance, tend to seek out nationally representative and person level data, leading them away from resources like CAHPS and NHQDRnet in favor of MEPS HC and IC. In addition to understanding ease of access of a dataset, it's also important to specify other characteristics and limitations of that dataset to help Pro Users easily determine whether it will be applicable to their research.

4.3.1.3. Understanding Methods & Variables

"I need to be able to quickly reference the linkage files, codebooks, and documentation files in any tool in order to conduct my research efficiently."

In addition to observing use of an AHRQ's data tool, the Atlas Team also observed Pro Users engaging with other AHRQ data tools with which they were not previously familiar. In these instances, it became apparent that in addition to seeking out data files, Pro Users will often seek out linkage files, codebooks, and documentation files when first engaging with a new data tool. Essential for first time visitors and for continued reference when consulting with a dataset, Pro Users will regularly consult with codebooks and documentation files to gain a granular understanding of how data was gathered. As such, it's important to have clear links to these resources on the homepage of any data tool, as well, perhaps, as featured alongside query results for quick, repeated reference.

"I need to be able to conduct customized searches of survey questionnaires and data dictionaries in order to fully understand the methods and variables behind a dataset."

Currently, Pro Users seeking to better understand the methods underlying certain datasets must painstakingly parse through multiple PDFs, each containing a different section of the larger survey; if, for instance, they'd like to seek a related question in the same year but a different topic area, they need to return to a previous page, conduct a new search, open the PDF and search it for the keywords they seek. Searchable formats of certain data – like reports and survey questionnaires – is the optimal solution, with an alternative being the option to download unified PDF versions of such documents.

4.3.1.4. Accessing Data Files

As it pertains to the data tools and summary data tables in question, gaining data file access is often – with the exception of using GitHub – a discrete, one-off activity that occurs after the latest data is released (after which the files most likely reside on the user's local computer). However, because Pro Users often forgo the use of the query tools in favor of running the data themselves, data access is nevertheless a central use case for this user type, and any insights are important to consider for the development of both a harmonized UI for the tools as well as a proposed COTS product.

"I need to be able to download data files in my preferred programming language in order to conduct analysis in the statistical analysis software of my choice."

Pro Users will have their own preferred statistical analysis software depending on their research background – from a business school graduate student using SAS to someone performing more traditional statistical analysis using SAS or Stata. As such, the ability to download raw data files (as well as query results) in the programming language of their choice is a crucial feature for Pro Users.

"I need up-to-date datasets hosted on GitHub in order to integrate the use of this tool with the rest of my GitHub workflow."

GitHub has emerged among a group of Pro Users – particularly graduate students – as an immensely useful tool for engaging with datasets for the purposes of analyses and understanding methods. This functionality has emerged as a best practice among those Pro Users who have engaged with MEPS Summary Tables (Household Component). Some users have even suggested the addition of a secure UI for GitHub in harmony with AHRQ's other data tools.

Addendum: *This requirement was deprioritized during the Mid-Phase Workshop by AHRQ staff due to its feasibility. Based on feedback from the project COR, the updated requirement – listed in the consolidated user stories below – is described as follows: "I need access to code on GitHub that will allow me to access up-to-date public-use files that live on the AHRQ website".*

"I need clear guidelines for microdata access and data use agreement procedures – and easy access to technical support – in order to engage with certain datasets."

While tools like HCUP have made some improvements to their data use agreement procedures, it remains an arduous and confusing process, even for Pro Users. In tandem with the other streamlined data access features suggested above and later in this section, clear guidelines that exhibit best practices in content design – with an ever-present option to submit a technical support request – are a must-have to provide, as seamlessly as possible, a data access portal for each of the data tools.

4.3.1.5. Formulating & Refining a Data Tool Query

"I need to be able to sort/refine query results – e.g., by condition and topic – in order to quickly answer my specific research question."

Due to Pro Users' reliance on running their own data, conducting queries in data tools or summary data tables is a rarity. It does, however, have some utility: this often takes the place of quickly referencing estimates for the purpose of verification and reference. In these cases, Pro Users' research query is often highly specific, necessitating a heightened ability to refine the results. And since the nature of this use case is often impromptu, a mobile-ready and aesthetically clean interface would best suit the Pro Users' needs.

"I need a one-page data query interface in order to efficiently create and edit data queries."

One of the most significant usability challenges for all users of the data query tools are the issues that arise from multi-page query interfaces. A best practice that arose was the use of single-screen interfaces, which allow for refinement or reformulating of data queries in a single screen.

4.3.1.6. Reviewing the Data

"I need the relationships between CCS, ICD-9, and ICD-10 codes to be designated whenever applicable in order to understand gaps, discrepancies, or shifts in the data."

For Pro Users, the transition between Clinical Classifications Software (CCS), International Classification of Diseases, 9th Revision (ICD-9), and International Classification of Diseases, 10th Revision (ICD-10) codes can lend some confusion to data results. Of AHRQ's data query tools, Fast Stats has provided a resolution to this issue by displaying the transition visually in graphs of data trends; others merely show an inexplicable dip in the data count at the time of transition between these codes. While Fast Stats provides a clever visual solution, there may be other opportunities to effectively signify the relationship between these diagnostic codes in a user-friendly fashion.

"I need visual signifiers of data – including graphs or other data visualizations – whenever applicable in order to make the review of complex data more efficient."

The Atlas Team observed the use of data visualizations, such as graphs, as signifiers for more complex information; this process typically entails first referring to a chart or graph to gain contextual understanding, followed by the review of tabulated data for more granular understanding. The use of visualization, while generally a usability best practice, also enables another user requirement: the comparison of data (e.g., mean and median expenditures per person) side-by-side.[3]

Agency Comparison: The Data Visualization Gallery of the National Center for Health Statistics (NCHS) is an excellent example of this use of visuals to quickly reference complex data. Their homepage simply features a grid of sample data visualizations each accompanied by the headline of the data it represents. The potential applications of such an interface are numerous, from featuring sample data visualizations from each tool on the home page as a preview of the capabilities of each, to offering a similar gallery to indicate different "view options" one can apply to the data in any given query tool using the embedded COTS product.

4.3.2. Intermediate Users

Intermediate Users are mainly distinguished by a limited fluency with manipulating and conducting analyses of raw data; thereby, they are the most heavily reliant on the data query tools out of all the External Users. The sophistication of their research skills and their level of effort is still relatively variable, and they will submit technical assistance requests to help resolve more complicated queries or to request special runs of data. This user type includes, but is not limited to, federal agencies, state government agencies, healthcare providers, advocacy groups, policymakers, and some researchers.

4.3.2.1. Knowing Where to Start My Query

"I need a query portal that allows me to engage by either clicking on prescribed topic areas or by filling a customizable search/query field in order to conduct a query of my choice."

Most users enter the tools in question with a specific research question or topic in mind; offering options for both allows for a versatility of engagement, while also allowing users who have a topic in mind that falls somewhere in between what's provided (e.g., on MEPS Summary Tables, if a query has to do with both population *and* prescribed drugs) to use the search field instead.

[3] See also:
"I need to be able to compare two queries side-by-side (e.g., mean and median expenditures per person) in order to gain an understanding of comparable variables for my research question."
"I need the ability to select from a variety of different data "view" formats – including different data visualizations – in order to effectively analysis query results."

4.3.2.2. Formulating & Refining a Data Tool Query

"I need access to an easy-to-use query tool that can quickly pull from a broad dataset in order to perform complex queries without having to manipulate the data files themselves."

The versatility and speed of the summary data tables such as those for MEPS (HC) are a useful tool for this particular group. However, as infrequent users of raw data files, intermediate users can find themselves hamstrung by the limited datasets of the summary tables. As such, a powerful product that can either pull from a larger sample of data or that is able to perform live calculations of full datasets with the speed of summary tables would be a great boon to Intermediate Users who may not have the necessary skill of Pro Users but have more complicated queries than Novice / Quick-Reference Users.

4.3.2.3. Accessing Data Files

"I need to be able to view a consolidated data release schedule in order to understand when to update my data files."

For those users who are operating off of some localized version of the data or who submit annual or semi-annual requests for special runs of the data, the ability to understand when they can expect the latest data to be released would be a useful feature on either a consolidated portal for all the tools or for each of the tools individually.

"I need to be able to understand alternate options to access the data in the event that the query tool I'm using doesn't yield the results I need in order to complete my research query."

While the selection of an optimal COTS product would presumably address the issue of inaccessible portions of a dataset, it's nevertheless important to remind the user during the query process that other options are available. An example is the observed behavior of MEPS (HC) users attempting to answer a query by first referring to the summary tables (e.g., to quickly reference a single-year estimate), then conducting the same query in MEPSnet (HC) to see a trend graph over time, and, in one case, then referencing the raw data files for more targeted analysis. For instance, beside the results of a query, such a feature could appear as a callout box offering alternate data query options including a link to the raw data files or alternate tools if applicable.

4.3.2.4. Reviewing the Data

"I need any footnotes or appendices that might accompany query results to be easily scannable in order to quickly grasp analyses."

Some AHRQ staff frequently mentioned that despite the significant effort that is put into documentation, users simply end up not reading it (though it's still requested in support queries / technical assistance requests). In some cases, this is because content is otherwise so dense and lacking a clear hierarchy that it's merely glazed over. This could be resolved by providing collapsible footnote text fields that might include descriptive analyses, sources, suggested citations, exclusion criteria, or potential limitations of the data under the data itself.

"I need to be able to return to a query at a later time in order to complete my research unabated."

For more complicated research questions, users would benefit from the ability to save a query and return to it at a later time. Some users interviewed requested a dashboard on which search history could be stored; however, given the unlikelihood of storing of private data on any of the sites, a good workaround would be the ability to share or save the link to a query for future reference.

"I need to be able to download query results as a printable 1-page PDF in order to distribute salient data with others or reference it in the future."

Some users requested the ability to download or print a succinct summary of query results. Specifically, such a PDF file could include a high-level, visual synopsis at top (e.g., line graph) and a table below to include both the data and standard errors.

4.3.2.5. Completing My Query Task

"I need to be able to copy and share the link to a query in order to collaborate and share with others."

For both Intermediate Users, and even other users who provide customer support, the ability to share the link to a query is a great boon for collaboration or customer support.

"I need to be able to download query results in multiple formats in order to collaborate and share with others."

In other cases, users need the ability to download query results in Excel, PDF, or PowerPoint format.

4.3.3. Novice & Quick-Reference Users

The foremost distinguishing characteristic of Novice and Quick-Reference Users is their reliance on AHRQ technical support requests to help find relevant data. While in many cases this reliance is due to the poor usability of some of AHRQ's data tools and summary data tables, it can also be due to time constraints or the need for custom queries of data that can't be completed in the current iteration of the tool. Nevertheless, the nature of the technical assistance and general support requests submitted by this group provide valuable insight into easy fixes to enhance the tools' usability. This user type includes, but is not limited to, the media, undergraduate students, economists, congressional staffers and some federal agencies, insurance companies, undergraduate students, and the general public.

4.3.3.1. Navigating the Data Tools with Ease

"I need harmonized UIs that have consistent look, feel, and terminology in order to seamlessly transition between AHRQ's various data tools."

While already integral to the purpose of this Environmental Scan, it's nevertheless important to reiterate the utility of this requirement for improving the experience of Novice users in particular. AHRQ staff who field technical assistance requests from this group report that many are submitted merely due to usability issues. By establishing some degree of consistency of both content and functionality, the harmonization of both branding and UI can significantly lessen the cognitive burden for these users navigating a series of seemingly disparate data tools.

4.3.3.2. Knowing Where to Start My Query

"I need the data query tools or summary tables and their function to be prominently featured in the UI in order to efficiently answer my research question."

Many technical assistance requests result from users having difficulty locating the query tool. The logical response is to ensure that the tools are centrally featured, and their functionality clearly designated in the interface.

4.3.3.3. Seeking Support

"I need to be able to easily submit a technical assistance request or a request for a special run of data in order to resolve my research question when my query results aren't sufficient."

According to some AHRQ staffers who respond to technical assistance requests, roughly half of the requests received are unique or involved. The ultimate goal is to create a user-friendly query interface / COTS product that would resolve technical assistance requests due to usability issues. In the event, however, that an Intermediate or Novice user still needs assistance with their query, that option should be readily available in the query tool.

4.4. Internal Users

While not representative of specific roles of AHRQ (internal) staff, these user groups represent two categories of use cases for these staff as it relates to AHRQ's data tools and summary data tables. Broadly, these users are responsible for developing, collecting, maintaining, and quality control of data, as well as making that data available to outside users. As it pertains to their specific use cases for the data tools and summary data tables, their activities include the support of tool users, administration of the tools, and ad hoc use of the tools for research and QC purposes. The resulting user types and corresponding themes of activity as it relates to the tools are as follows:

- **Support**
    - Navigating the Data Tools with Ease

- Receiving & Resolving User Queries
- General Support

- **Administrator**
  - Data Tool Management
  - Quality Control

This user group includes (but is not limited to) statisticians, economists, researchers / research scientists, research division directors, project managers, writer-editors, media and policy relations staff, and database operations contractors.

4.4.1. Support

Support users include (but are not limited to) statisticians, economists, researchers, media and policy relations staff, and database contractors who help field Technical Assistance Requests, special runs, and general support and outreach as it relates to the data query tools and summary data tables. As a result, they have important insight into the needs and challenges of External Users, much of which has been incorporated into that section of this document. For the purpose of providing customer support, this user type is most similar to Intermediate Users in their reliance on and fluency with the query tools for quick reference in response to these requests.

4.4.1.1. Navigating the Data Tools with Ease

"I need the website and its core functionality to be optimized for all browser types in order to include its regular usage in my normal workflow."

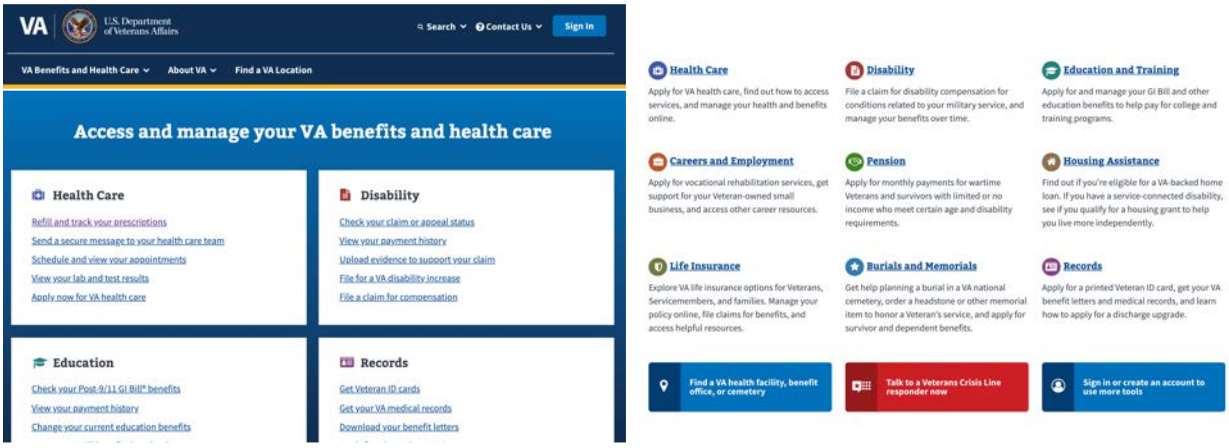
As they respond to multiple users on different types of browsers, support currently needs to understand the variations in user experience across browsers. The functionality of the tool or site can be impeded as a whole – for instance, in IE is notoriously slow – or certain functionality is variable – for instance, as described earlier, certain mobile browsers are difficult to maneuver, particularly for downloading excel files or plots.

"I need simple, easy-to-understand terminology and descriptions of the tools' utility in order to quickly grasp which tool can be used to answer my research question."

Currently, some Support Users responsible for fielding questions from Novice Users spend significant effort describing each tool and its utility in laymen's terms. This correspondence takes the form of individual emails with customers, blog posts, and other forms of outreach. In considering harmonized branding, this requirement reinforces the need to create streamlined content and a visual hierarchy prioritized by user type: high-level descriptions should be easy to understand for Novice Users, while more granular, technical text and descriptions should be deprioritized and streamlined but still readily apparent to Intermediate and Pro Users.

Agency Comparison: Based on extensive research with Veterans, the Department of Veterans Affairs (VA) consolidated multiple Veteran-facing tools into one portal, VA.gov. The homepage for this portal features a user-centered information hierarchy for its 9 consolidated, color-coded topic areas for Veteran services: after a simple 10-word call to action, it is followed by two portal

sections, the first featuring the most frequently-used topics and subpages – each with a 1-2 word headline and 5 subpage links at 5 words each – followed by the full list of topics in the section below. No single block of text exceeds 30 words.



4.4.1.2. Receiving & Resolving User Queries

According to some Support users we interviewed (and as stated previously in the Novice User section), the amount of Technical Assistance Requests that are due to usability issues and those that are regarding complex queries are roughly equivalent. As a result, the comprehensive skillsets of Support users are being underutilized; if basic usability issues were resolved, those skills could be put to better use in resolving more complex queries. Here are a few requirements that both satisfy the needs of External Users and the internal Support Users assisting them.

"I need to easily modify my query on one page – switching between variable groups, variable selection, and potential analyses – in order to answer a research question efficiently."

As stated previously, users make frequent mistakes in data query tool with multi-page formats. An ideal solution would provide a simple query portal, followed by a single-page query tool that enables rapid customization while using effective visual hierarchy to ensure that Novice Users are not overwhelmed by a barrage of options. This would significantly benefit Support Users in two ways: by reducing the margin of error for conducting queries on behalf of users it would reduce time to resolve those queries, and by reducing the margin of error for External Users, Support staff would receive more substantive, complex queries as a result, making better use of their expertise and thereby providing more value to the customer.

4.4.1.3. General Support

"I need user queries to be easily and efficiently routed through the UI to the right Support Users in order to provide substantive customer support."

While the routing of customer queries at AHRQ was not within the scope of this project, there exist strategic opportunities in External Users' workflow within the tools to route customer support requests to the right people. This can take the form of a prominent customer support option. We suggest that further research be conducted into the customer query process –

including the most efficient forms of communication to Support Users – in order to provide value to External Users and efficiency and ease for Internal Users.

4.4.2. Administrator

This user type is characterized by all use cases of the data tools that are not directly related to customer support. Stakeholders included in this user type include (but are not limited to): statisticians, researchers / research scientists, research division directors, project managers, writer-editors, and database operations contractors.

Their use of the data tools is most similar to Pro Users in that they largely forgo the tools and summary tables in favor of running their own data. However, the themes of user behavior as it pertains to the data tools and summary data tables encompass general usability needs, the management of the databases and corresponding tools, ad hoc or miscellaneous usage of the tools, and quality control of content including stat briefs, reports, and data tool content.

4.4.2.1. Data Tool Management

“I need backend management of the datasets to be user-friendly for both contractors and administrators in order to ensure that quality control efforts are carried out seamlessly every time the datasets are updated.”

Extensive backend quality control efforts go into the updating of the web-based data tables themselves every time the datasets are updated. While this aspect of the data tools is on the margins of the scope of this project, we highly recommend further research go into this process to ensure that a harmonized user interface and COTS product keep this use case in mind.

4.4.2.2. Quality Control

While the query tools are by no means central to Quality Control activities, Administrator Users will use them to quickly validate estimates in reports and statistical briefs before publishing them. Additionally, extensive backend quality control efforts go into the updating of the web-based data tables themselves every time the datasets are updated.

“I need the query tool to be able to conduct statistical tests, so I can easily determine what's statistically significant for my research question.”

When retrieving quick estimates for the purpose of quality control, Administrator Users will use the data query tools for quality control to find quick estimates

4.5. Consolidated User Stories

The following tables contain the consolidated user requirements organized by theme and cross-referenced with the user type each requirement satisfies. The abbreviations for each user type are as follows:

- **External Users:**
    - **Pro User:** "Pro"
    - **Intermediate User:** "Int"
    - **Novice / Quick-Reference User:** "Nov"
- **Internal Users:**
    - **Customer Support:** "Sup"
    - **General Administrator:** "Adm"

Each requirement is written in Agile-compliant "user story" format for simplicity and ease of communication with AHRQ IT and COTS product representatives during the prototyping phase of this project.

4.5.1. Navigating the Data Tools with Ease

User Story	Pro	Int	Nov	Sup	Adm
"I need to be able to access basic query functionality on both my laptop and my mobile device in order to spontaneously reference information."	X	X	X	X	X
"I need a user-friendly interface with highly streamlined & visualized content – preferably that can be viewed with minimal scrolling and scanning of extraneous text – in order to reduce the cognitive burden of browsing complex information."	X	X	X		
"I need access to the archived version of original site in order to reference where I know information was while transitioning to a new interface."	X	X			
"I need a similar site map and terminology to the original site in order to continue my use of the site relatively unabated."	X	X			
"I need harmonized UIs that have consistent look, feel, and terminology in order to seamlessly transition between AHRQ's various data tools."		X	X		
"I need a website that's optimized for all browser types in order to include its regular usage in my normal workflow."	X	X	X	X	X

"I need simple, easy-to-understand terminology and descriptions of the tools' utility in order to quickly grasp which tool can be used to answer my research question."			X	X	

4.5.2. Knowing Where to Start My Query

User Story	Pro	Int	Nov	Sup	Adm
"I need a unified data portal from which to access all of AHRQ's data tools in order to seamlessly integrate multiple data sources in my research workflow."	X	X	X		
"I need to view or search what tools are available by topic area in order to determine which datasets can answer my research question."	X	X	X		
"I need a customizable search engine that enables me to sort/refine results – including by topic, data tool, date, and document type – in order to find potential answers to my research questions across multiple data tools and summary data tables."	X	X	X		
"I need the data tools I use to be optimized for popular search engines, so I can continue to engage with the datasets via my search engine of choice."	X	X	X		
"I need simple, intuitive URLs to access the data tools as a whole or individually in order to quickly navigate to."	X	X			
"I need a query portal that allows me to engage by either clicking on prescribed topic areas or by filling a customizable search/query field in order to conduct a query of my choice."		X	X		
"I need the data query tools or summary tables and their function to be prominently featured in the UI in order to efficiently answer my research question."		X	X		
"I need a simple, predictive search engine in order to find everything from estimates to survey questions."	X	X	X		
I need all branding and sub-branding of my tool to remain prominent in a user interface in order to retain the well-established reputation and userbase of my dataset.		X	X		X

4.5.3. Understanding Methods & Variables

User Story	Pro	Int	Nov	Sup	Adm
"I need to be able to quickly reference the linkage files, codebooks, and documentation files in any tool in order to conduct my research efficiently."	X				
"I need to be able to conduct customized searches of survey questionnaires and data dictionaries in order to fully understand the methods and variables behind a dataset."	X	X			
"I need the relationships between CCS, ICD-9, and ICD-10 codes to be designated whenever applicable in order to understand gaps, discrepancies, or shifts in the data."	X	X			

4.5.4. Accessing Data Files

User Story	Pro	Int	Nov	Sup	Adm
"I need to be able to download data files in my preferred programming language in order to conduct analysis in the statistical analysis software of my choice."	X				
"I need access to code on GitHub that will allow me to access up-to-date public-use files that live on the AHRQ website."	X				
"I need clear guidelines for microdata access and data use agreement procedures – and easy access to technical support – in order to engage with certain datasets."	X				
"I need to be able to view a consolidated data release schedule in order to understand when to update my data files."	X	X	X		
"I need to quickly assess the characteristics of a data source – particularly if it provides nationally-representative, person-level data – in order to determine the LOE of, and applicability of that data, to my research."	X	X			
"I need to be able to quickly determine the ease of access to data files for a particular data source in order to determine the LOE and feasibility of its use for my research."	X				

4.5.5. Formulating & Refining a Data Tool Query

User Story	Pro	Int	Nov	Sup	Adm
"I need to be able to sort/refine query results – e.g., by condition and topic – in order to quickly answer my specific research question."	X	X	X		
"I need a one-page data query interface in order to efficiently create and edit data queries."	X	X	X		
"I need to have greater customization of age groups and date ranges in my queries in order to achieve the granular results I require."	X	X	X		
"I need access to an easy-to-use query tool that can quickly pull from a broad dataset in order to perform complex queries without having to manipulate the data files themselves."		X	X		
"I need to be able to understand alternate options to access the data in the event that the query tool I'm using doesn't yield the results I need in order to complete my research query."	X	X	X		
"I need to easily modify my query on one page – switching between variable groups, variable selection, and potential analyses – in order to answer a research question efficiently."	X	X	X	X	X

4.5.6. Reviewing the Data

User Story	Pro	Int	Nov	Sup	Adm
"I need visual signifiers of data – including graphs or other data visualizations – whenever applicable in order to make the review of complex data more efficient."	X	X	X		
"I need to be able to compare two queries side-by-side (e.g., mean and median expenditures per person) in order to gain an understanding of comparable variables for my research question."	X	X			
"I need the ability to select from a variety of different "view" formats – including different data visualizations – in order to effectively analysis query results."	X	X	X		

"I need access to be able to select diagnostic and procedural variables in all the data tools in which its available."		X	X		
"I need any footnotes or appendices that might accompany query results to be easily scannable in order to quickly grasp analyses."	X	X	X		
"I need to be able to return to a query at a later time in order to complete my research unabated."		X	X		
"I need to be able to download query results as a printable 1-page PDF in order to distribute salient data with others or reference it in the future."		X	X		
"I need descriptions for all variable labels – including examples – in order to understand the results of my query."		X	X		

4.5.7. Seeking Support

User Story	Pro	Int	Nov	Sup	Adm
"I need to be able to easily submit a technical assistance request or a request for a special run of data in order to resolve my research question when my query results aren't sufficient."		X	X		
"I need user queries to be easily and efficiently routed through the UI to the right Support Users in order to provide substantive customer support."				X	

4.5.8. Completing My Query Task

User Story	Pro	Int	Nov	Sup	Adm
"I need to be able to copy and share the link to a query in order to collaborate and share with others."		X	X		
"I need to be able to download query results in multiple formats in order to collaborate and share with others."		X	X		

4.5.9. Data Tool Management

User Story	Pro	Int	Nov	Sup	Adm
"I need backend management of the datasets to be user-friendly for both contractors and administrators in order to ensure that quality control efforts are carried out seamlessly every time the datasets are updated."					X

4.5.10. Quality Control

User Story	Pro	Int	Nov	Sup	Adm
"I need the query tool to be able to conduct statistical tests, so I can easily determine what's statistically significant for my research question."		X	X	X	X

4.5.11. ADDENDUM: Mid-Phase Workshop Requirements Prioritization

During the January 16th Mid-Phase Workshop, POCs from each of the data tools and summary data tables voted on the accumulated requirements according to their desirability for their specific tools. Below are those requirements that were most voted-upon, listed by the data tool whose representative cast that vote (please note that votes marked "General" were general administrative AHRQ staff without clear designation for a particular data tool); for each vote, participants provided a "priority" designation, with Priority #1 being most important and #3 being of lowest priority.

Any requirements that are missing from this list do not indicate that they are of no import to data tools POCs or users; some lack votes merely because their utility is considered a given. Additionally, some requirements were added during the workshop; these new requirements are in bold. You can find more participant feedback from the workshop in the Conclusion section (Section 5) and Appendix C.

User Story	#1	#2	#3	Total Votes
"I need a website that's optimized for all browser types in order to include its regular usage in my normal workflow."	MEPS HC MEPS IC Fast Stats General	MEPS HC CAHPS		6

User Story	#1	#2	#3	Total Votes
"I need the ability to select from a variety of different "view" formats – including different data visualizations – in order to effectively analysis query results."	MEPS IC Fast Stats	MEPS IC General	MEPS HC	5
"I need the query tool to be able to conduct statistical tests, so I can easily determine what's statistically significant for my research question."		MEPS HC MEPS IC	MEPS HC MEPS IC	4
"I need to be able to access basic query functionality on both my laptop and my mobile device in order to spontaneously reference information."		MEPS HC	MEPS IC Fast Stats CAHPS	4
"I need the data tools I use to be optimized for popular search engines, so I can continue to engage with the datasets via my search engine of choice."	MEPS IC		MEPS IC MEPS HC	3
I need a cover page that displays both program descriptions and available data in order to properly demonstrate the utility of my tool to users.	**CAHPS**	**General MEPS IC**		**3**
"I need access to an easy-to-use query tool that can quickly pull from a broad dataset in order to perform complex queries without having to manipulate the data files themselves."	CAHPS	MEPS IC (2)		3
"I need to easily modify my query on one page – switching between variable groups, variable selection, and potential analyses – in order to answer a research question efficiently."	MEPS IC	MEPS IC	MEPS IC	3
"I need backend management of the datasets to be user-friendly for both contractors and administrators in order to ensure that quality control efforts are carried out seamlessly every time the datasets are updated."	MEPS HC	CAHPS Fast Stats		3

User Story	#1	#2	#3	Total Votes
"I need a unified data portal from which to access all of AHRQ's data tools in order to seamlessly integrate multiple data sources in my research workflow."	General	MEPS IC		2
"I need a user-friendly interface with highly streamlined & visualized content – preferably that can be viewed with minimal scrolling and scanning of extraneous text – in order to reduce the cognitive burden of browsing complex information."	MEPS HC		QDR	2
"I need a query portal that allows me to engage by either clicking on prescribed topic areas or by filling a customizable search/query field in order to conduct a query of my choice."	MEPS IC		CAHPS	2
"I need a simple, predictive search engine in order to find everything from estimates to survey questions."	Fast Stats	MEPS HC		2
"I need to be able to quickly reference the linkage files, codebooks, and documentation files in any tool in order to conduct my research efficiently."		MEPS HC	HCUPnet	2
"I need to quickly assess the characteristics of a data source – particularly if it provides nationally-representative, person-level data – in order to determine the LOE of, and applicability of that data, to my research."			MEPS HC HCUPnet	2
"I need to be able to quickly determine the ease of access to data files for a particular data source in order to determine the LOE and feasibility of its use for my research."	MEPS HC	HCUPnet		2
"I need to be able to compare two queries side-by-side (e.g., mean and median expenditures per person) in order to gain an understanding of comparable variables for my research question."	MEPS HC (2)			2

User Story	#1	#2	#3	Total Votes
“I need any footnotes or appendices that might accompany query results to be easily scannable in order to quickly grasp analyses.”		Fast Stats	Fast Stats	2
“I need to be able to download query results as a printable 1-page PDF in order to distribute salient data with others or reference it in the future.”	MEPS IC	Fast Stats		2
"I need the option to view map visualizations for any relevant data in order to illustrate the richness of my dataset to users."			**MEPS HC Fast Stats**	**2**
“I need to be able to download query results in multiple formats in order to collaborate and share with others.”		MEPS IC	QDR	2
“I need access to the archived version of original site in order to reference where I know information was while transitioning to a new interface.”			MEPS IC	1
“I need harmonized UIs that have consistent look, feel, and terminology in order to seamlessly transition between AHRQ’s various data tools.”			MEPS IC	1
“I need simple, easy-to-understand terminology and descriptions of the tools’ utility in order to quickly grasp which tool can be used to answer my research question.”		Fast Stats		1
“I need a customizable search engine that enables me to sort/refine results – including by topic, data tool, date, and document type – in order to find potential answers to my research questions across multiple data tools and summary data tables.”	QDR			1
“I need to be able to conduct customized searches of survey questionnaires and data dictionaries in order to fully understand the methods and variables behind a dataset.”			MEPS HC	1

User Story	#1	#2	#3	Total Votes
"I need the relationships between CCS, ICD-9, and ICD-10 codes to be designated whenever applicable in order to understand gaps, discrepancies, or shifts in the data."		Fast Stats		1
"I need to be able to download raw data files in order to conduct analysis in the statistical analysis software of my choice."			MEPS IC	1
"I need clear guidelines for microdata access and data use agreement procedures – and easy access to technical support – in order to engage with certain datasets."	HCUPnet			1
"I need to be able to sort/refine query results – e.g., by condition and topic – in order to quickly answer my specific research question."	HCUPnet			1
"I need a one-page data query interface in order to efficiently create and edit data queries."	MEPS IC			1
"I need visual signifiers of data – including graphs or other data visualizations – whenever applicable in order to make the review of complex data more efficient."		HCUPnet		1
"I need to be able to return to a query at a later time in order to complete my research unabated."		MEPS HC		1
"I need descriptions for all variable labels – including examples – in order to understand the results of my query."		MEPS HC		1

4.6. Card Sort Analysis

Card sorts are useful tools for understanding users' mental models of how information is categorized and prioritized. The activity enables users to categorize content and functionality – in this case, from multiple, disparate data tools – upon which a site map, information hierarchy, and streamlined content terminology for a redesigned product can be built.

In preparation for our card sort activity, we consolidated 85 high-level terms – primarily indicating navigation and core functionality – from all of the data tools in the scope of this project (and a few from some of our "agency comparisons" task). The card sort activity itself was conducted with 21 total participants and consisted of two tasks: participants were asked to (1) sort the various terms cards into categories that made sense to them, and then (2) to prioritize those categories according to what was most important for their particular use of the tools.

Below are selected results of this card sort activity. It's important to note that while the relatively small sample size of responses – particularly those responses that indicate consensus – warrant continued validation, the results of this card sort are nevertheless helpful for informing the *initial* concepts for high-level navigation for a unified portal or harmonized user interfaces.

4.6.1. Top Categories by Priority

The most common categories in the top 5 priority groups consist of the following:

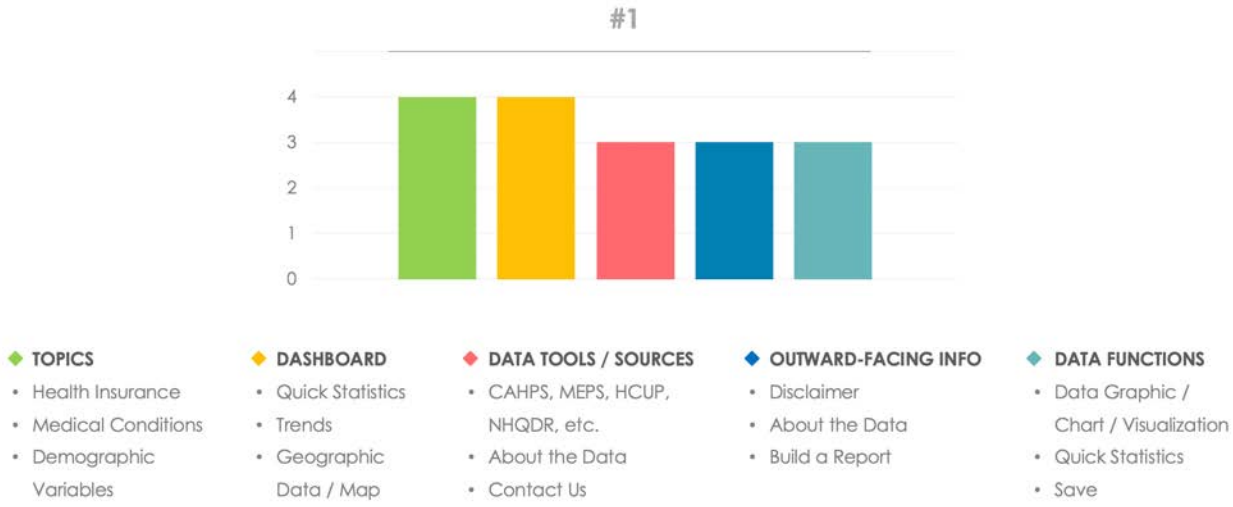
- **Dashboard / Customization**: all content and functionality related to a dashboard – in other words, a highly visual, easy-to-read landing page of high-level, summarized information – that potentially includes some level of personalization / customization based on the user
- **Data Functions / Tool Interface:** functionality and look/feel for the user interface of a basic Data Query Tool
- **Data Tools / Sources:** the specific data tools themselves and the data sources upon which they are based
- **Data Types:** specific variables and data characteristics
- **Programming / Customization:** functionality related to manipulation of raw data files
- **Outward-Facing Information:** all outward-facing summaries of data, including reports, chart books, and stat briefs
- **Topics:** broad topic areas touched upon by the various data tools

These categories mapped onto the top five priority groups as follows:

Among the potential applications of these ranked content categories is a data portal landing page. For example, the prominence of a user-friendly dashboard among the priority #1 categories indicates a need for a landing page that contains high-level, summarized content displayed in an easy-to-digest fashion in one place. Furthermore, the prominence of the "Data Tools," "Topics," and "Data Functions" categories in priority groups 1 and 2 indicates a potential solution to how data tools could be displayed on such a dashboard: each tool could be described by their name, function, and the topic areas they address.

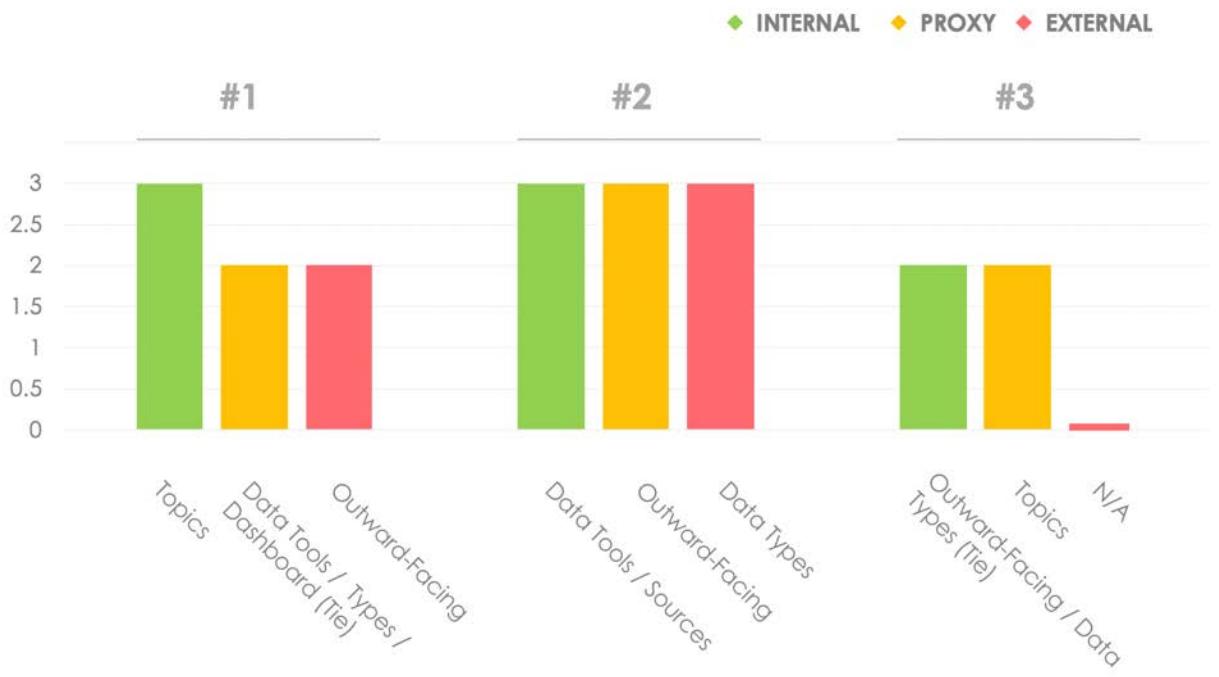
4.6.2. Priority #1 Categories & Terms

Competing for the #1 priority group are Topics, Dashboard, Data Tools / Sources, Outward-Facing Information, and Data Functions categories. Below is a breakdown of these categories and the most popular terms for each category.



4.6.3. Top Categories by User Type

The prompt for the card sort activity was slightly different depending on whether the participant was an internal/external user or a "proxy" user (read: "Support"); for the former, the participant was asked to prioritize terms according to what was most important for their own use of the data tools, whereas proxy users were asked to prioritize categories based on their understanding of their importance for external users.



The Atlas team calculated the top three priority groups for content and functionality for each of these user types: internal users, proxy users, and external users. The Atlas team found variable and limited consensus among these user types. This variability can be interpreted as speaking to the unique needs of each user type across all the tools and reinforces the need for further validation of prioritized terminology for a portal/harmonized interfaces.

As with all of the card sort results, it is their application to concepts – wireframes and prototypes – and subsequent user feedback that will yield the greatest insights.

4.7. General Usability Best Practices

In addition to the requirements outlined above, the Atlas Team recommends the following high-level usability best practices:

4.7.1. Brand Consistency

While maintaining branding for each of the individual data tools, establish a common style guide for each, including fonts, colors, formatting, basic UI, and core terminology (particularly for high-level navigation). This will lend to an improved UX and will reinforce AHRQ as an overarching entity and an accountable customer service resource for all tools.

4.7.2. Visual Hierarchy

Overall, AHRQ's data tools lack effective visual hierarchy, which refers to the delineation of levels of importance among text fields – from headings to body text to captions – largely using text size, weight, color, white space, nesting, collapsible text, and highlighting of key terms. This goes hand-in-hand with the next usability principle: Streamlined Content.

4.7.3. Streamlined Content

This principle is perhaps the most important. While much explanatory text might seem absolutely necessary, from a usability perspective it presents unnecessary noise for users who are seeking familiar or desired functionality only. The Card Sort analysis can provide a sense of how this content can be streamlined, but generally it's advisable to omit needless words, keep paragraphs short, and use bulleted lists. When dense explanatory content cannot be reduced, effective visual hierarchy can be applied.

4.7.4. Visual Signifiers

As noted in some user requirements, visual signifiers are proven to reduce cognitive load of information, allowing a user to parse larger volumes of complex information. Visual signifiers can take the form of data visualizations, icons, and effective color coding.

4.7.5. UI Conventions

Users rely on familiar UI conventions to quickly grasp where they are and how to use a new website (e.g., sidebar navigation, or upper-right hand search bars).

5. CONCLUSION

Based on the findings from the Environmental Scan, the Atlas Team recommended that AHRQ prioritize the requirements contained herein based on their viability and feasibility for the various data tool POCs. That prioritized list can be used to inform the selection of a COTS data visualization and analysis product and a harmonized UI within which it will operate. The Atlas Team will then validate and iterate on those prioritized requirements in collaboration with the users identified over the course of the Environmental Scan during the subsequent Style Examples and Prototyping phases.

To execute this recommendation, the Atlas Team hosted a Mid-Phase Workshop at AHRQ on January 16, 2019. During the workshop, the Atlas Team presented an executive summary presentation of this report at an AHRQ Data Tools Mid-Phase Presentation and Co-Design Workshop to the POCS of all the relevant data tools and summary data tables in the scope of this project. The purpose of this workshop was to present and validate the findings in this report, to gather feedback and ideas, and to prioritize and elaborate on individual and collective user requirements.

There were two feedback exercises during this presentation and workshop. The first exercise was an opportunity for meeting attendees to share which slides they wanted to discuss further as a group because they had questions, didn't understand something, liked what they saw or heard and wanted to hear more, or wanted to discuss a connection between the slide content and the current AHRQ Data Tool Harmonization Project.

This first exercise invited all meeting participants to place color stickers (colors related to the specific data tool they represented in the meeting) on slide print-outs. Each attendee received 6 stickers. The slides receiving the most stickers were discussed further. The top three slides were:

1. NCHS Data Visualization Gallery
2. Requirement: Knowing Where to Start My Query - I need a simple, predictive search engine in order to find everything from estimates to survey questions.
3. Requirement: Knowing Where to Start My Query - data.HRSA.gov Website Example

The NCHS Data Visualization Gallery example sparked a rich discussion about the role of visualizations in the research process (e.g., do visitors choose the visualization first?), the power of visualizations to make data interesting (e.g., many remarked on how interesting the visualizations looked and that they voted to discuss this slide because they were curious to see how these visualizations worked), how data visualizations, especially maps, communicate statistical significance, and how to address that not all data within AHRQ tools is applicable to a map visualization. There was consensus that a map visualization should be a key requirement in choosing a data visualization product but that further discussion is needed to reach agreement on how to display differences across US states.

Those that voted to discuss the need for "a simple, predictive search engine" wanted to better understand what predictive search is and whether "search" functionality is within the scope of this data tool harmonization project. When we reference "predictive search" we are referring to the use of an algorithm that uses user data to auto-fill the rest of the search query or to offer suggestions of what the user is attempting to do. The current project is intended to produce a prototype of the harmonized data tool and one additional page related to the use of this new tool. Therefore, site wide functionality or even the implementation of search within the harmonized data tool is out of the scope of this current project. This conversation did raise many user-centric comments, such as suggesting that the current site experience is focused on using the standard language used within AHRQ but not commonly used outside of AHRQ. It also raised the question of the user journeys to the information they want. How do they ask their research questions? Finally, it raised interesting user experience options, such as using search results to make suggestions of data or content the user might be interested in based on their search terms (e.g., "you may be interested in" or "others searching X also viewed Y"). The project team will keep these thoughts with us as we continue to explore data visualization products and design the information architecture and user journeys through the tool.

Finally, attendees voted to discuss the data.HRSA.gov website example because they liked how easy it is to see the paths to different parts of the organization as well as a mix of showing all the topics they cover as well as how to highlight specific data points or statistics from the data that would interest and help users connect the data to their queries.

The second workshop activity, "Ideation," where attendees were given numbered, colored dots and asked to prioritize user requirements based on what they feel is important for the individual data tool they represent as well as what is important for the collective harmonized tool. Stickers were numbered 1-3, and each attendee received 2 of each number.

The themes for the requirements that received the most votes across all tools were:

- Accessibility Related (across different devices, operating systems (OS), and browsers)
- Visualization Options (ability to select different "view" formats)
- Supports Multiple Paths to a Result (allows search and topic selection)
- Conducts Statistical Tests
- Saving and Sharing Results (ability to export visualizations, save them, and use them in reports, presentations, or other formats)

- Data Download Options (ability to download data sets into excel or other formats for different statistical tools like SPSS or SAS)

While it was not on a card, the discussion did net an additional requirement that the webpage containing the harmonized tool would also include high-level descriptions of the AHRQ program and the different data sets. The content would address the following topics:

- What the program does
- What the surveys measure
- Which tool do you go to with which type of questions

These user requirements will be elevated when reviewing data visualization products and when considering style requirements and UX design.

Please refer to Appendix C for photos taken of the activities during the workshop.

A. ACRONYMS AND DEFINITIONS

Acronym	Description
Adm	Administrator Users
AHRQ	Agency for Healthcare Research and Quality
CAHPS	Consumer Assessment of Healthcare Providers and Systems
CCS	Clinical Classifications Software
COTS	Commercial off-the-Shelf
HC	Household Component
HCD	Human Centered Design
HCUP	Health Cost and Utilization Project
HRSA	Health Resources and Services Administration
IC	Insurance Component
ICD-9	International Classification of Diseases, 9th Revision
ICD-10	International Classification of Diseases, 10th Revision
Int	Intermediate Users (External)
IPUMS	Integrated Public Use Microdata Series
LOE	Level of Effort
MEPS	Medical Expenditure Panel Survey
NHQDRnet	National Healthcare Quality and Disparities Reports Data Query Tool
NCHS	National Center for Health Statistics
NIH	National Institutes of Health
Nov	Novice / Quick-Reference Users (External)
OS	Operating System
PDF	Portable Document Format (File Extension)
POC	Point of Contact
PhD	Doctor of Philosophy
Pro	Pro Users (External)
R	R Programming Language
SAS	Statistical Analysis Software®

Acronym	Description
Sup	Customer Support Users (Internal)
UI	User Interface
URL	Uniform Resource Locator
UX	User Experience
VA	Department of Veterans Affairs

B. BACKGROUND USER RESEARCH

B.1. Sample Discussion Guide

The following questions comprise our standardized discussion guide. Questions were adjusted where appropriate to suit the relevant interviewee / user type.

- General / Contextual Questions
    - Please *state your name, role and history in the organization.* What are the core activities / goals of your day-to-day work?
    - In a sentence or two, describe how all of AHRQ's data tools function in your day-to-day work. For what tasks are they useful? How vital are they to these various duties? Which do you use concurrently? Are you aware of what the other tools are and what they do? What are they? Of the ones you use, how could they be better consolidated if at all? Can you speak to some of the potential benefits / use cases of harmonizing any of these tools? Do you need guidance about which tool would best fit your needs? In what cases? Do you have any concerns or suggested considerations regarding the prospect of greater harmonization between the tools? About adopting a COTS product? Do you see AHRQ as an overarching entity for these tools? Or do you view each particular tool as independent from the others?
    - Administrators: Who are the users of this tool? Do you have web traffic metrics? Customer feedback? What questions would you have for the users of these tools?
- Simulated Scenarios / Observation
    - Please describe a recent or hypothetical situation in which you needed to use the AHRQ data tools for non-administrative purposes (e.g., research and analysis). How important is the use of this tool for this purpose? What was the initial impetus / problem(s)? How frequently does it occur? In what context does this usually happen? Are you at any point using a phone or desktop device to complete part of this task? What sub-tasks do you need to accomplish, in sequence? What is the frequency of each? What is the relative importance of

each? What other tools or resources do you leverage (e.g., software, other websites) to accomplish this task, in sequence? What other AHRQ data tools? Are you reliant on other stakeholders at any point, or communication with them? If so, please describe them, their role, and what specific sub-tasks they need to accomplish. What analyses of data are you seeking to accomplish? What's feasible within the tools available to you, and what's not? What would be ideal? How do you record the information you find? What output do you require to successfully accomplish the data analysis portion of the task? Do you consider the data / analyses provided to be accurate & reliable? Why or why not?

- Please walk us through a recent scenario in which you used the AHRQ tool. We'd like to watch as you simulate achieving that prompt/query; we ask that you describe what you're doing as you do it and why. How often do you use this specific data tool? Do you find you repeat certain queries? Are you looking for the same thing consistently, or a new query every time? How often a day, week, year? What are common mistakes / corrections you make in this process? What are the risks of failure?

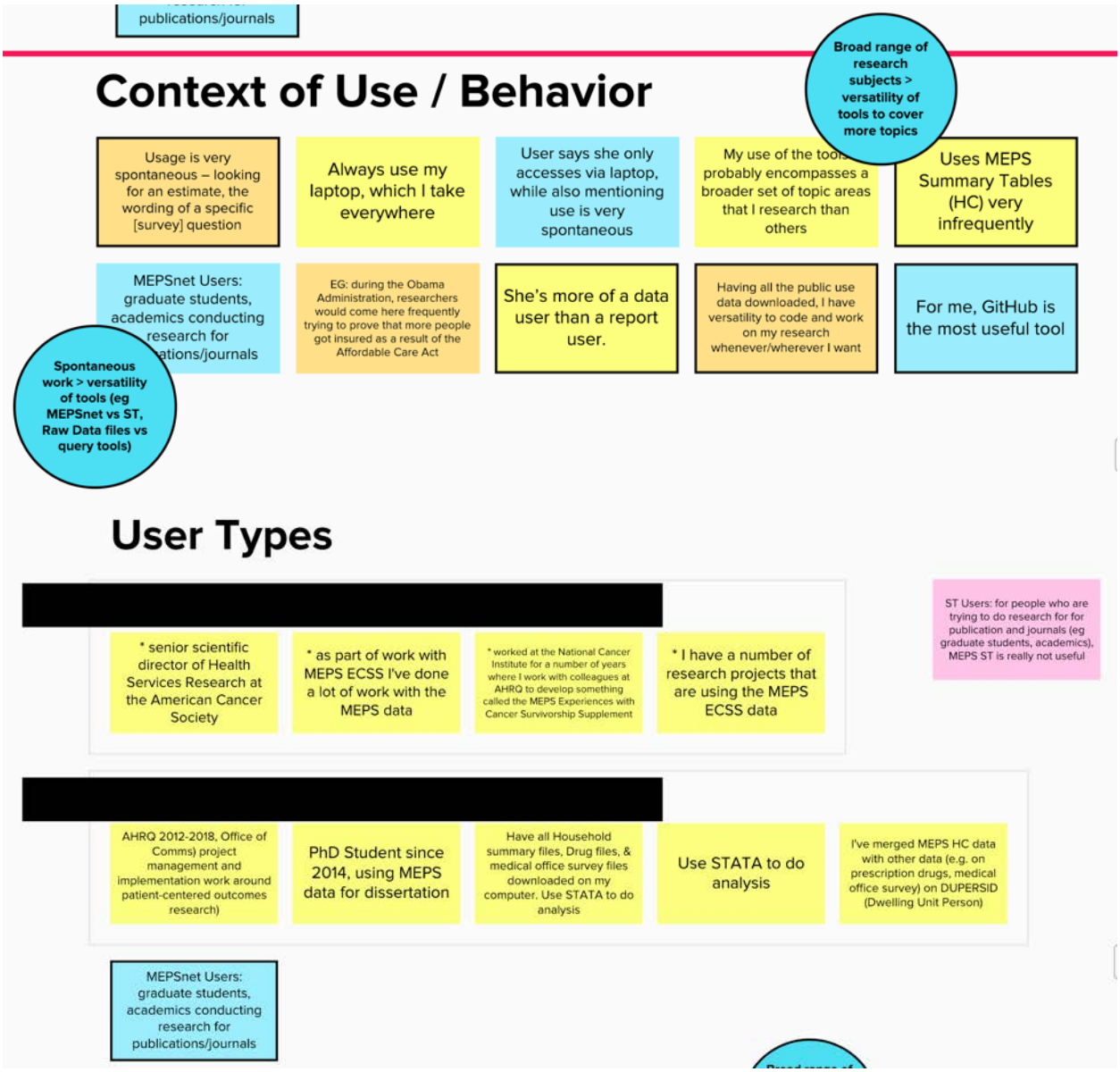
- Do you use this tool exclusively, or in combination with any of the other tools? Please demonstrate a scenario where you'd use this in tandem with other tools. What features of each tool do you use in tandem? In what sequence? Can you show us?

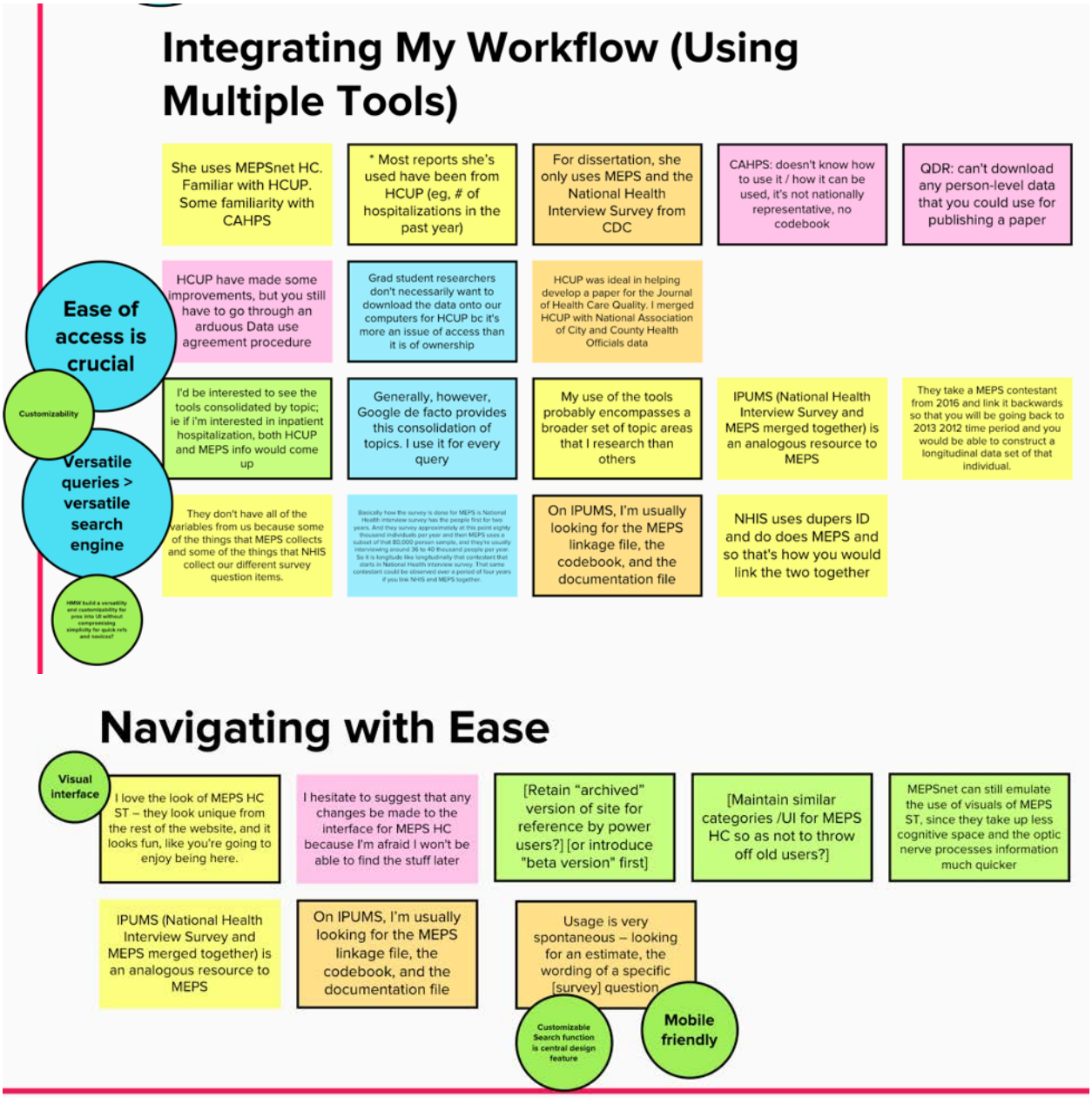
- What features do you love? Which features do you dislike? Which features would be ideal? Let's take a step back and consider best practices. In your mind, what is the best example of a data set or data table that was easy to use and understand, and best aided you in accomplishing a vital task? Please show us. What do you expect in a great tool for conducting analysis?

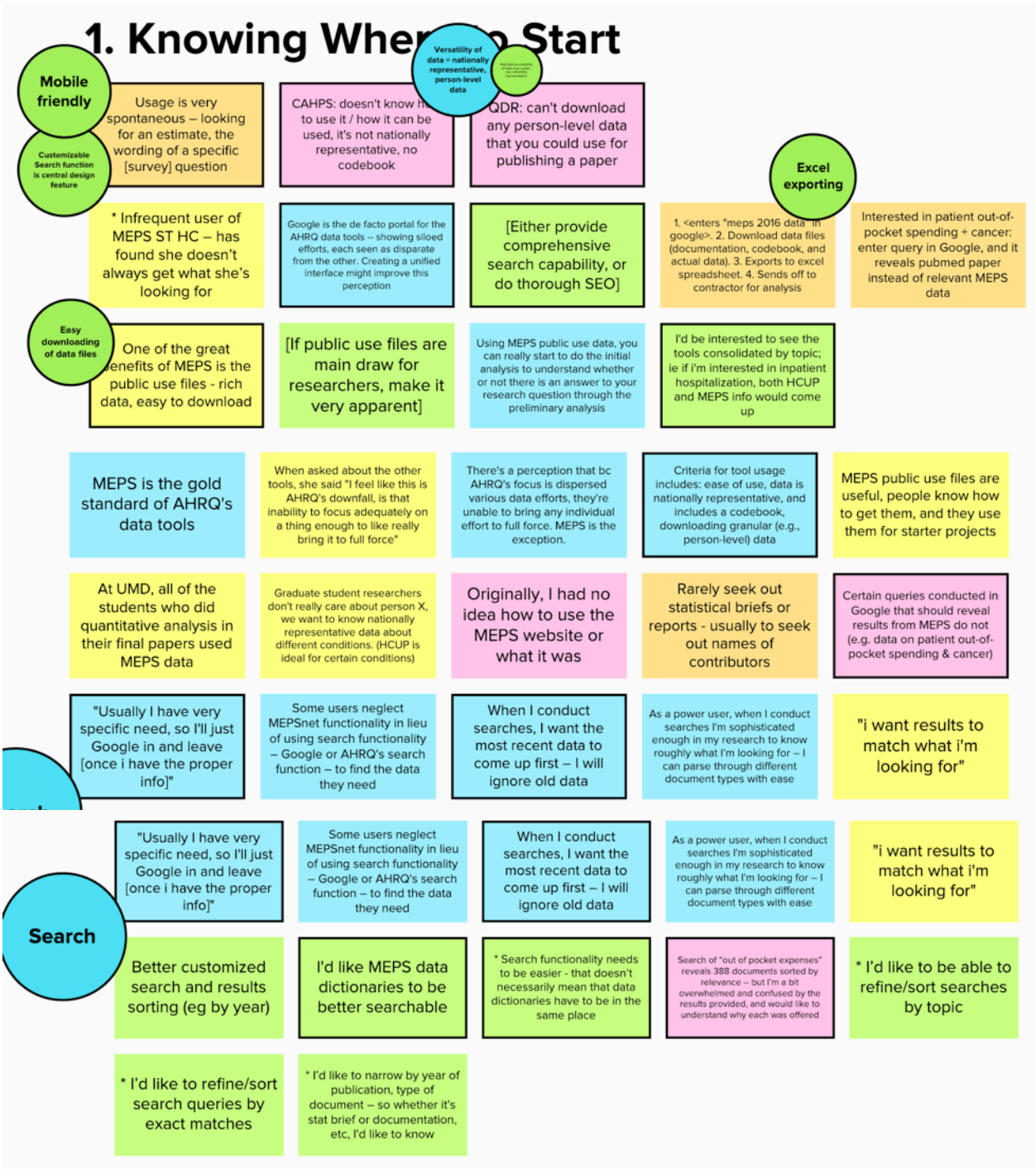
- Administrators: Please describe 1-2 real or hypothetical situations in which you accessed the tools for administrative/backend use (e.g., maintenance, content management).

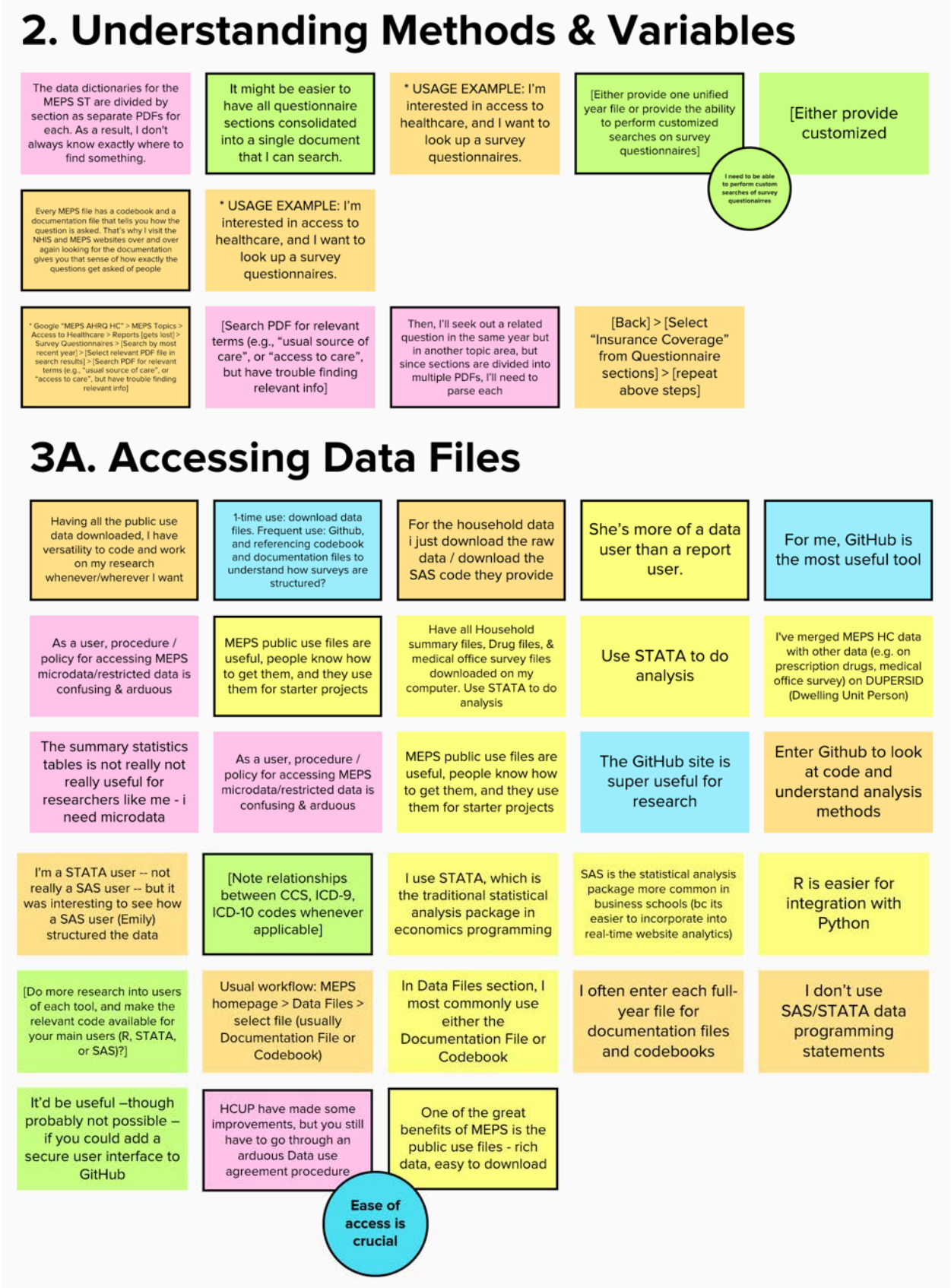
B.2. User Group Data Points

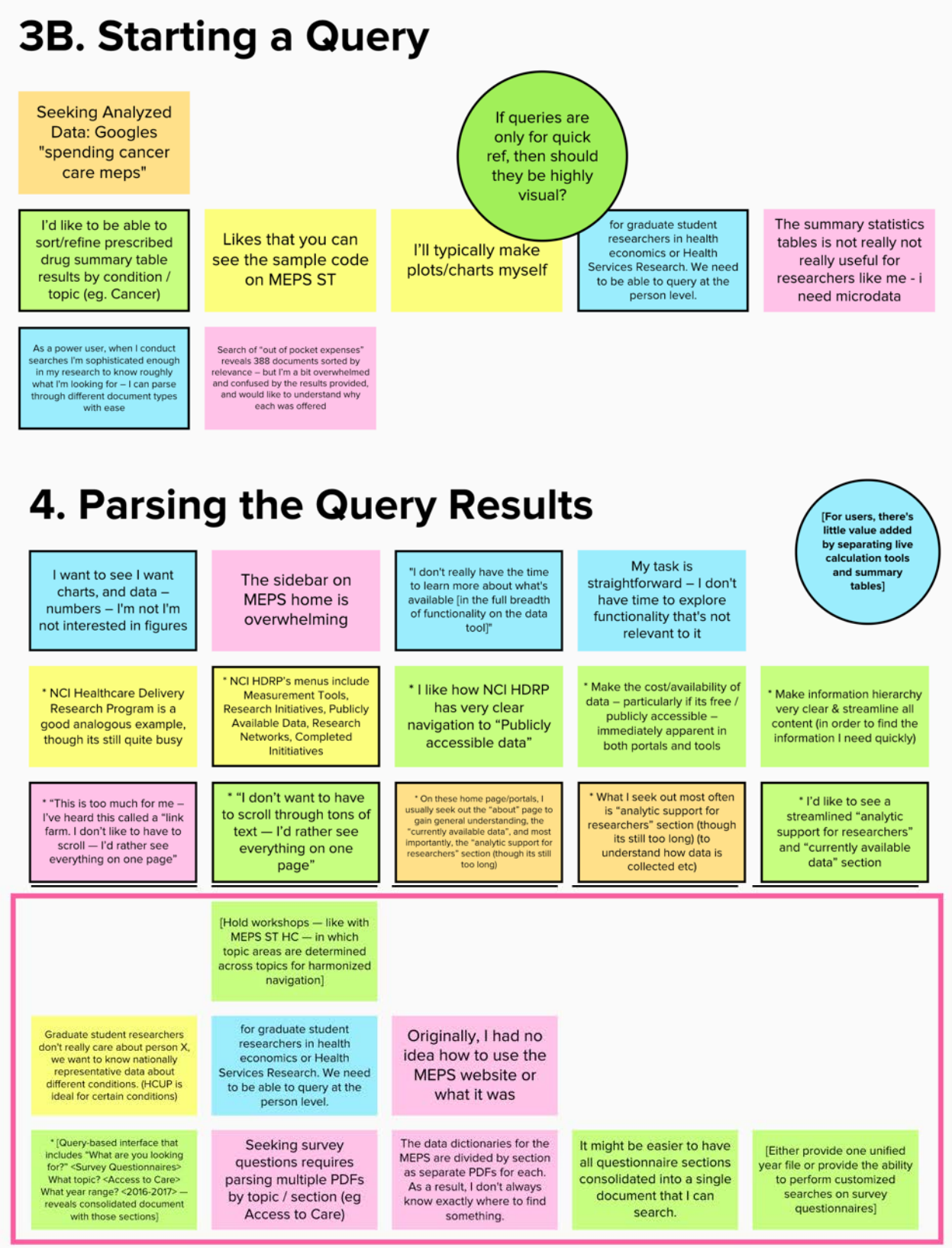
The following selected data points were gathered and then synthesized using a Design Thinking qualitative analysis method called Affinity Diagramming. The data points were color coded (see key below). Below, we've listed some of the most insightful data points consolidated from our interviews with this user group.

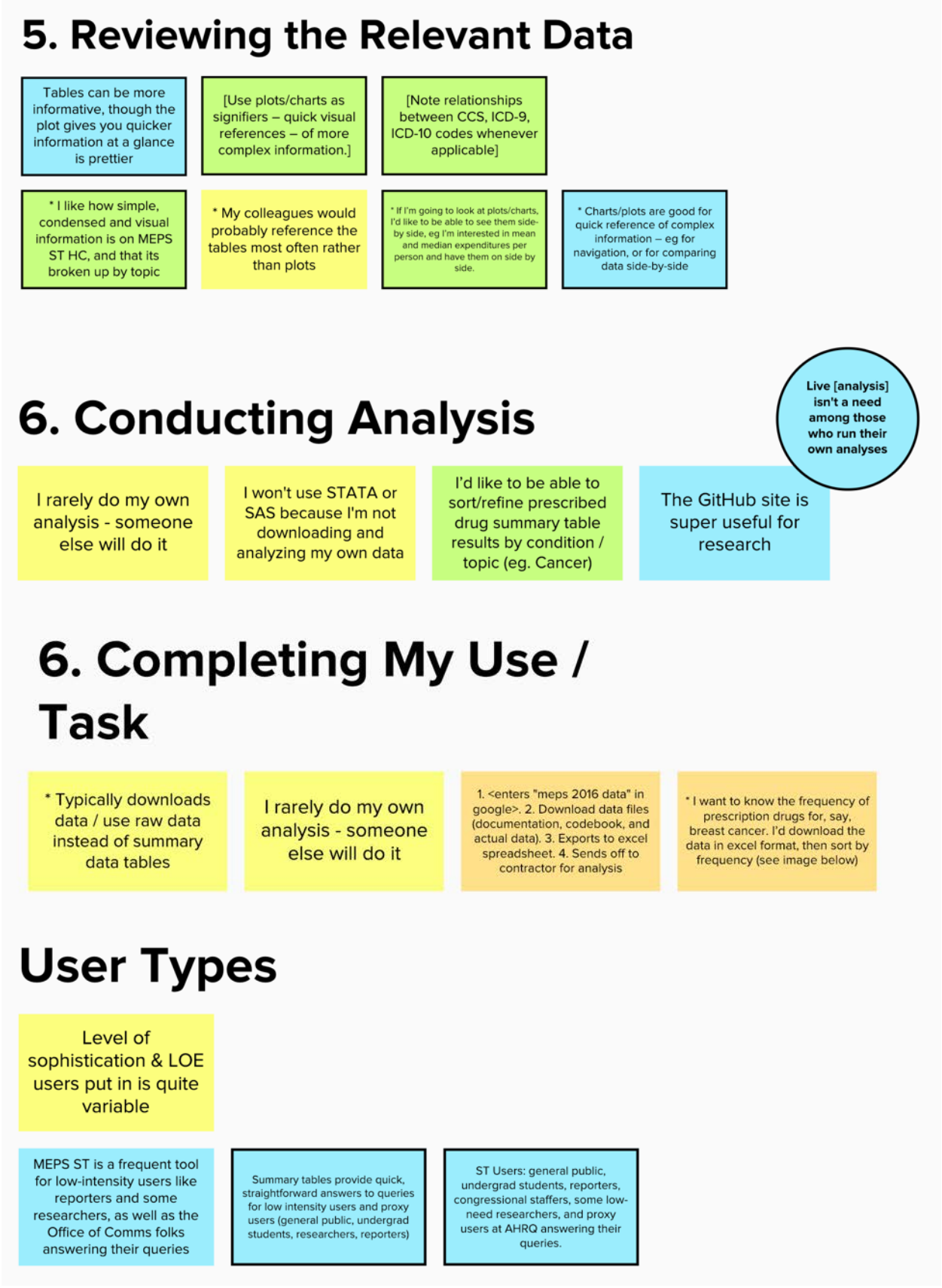

publications/journals
Context of Use / Behavior
Broad range of research subjects > versatility of tools to cover more topics
Usage is very spontaneous – looking for an estimate, the wording of a specific [survey] question
Always use my laptop, which I take everywhere
User says she only accesses via laptop, while also mentioning use is very spontaneous
My use of the tool probably encompasses a broader set of topic areas that I research than others
Uses MEPS Summary Tables (HC) very infrequently
MEPSnet Users: graduate students, academics conducting research for ations/journals
EG: during the Obama Administration, researchers would come here frequently trying to prove that more people got insured as a result of the Affordable Care Act
She's more of a data user than a report user.
Having all the public use data downloaded, I have versatility to code and work on my research whenever/wherever I want
For me, GitHub is the most useful tool
Spontaneous work > versatility of tools (eg MEPSnet vs ST, Raw Data files vs query tools)
User Types
[redacted]
* senior scientific director of Health Services Research at the American Cancer Society
* as part of work with MEPS ECSS I've done a lot of work with the MEPS data
* worked at the National Cancer Institute for a number of years where I work with colleagues at AHRQ to develop something called the MEPS Experiences with Cancer Survivorship Supplement
* I have a number of research projects that are using the MEPS ECSS data
ST Users: for people who are trying to do research for for publication and journals (eg graduate students, academics), MEPS ST is really not useful
[redacted]
AHRQ 2012-2018, Office of Comms) project management and implementation work around patient-centered outcomes research)
PhD Student since 2014, using MEPS data for dissertation
Have all Household summary files, Drug files, & medical office survey files downloaded on my computer. Use STATA to do analysis
Use STATA to do analysis
I've merged MEPS HC data with other data (e.g. on prescription drugs, medical office survey) on DUPERSID (Dwelling Unit Person)
MEPSnet Users: graduate students, academics conducting research for publications/journals


Integrating My Workflow (Using Multiple Tools)
She uses MEPSnet HC. Familiar with HCUP. Some familiarity with CAHPS
* Most reports she's used have been from HCUP (eg, # of hospitalizations in the past year)
For dissertation, she only uses MEPS and the National Health Interview Survey from CDC
CAHPS: doesn't know how to use it / how it can be used, it's not nationally representative, no codebook
QDR: can't download any person-level data that you could use for publishing a paper
Ease of access is crucial
HCUP have made some improvements, but you still have to go through an arduous Data use agreement procedure
Grad student researchers don't necessarily want to download the data onto our computers for HCUP bc it's more an issue of access than it is of ownership
HCUP was ideal in helping develop a paper for the Journal of Health Care Quality. I merged HCUP with National Association of City and County Health Officials data
Customizability
I'd be interested to see the tools consolidated by topic; ie if i'm interested in inpatient hospitalization, both HCUP and MEPS info would come up
Generally, however, Google de facto provides this consolidation of topics. I use it for every query
My use of the tools probably encompasses a broader set of topic areas that I research than others
IPUMS (National Health Interview Survey and MEPS merged together) is an analogous resource to MEPS
They take a MEPS contestant from 2016 and link it backwards so that you will be going back to 2013 2012 time period and you would be able to construct a longitudinal data set of that individual.
Versatile queries > versatile search engine
They don't have all of the variables from us because some of the things that MEPS collects and some of the things that NHIS collect our different survey question items.
On IPUMS, I'm usually looking for the MEPS linkage file, the codebook, and the documentation file
NHIS uses dupers ID and do does MEPS and so that's how you would link the two together
Navigating with Ease
Visual interface
I love the look of MEPS HC ST – they look unique from the rest of the website, and it looks fun, like you're going to enjoy being here.
I hesitate to suggest that any changes be made to the interface for MEPS HC because I'm afraid I won't be able to find the stuff later
[Retain "archived" version of site for reference by power users?] [or introduce "beta version" first]
[Maintain similar categories /UI for MEPS HC so as not to throw off old users?]
MEPSnet can still emulate the use of visuals of MEPS ST, since they take up less cognitive space and the optic nerve processes information much quicker
IPUMS (National Health Interview Survey and MEPS merged together) is an analogous resource to MEPS
On IPUMS, I'm usually looking for the MEPS linkage file, the codebook, and the documentation file
Usage is very spontaneous – looking for an estimate, the wording of a specific [survey] question
Customizable Search function is central design feature
Mobile friendly


1. Knowing Where to Start
Mobile friendly
Customizable Search function is central design feature
Usage is very spontaneous – looking for an estimate, the wording of a specific [survey] question
CAHPS: doesn't know how to use it / how it can be used, it's not nationally representative, no codebook
Versatility of data = nationally representative, person-level data
QDR: can't download any person-level data that you could use for publishing a paper
Excel exporting
* Infrequent user of MEPS ST HC – has found she doesn't always get what she's looking for
Google is the de facto portal for the AHRQ data tools -- showing siloed efforts, each seen as disparate from the other. Creating a unified interface might improve this perception
[Either provide comprehensive search capability, or do thorough SEO]
1. <enters "meps 2016 data" in google>. 2. Download data files (documentation, codebook, and actual data). 3. Exports to excel spreadsheet. 4. Sends off to contractor for analysis
Interested in patient out-of-pocket spending + cancer: enter query in Google, and it reveals pubmed paper instead of relevant MEPS data
Easy downloading of data files
One of the great benefits of MEPS is the public use files - rich data, easy to download
[If public use files are main draw for researchers, make it very apparent]
Using MEPS public use data, you can really start to do the initial analysis to understand whether or not there is an answer to your research question through the preliminary analysis
I'd be interested to see the tools consolidated by topic; ie if i'm interested in inpatient hospitalization, both HCUP and MEPS info would come up
MEPS is the gold standard of AHRQ's data tools
When asked about the other tools, she said "I feel like this is AHRQ's downfall, is that inability to focus adequately on a thing enough to like really bring it to full force"
There's a perception that bc AHRQ's focus is dispersed various data efforts, they're unable to bring any individual effort to full force. MEPS is the exception.
Criteria for tool usage includes: ease of use, data is nationally representative, and includes a codebook, downloading granular (e.g., person-level) data
MEPS public use files are useful, people know how to get them, and they use them for starter projects
At UMD, all of the students who did quantitative analysis in their final papers used MEPS data
Graduate student researchers don't really care about person X, we want to know nationally representative data about different conditions. (HCUP is ideal for certain conditions)
Originally, I had no idea how to use the MEPS website or what it was
Rarely seek out statistical briefs or reports - usually to seek out names of contributors
Certain queries conducted in Google that should reveal results from MEPS do not (e.g. data on patient out-of-pocket spending & cancer)
"Usually I have very specific need, so I'll just Google in and leave [once i have the proper info]"
Some users neglect MEPSnet functionality in lieu of using search functionality – Google or AHRQ's search function – to find the data they need
When I conduct searches, I want the most recent data to come up first – I will ignore old data
As a power user, when I conduct searches I'm sophisticated enough in my research to know roughly what I'm looking for – I can parse through different document types with ease
"i want results to match what i'm looking for"
Search
"Usually I have very specific need, so I'll just Google in and leave [once i have the proper info]"
Some users neglect MEPSnet functionality in lieu of using search functionality – Google or AHRQ's search function – to find the data they need
When I conduct searches, I want the most recent data to come up first – I will ignore old data
As a power user, when I conduct searches I'm sophisticated enough in my research to know roughly what I'm looking for – I can parse through different document types with ease
"i want results to match what i'm looking for"
Better customized search and results sorting (eg by year)
I'd like MEPS data dictionaries to be better searchable
* Search functionality needs to be easier - that doesn't necessarily mean that data dictionaries have to be in the same place
Search of "out of pocket expenses" reveals 388 documents sorted by relevance – but I'm a bit overwhelmed and confused by the results provided, and would like to understand why each was offered
* I'd like to be able to refine/sort searches by topic
* I'd like to refine/sort search queries by exact matches
* I'd like to narrow by year of publication, type of document – so whether it's stat brief or documentation, etc, I'd like to know

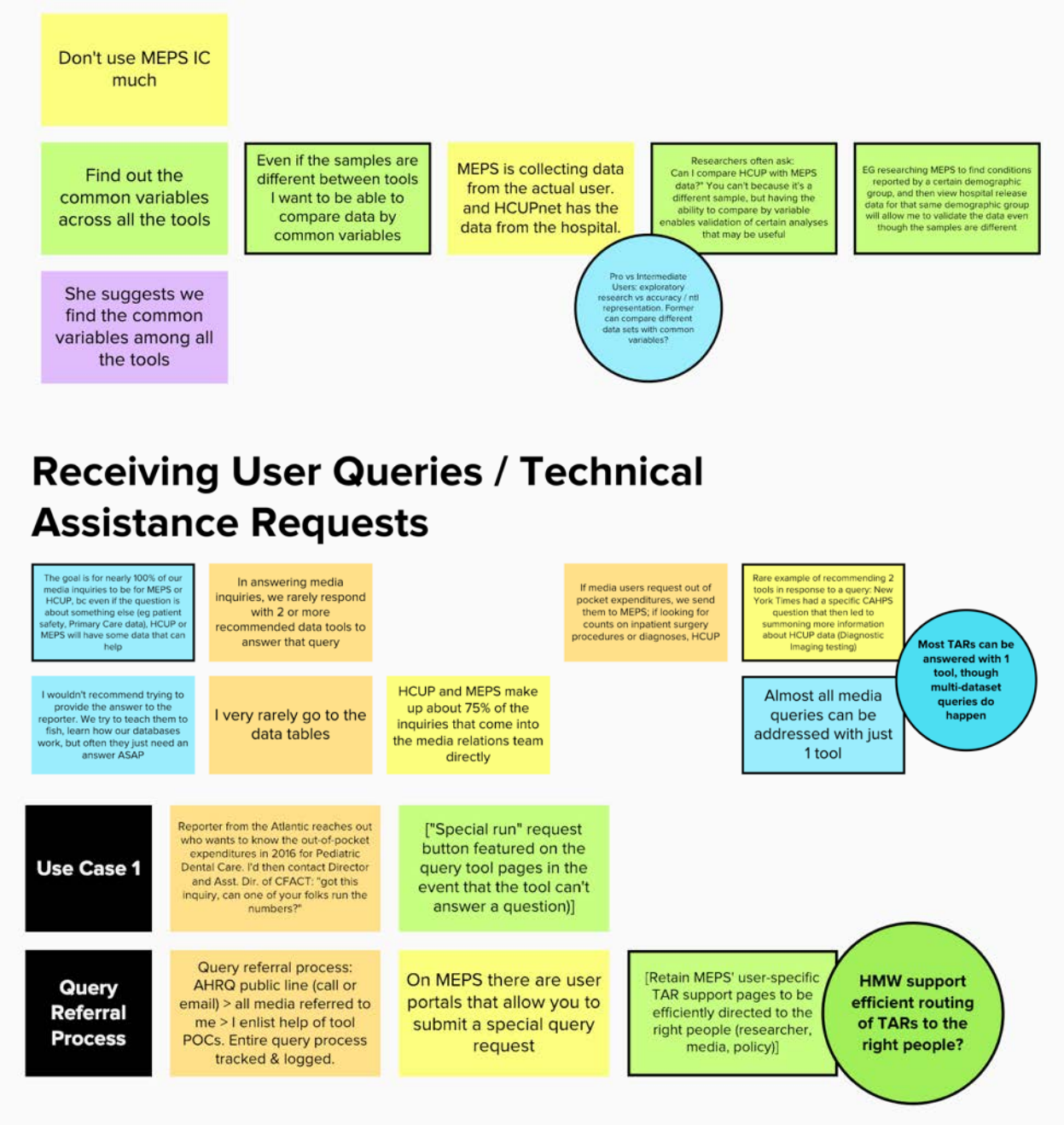

2. Understanding Methods & Variables
The data dictionaries for the MEPS ST are divided by section as separate PDFs for each. As a result, I don't always know exactly where to find something.
It might be easier to have all questionnaire sections consolidated into a single document that I can search.
* USAGE EXAMPLE: I'm interested in access to healthcare, and I want to look up a survey questionnaires.
[Either provide one unified year file or provide the ability to perform customized searches on survey questionnaires]
[Either provide customized
I need to be able to perform custom searches of survey questionairres
Every MEPS file has a codebook and a documentation file that tells you how the question is asked. That's why I visit the NHIS and MEPS websites over and over again looking for the documentation gives you that sense of how exactly the questions get asked of people
* USAGE EXAMPLE: I'm interested in access to healthcare, and I want to look up a survey questionnaires.
* Google "MEPS AHRQ HC" > MEPS Topics > Access to Healthcare > Reports [gets lost] > Survey Questionnaires > [Search by most recent year] > [Select relevant PDF file in search results] > [Search PDF for relevant terms (e.g., "usual source of care", or "access to care", but have trouble finding relevant info]
[Search PDF for relevant terms (e.g., "usual source of care", or "access to care", but have trouble finding relevant info]
Then, I'll seek out a related question in the same year but in another topic area, but since sections are divided into multiple PDFs, I'll need to parse each
[Back] > [Select "Insurance Coverage" from Questionnaire sections] > [repeat above steps]
3A. Accessing Data Files
Having all the public use data downloaded, I have versatility to code and work on my research whenever/wherever I want
1-time use: download data files. Frequent use: Github, and referencing codebook and documentation files to understand how surveys are structured?
For the household data i just download the raw data / download the SAS code they provide
She's more of a data user than a report user.
For me, GitHub is the most useful tool
As a user, procedure / policy for accessing MEPS microdata/restricted data is confusing & arduous
MEPS public use files are useful, people know how to get them, and they use them for starter projects
Have all Household summary files, Drug files, & medical office survey files downloaded on my computer. Use STATA to do analysis
Use STATA to do analysis
I've merged MEPS HC data with other data (e.g. on prescription drugs, medical office survey) on DUPERSID (Dwelling Unit Person)
The summary statistics tables is not really not really useful for researchers like me - i need microdata
As a user, procedure / policy for accessing MEPS microdata/restricted data is confusing & arduous
MEPS public use files are useful, people know how to get them, and they use them for starter projects
The GitHub site is super useful for research
Enter Github to look at code and understand analysis methods
I'm a STATA user -- not really a SAS user -- but it was interesting to see how a SAS user (Emily) structured the data
[Note relationships between CCS, ICD-9, ICD-10 codes whenever applicable]
I use STATA, which is the traditional statistical analysis package in economics programming
SAS is the statistical analysis package more common in business schools (bc its easier to incorporate into real-time website analytics)
R is easier for integration with Python
[Do more research into users of each tool, and make the relevant code available for your main users (R, STATA, or SAS)?]
Usual workflow: MEPS homepage > Data Files > select file (usually Documentation File or Codebook)
In Data Files section, I most commonly use either the Documentation File or Codebook
I often enter each full-year file for documentation files and codebooks
I don't use SAS/STATA data programming statements
It'd be useful –though probably not possible – if you could add a secure user interface to GitHub
HCUP have made some improvements, but you still have to go through an arduous Data use agreement procedure
One of the great benefits of MEPS is the public use files - rich data, easy to download
Ease of access is crucial

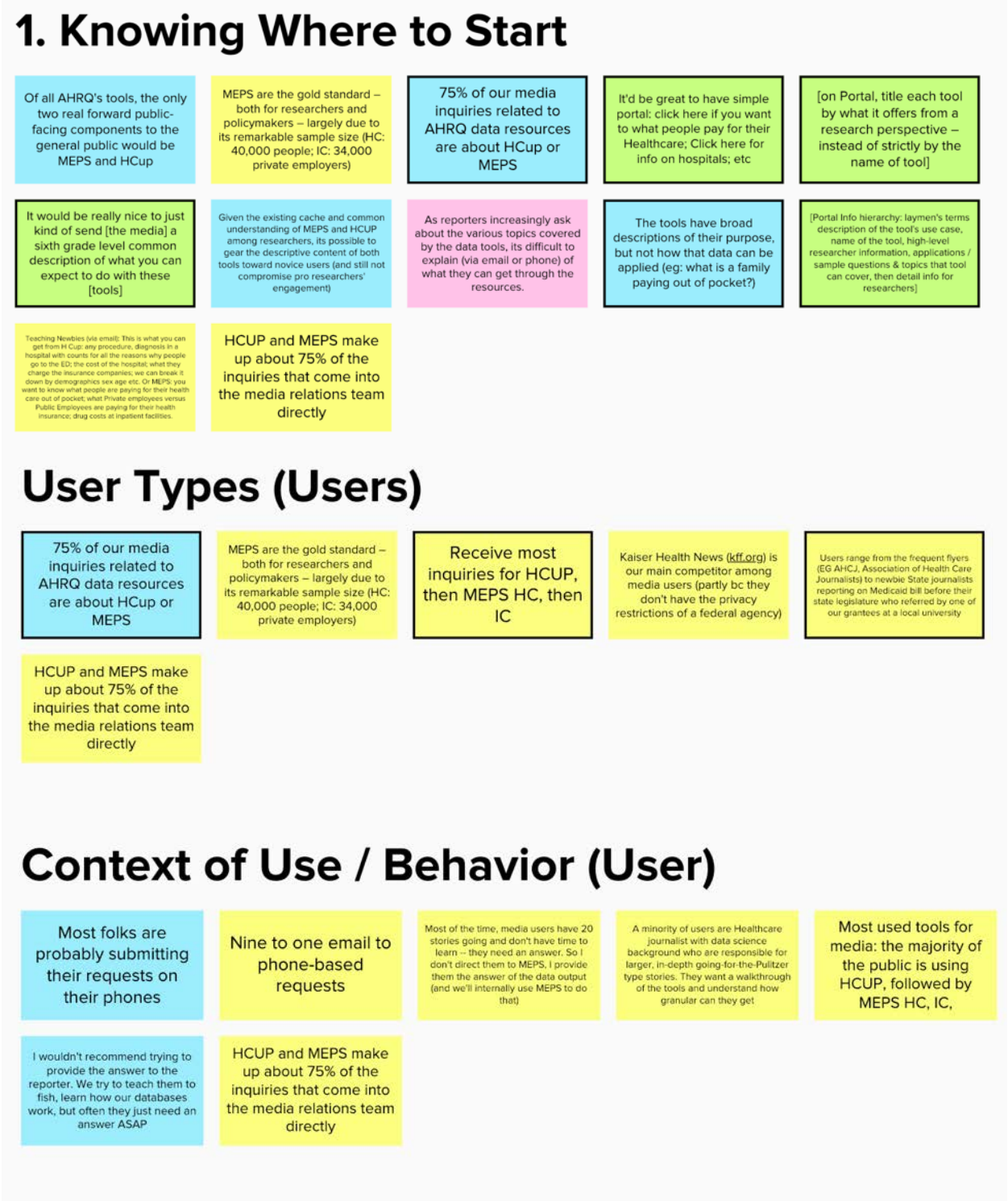

3B. Starting a Query
Seeking Analyzed Data: Googles "spending cancer care meps"
If queries are only for quick ref, then should they be highly visual?
I'd like to be able to sort/refine prescribed drug summary table results by condition / topic (eg. Cancer)
Likes that you can see the sample code on MEPS ST
I'll typically make plots/charts myself
for graduate student researchers in health economics or Health Services Research. We need to be able to query at the person level.
The summary statistics tables is not really not really useful for researchers like me - i need microdata
As a power user, when I conduct searches I'm sophisticated enough in my research to know roughly what I'm looking for – I can parse through different document types with ease
Search of "out of pocket expenses" reveals 388 documents sorted by relevance – but I'm a bit overwhelmed and confused by the results provided, and would like to understand why each was offered
4. Parsing the Query Results
[For users, there's little value added by separating live calculation tools and summary tables]
I want to see I want charts, and data – numbers – I'm not I'm not interested in figures
The sidebar on MEPS home is overwhelming
"I don't really have the time to learn more about what's available [in the full breadth of functionality on the data tool]"
My task is straightforward – I don't have time to explore functionality that's not relevant to it
* NCI Healthcare Delivery Research Program is a good analogous example, though its still quite busy
* NCI HDRP's menus include Measurement Tools, Research Initiatives, Publicly Available Data, Research Networks, Completed Inititiatives
* I like how NCI HDRP has very clear navigation to "Publicly accessible data"
* Make the cost/availability of data – particularly if its free / publicly accessible – immediately apparent in both portals and tools
* Make information hierarchy very clear & streamline all content (in order to find the information I need quickly)
* "This is too much for me – I've heard this called a "link farm. I don't like to have to scroll — I'd rather see everything on one page"
* "I don't want to have to scroll through tons of text — I'd rather see everything on one page"
* On these home page/portals, I usually seek out the "about" page to gain general understanding, the "currently available data", and most importantly, the "analytic support for researchers" section (though its still too long)
* What I seek out most often is "analytic support for researchers" section (though its still too long) (to understand how data is collected etc)
* I'd like to see a streamlined "analytic support for researchers" and "currently available data" section
[Hold workshops — like with MEPS ST HC — in which topic areas are determined across topics for harmonized navigation]
Graduate student researchers don't really care about person X, we want to know nationally representative data about different conditions. (HCUP is ideal for certain conditions)
for graduate student researchers in health economics or Health Services Research. We need to be able to query at the person level.
Originally, I had no idea how to use the MEPS website or what it was
* [Query-based interface that includes "What are you looking for?" <Survey Questionnaires> What topic? <Access to Care> What year range? <2016-2017> — reveals consolidated document with those sections]
Seeking survey questions requires parsing multiple PDFs by topic / section (eg Access to Care)
The data dictionaries for the MEPS are divided by section as separate PDFs for each. As a result, I don't always know exactly where to find something.
It might be easier to have all questionnaire sections consolidated into a single document that I can search.
[Either provide one unified year file or provide the ability to perform customized searches on survey questionnaires]


5. Reviewing the Relevant Data
Tables can be more informative, though the plot gives you quicker information at a glance is prettier
[Use plots/charts as signifiers – quick visual references – of more complex information.]
[Note relationships between CCS, ICD-9, ICD-10 codes whenever applicable]
* I like how simple, condensed and visual information is on MEPS ST HC, and that its broken up by topic
* My colleagues would probably reference the tables most often rather than plots
* If I'm going to look at plots/charts, I'd like to be able to see them side-by side, eg I'm interested in mean and median expenditures per person and have them on side by side.
* Charts/plots are good for quick reference of complex information – eg for navigation, or for comparing data side-by-side
6. Conducting Analysis
Live [analysis] isn't a need among those who run their own analyses
I rarely do my own analysis - someone else will do it
I won't use STATA or SAS because I'm not downloading and analyzing my own data
I'd like to be able to sort/refine prescribed drug summary table results by condition / topic (eg. Cancer)
The GitHub site is super useful for research
6. Completing My Use / Task
* Typically downloads data / use raw data instead of summary data tables
I rarely do my own analysis - someone else will do it
1. <enters "meps 2016 data" in google>. 2. Download data files (documentation, codebook, and actual data). 3. Exports to excel spreadsheet. 4. Sends off to contractor for analysis
* I want to know the frequency of prescription drugs for, say, breast cancer. I'd download the data in excel format, then sort by frequency (see image below)
User Types
Level of sophistication & LOE users put in is quite variable
MEPS ST is a frequent tool for low-intensity users like reporters and some researchers, as well as the Office of Comms folks answering their queries
Summary tables provide quick, straightforward answers to queries for low intensity users and proxy users (general public, undergrad students, researchers, reporters)
ST Users: general public, undergrad students, reporters, congressional staffers, some low-need researchers, and proxy users at AHRQ answering their queries.


Navigating with Ease
I'm familiar with MEPS ST in my role at AHRQ Office of Comms, answering reporter queries
Office of Comms would use this website to reference #'s before they emailing researchers/reporters with an answer to their query
ST Users: for people who are trying to do research for for publication and journals (eg graduate students, academics), MEPS ST is really not useful
MEPS is a complicated survey & database, so it's inherently hard to make user friendly
[Better info hierarchy]
data.hrsa.gov is a best practice
The flexibility of MEPS HC ST makes it the foremost best practice
look into IPUMS too
1. Knowing Where to Start
Most research scenarios for my use case are regarding total expenditures or medical condition expenditures
Complicated Queries
Not all questions are about the actual numbers - some questions are about analyses
Some users go to the papers we write to see a summary/analysis of the data we have
People who ask for more complicated stuff related to their research - eg people who have diabetes and blood pressure -
~50% of queries are unique/complicated – users have a specific project ie they're not looking
3. Starting / Formulating a Query
User was looking to write about hypertension, and was looking for the usage of different drugs to treat hypertension over the years
Typical TAR request: user doing research on ER Admissions and the cost of those admissions whether it be for a specific condition or just the average
It would be helpful, tho not necessary, to have more customizable age groups on MEPS ST HC (instead of just age ranges)
Some users mights find MEPS ST user friendly, requiring little guidance - but it doesn't give them everything they need
MEPS ST: I grab the tables for total expenditures by therapeutic class across a certain year span, then copy that link and send it to the user
Sometimes i'll go to the actual data files to search for a particular variable that's not in the MEPS ST, then send to user and tell them how to use it
Responding to more complicated queries, I can calculate it separately and do my own programming or download the data and run it. That data is not on the tool [MEPS ST] because there are so many different combinations that people may be looking, and we don't have any tool that allows them to quickly calculate those numbers
Ability to customize MEPS HC ST searches
You can't combine certain data (e.g. people 40+ diabetes, blood pressure what are the expenditures per year) and MEPSnet doesn't have conditions


4. Parsing Query Results
It's hard to get users to read the descriptive analysis; I'll have to paste it into an email with the link
[Better info hierarchy]
5. Reviewing Data / Query Results
Users complain they can't get data on each condition on MEPSnet
Lack of data on each condition: EG you can collect data on total spending from year to year between different demographic groups, but if you're looking for expenditures on Diabetes among people 45 and older you can't get that
We guide MEPSnet users to ST when they want data on conditions etc (tho it's a different methodology)
I'd like to see a MEPS ST table that represents just one year at a time
One of favorite MEPSnet functionality is that if a user would like to return to a Query / retrace steps, they can save the link
7. Completing My Task
I want to be able to email a previous MEPS ST/MEPSnet query to myself
I'd like to be able to save a previous MEPS ST search to a dashboard, though it might not be practical because of gov't concerns over storing user data
Ability to amend previous searches on MEPS ST/MEPSnet
Way to create a dashboard without storing user data
Integrating Multiple Tools into Workflow
MEPS IC & HC
Experience limited to MEPS tools, very rare use of HCUP
Proxy Users vary widely in their knowledge of other AHRQ tools (potentially limiting their ability to provide cross-functional tech assistance?)
Fully aware of all the data tools
I do talk to the participants in HCUP workshops about hcupnet, but it's an entirely different type of data set - as well as CAHPS, QDR
The data sets of the AHRQ's tools are vastly different, but the users and their experience have some overlap in basic functionality and usability needs
I'll google each of the tools to navigate to them, unless i know the url
In addition to unified portal, urls for each tool, so they don't need to nav through google
I see the data tools as being independent – individual identities – rather than unified under AHRQ's brand
Significant differentiators among AHRQ's tools: free or at cost
Icons indicating whether a dataset is free or not


Don't use MEPS IC much
Find out the common variables across all the tools
Even if the samples are different between tools I want to be able to compare data by common variables
MEPS is collecting data from the actual user. and HCUPnet has the data from the hospital.
Researchers often ask: Can I compare HCUP with MEPS data?" You can't because it's a different sample, but having the ability to compare by variable enables validation of certain analyses that may be useful
EG researching MEPS to find conditions reported by a certain demographic group, and then view hospital release data for that same demographic group will allow me to validate the data even though the samples are different
She suggests we find the common variables among all the tools
Pro vs Intermediate Users: exploratory research vs accuracy / ntl representation. Former can compare different data sets with common variables?
Receiving User Queries / Technical Assistance Requests
The goal is for nearly 100% of our media inquiries to be for MEPS or HCUP, bc even if the question is about something else (eg patient safety, Primary Care data), HCUP or MEPS will have some data that can help
In answering media inquiries, we rarely respond with 2 or more recommended data tools to answer that query
If media users request out of pocket expenditures, we send them to MEPS; if looking for counts on inpatient surgery procedures or diagnoses, HCUP
Rare example of recommending 2 tools in response to a query: New York Times had a specific CAHPS question that then led to summoning more information about HCUP data (Diagnostic Imaging testing)
Most TARs can be answered with 1 tool, though multi-dataset queries do happen
I wouldn't recommend trying to provide the answer to the reporter. We try to teach them to fish, learn how our databases work, but often they just need an answer ASAP
I very rarely go to the data tables
HCUP and MEPS make up about 75% of the inquiries that come into the media relations team directly
Almost all media queries can be addressed with just 1 tool
Use Case 1
Reporter from the Atlantic reaches out who wants to know the out-of-pocket expenditures in 2016 for Pediatric Dental Care. I'd then contact Director and Asst. Dir. of CFACT: "got this inquiry, can one of your folks run the numbers?"
["Special run" request button featured on the query tool pages in the event that the tool can't answer a question)]
Query Referral Process
Query referral process: AHRQ public line (call or email) > all media referred to me > I enlist help of tool POCs. Entire query process tracked & logged.
On MEPS there are user portals that allow you to submit a special query request
[Retain MEPS' user-specific TAR support pages to be efficiently directed to the right people (researcher, media, policy)]
HMW support efficient routing of TARs to the right people?


1. Knowing Where to Start
Of all AHRQ's tools, the only two real forward public-facing components to the general public would be MEPS and HCup
MEPS are the gold standard – both for researchers and policymakers – largely due to its remarkable sample size (HC: 40,000 people; IC: 34,000 private employers)
75% of our media inquiries related to AHRQ data resources are about HCup or MEPS
It'd be great to have simple portal: click here if you want to what people pay for their Healthcare; Click here for info on hospitals; etc
[on Portal, title each tool by what it offers from a research perspective – instead of strictly by the name of tool]
It would be really nice to just kind of send [the media] a sixth grade level common description of what you can expect to do with these [tools]
Given the existing cache and common understanding of MEPS and HCUP among researchers, its possible to gear the descriptive content of both tools toward novice users (and still not compromise pro researchers' engagement)
As reporters increasingly ask about the various topics covered by the data tools, its difficult to explain (via email or phone) of what they can get through the resources.
The tools have broad descriptions of their purpose, but not how that data can be applied (eg: what is a family paying out of pocket?)
[Portal Info hierarchy: laymen's terms description of the tool's use case, name of the tool, high-level researcher information, applications / sample questions & topics that tool can cover, then detail info for researchers]
Teaching Newbies (via email): This is what you can get from H Cup: any procedure, diagnosis in a hospital with counts for all the reasons why people go to the ED; the cost of the hospital; what they charge the insurance companies; we can break it down by demographics sex age etc. Or MEPS: you want to know what people are paying for their health care out of pocket; what Private employees versus Public Employees are paying for their health insurance; drug costs at inpatient facilities.
HCUP and MEPS make up about 75% of the inquiries that come into the media relations team directly
User Types (Users)
75% of our media inquiries related to AHRQ data resources are about HCup or MEPS
MEPS are the gold standard – both for researchers and policymakers – largely due to its remarkable sample size (HC: 40,000 people; IC: 34,000 private employers)
Receive most inquiries for HCUP, then MEPS HC, then IC
Kaiser Health News (kff.org) is our main competitor among media users (partly bc they don't have the privacy restrictions of a federal agency)
Users range from the frequent flyers (EG AHCJ, Association of Health Care Journalists) to newbie State journalists reporting on Medicaid bill before their state legislature who referred by one of our grantees at a local university
HCUP and MEPS make up about 75% of the inquiries that come into the media relations team directly
Context of Use / Behavior (User)
Most folks are probably submitting their requests on their phones
Nine to one email to phone-based requests
Most of the time, media users have 20 stories going and don't have time to learn -- they need an answer. So I don't direct them to MEPS, I provide them the answer of the data output (and we'll internally use MEPS to do that)
A minority of users are Healthcare journalist with data science background who are responsible for larger, in-depth going-for-the-Pulitzer type stories. They want a walkthrough of the tools and understand how granular can they get
Most used tools for media: the majority of the public is using HCUP, followed by MEPS HC, IC,
I wouldn't recommend trying to provide the answer to the reporter. We try to teach them to fish, learn how our databases work, but often they just need an answer ASAP
HCUP and MEPS make up about 75% of the inquiries that come into the media relations team directly


Integrating Multiple Tools in Workflow
From the day I got here, it'd really be nice if there was a way that these [tools] could talk to each other. But as I've gotten to be here a little bit longer and gotten to know the people who work on these programs, they're not really designed in a way to be cross-referenced
Rare example of recommending 2 tools in response to a query: New York Times had a specific CAHPS question that then led to summoning more information about HCUP data (Diagnostic Imaging testing)
Rare example of recommending 2 tools in response to a query: New York Times had a specific CAHPS question that then led to summoning more information about HCUP data (Diagnostic Imaging testing)
Navigating with Ease
Simpler description, less text and instead more click-through for heavier description
Visit the AHRQ Views blog for summations of each of the tools in plain language
0. Impetus for a Query
Participants reporting using QDR products for informational and research purposes, including presentations, publications, grant applications, and teaching
Participants reporting using QDR products for influential purposes to highlight gaps in care to advocate for resources and action
Participants reporting using QDR products for health care improvement activities.
1. Knowing Where to Start
Recommended to carefully document the sources, exclusion criteria, and potential limitations of the data in footnotes or appendices.
Recommended to develop easier search functions to facilitate the process of searching for specific data or data extraction.
Recommended to make searchable formats of data available.
Develop one or multiple Web sites with similar look and feel to make tools seem more integrated.
Link to other relevant AHRQ tools, including the Innovations Exchange, health literacy tools, quality indicators, and the Patient Safety Network.
Navigating with Ease
Provide additional visual and graphic displays, including infographics.
Improve clarity of messages through bullets, call-out boxes, and providing more impact statements.
Link to case studies and stories and provide key details about implementation.
Consider updating QDR Web site to include more user-friendly features, pictures, and customizable data tools.

Navigating with Ease

Would like to see IC emulate UI of MEPS ST HC

Census' Fast Stats ("American Fact Finder") is a best practice as it uses a step-by-step query process in one page (see image), allowing you to retrace your steps

Best practices in Census American Fact finder is quick actions on any page: Modify Table, Add/Remove Geographies, Bookmark/Save, Print, Download, Create a map

Most people want to see trends

More unsophisticated users like reporters etc want to see descriptive analyses

[He scrolls a page using data visualizations as an indicator of the info/tables he's seeking on Kaiser]

[Visualizations of data act as quick-reference signifiers / reminders of where more in-depth tabulated data is]

ST IC

I like that MEPS ST HC is grouped into topics. But it needs instructions on where to start (query)

MEPS ST HC critique: Some queries might fall between the topic areas (eg how do you get cross-tab information, eg population and prescribed drugs)?

Allow users to compare two tables side by side

Main competitor for reporter users: Kaiser survey

[Sidebar of summary table sections a high-level search through all section (see Kaiser Employer Health Benefits Survey image)]

[Small button for quick reference to methods of survey]

[Sidebar of summary table sections a high-level search through all section (see Kaiser Employer Health Benefits Survey image)]

KAISER EHBS: Click on report > click on deductibles section > scroll down to the table I need > find percentage of workers covered by an HSA plan > download this into a PowerPoint

[Ability to download summary tables into multiple formats: PDF or Powerpoint]

Mercer is another example

5. Reviewing Data / Query Results

Aspects of the data need to be more user friendly to external users

Use MEPSnet IC tools if I want to know a time trend, I want a nice graph – it's not nice enough, but it has a decent graph and shows me all of the numbers for all years in one place.

Use MEPS IC tools time trend graphs not nice enough

Can't do time trends in MEPS ST IC - just single years

Want to be able to download/print a line graph of trends over a # of years and have everything come out on one single page (for quick reference or distribution)

[all query results have a high-level, visual synopsis at top (eg line graph), table below with #'s and standard errors, which you can download / print / share as 1-pager]

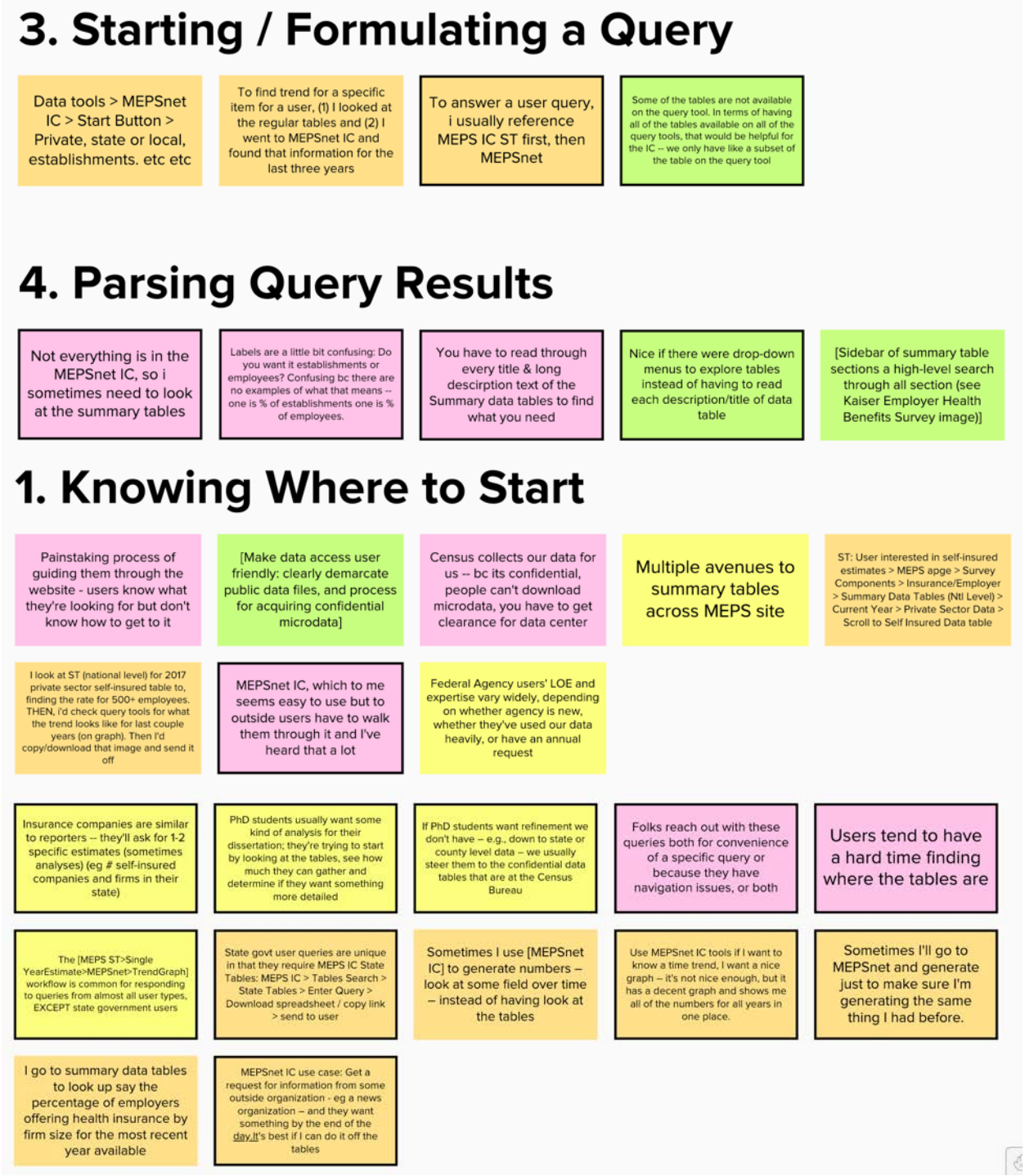
MEPS ST HC code section wouldn't be applicable to MEPS IC ST, since microdata isn't publicly available

To answer a query, she first uses MEPS ST to quickly reference a single-year estimate with granular characteristics, then conducts same query in MEPSnet to see a trend graph over time, which she copy-pastes in response to TAR

Need both granular detail for single-year estimates (currently, ST) as well as the ability to see multi-year trends w/ data visualization (currently, MEPSnet)

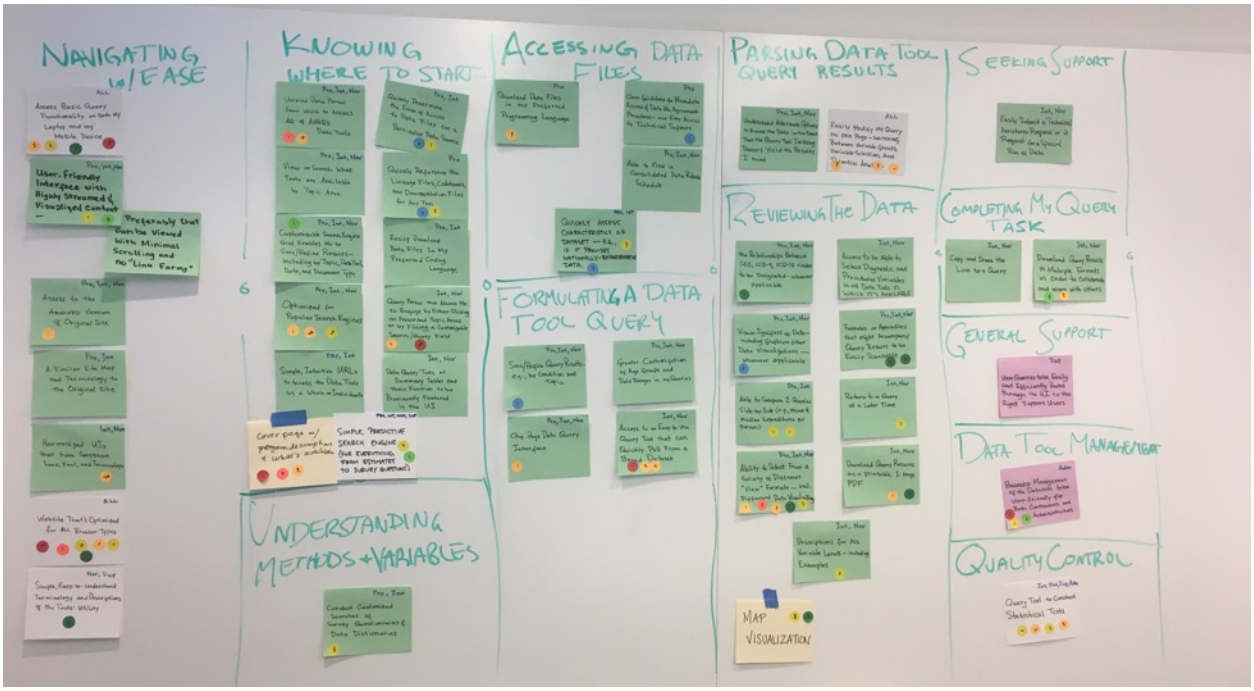
Pull up trends on MEPSnet and see all firm sizes on one chart (instead of only being able to view size groupings 1 at a time) with one table below it

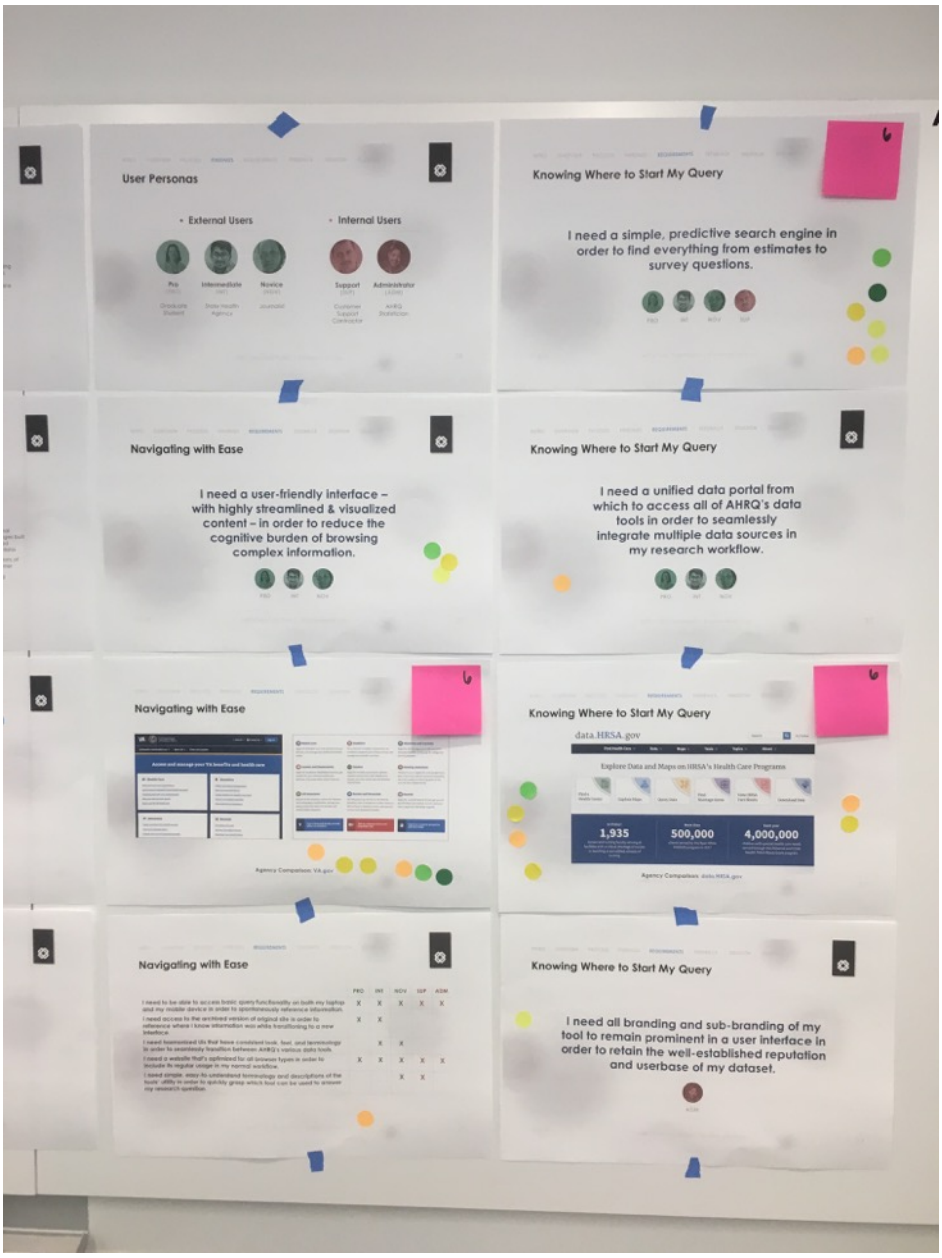
sometimes I'll use the I'll go back in time to make sure I'm using the right table just as a check on myself

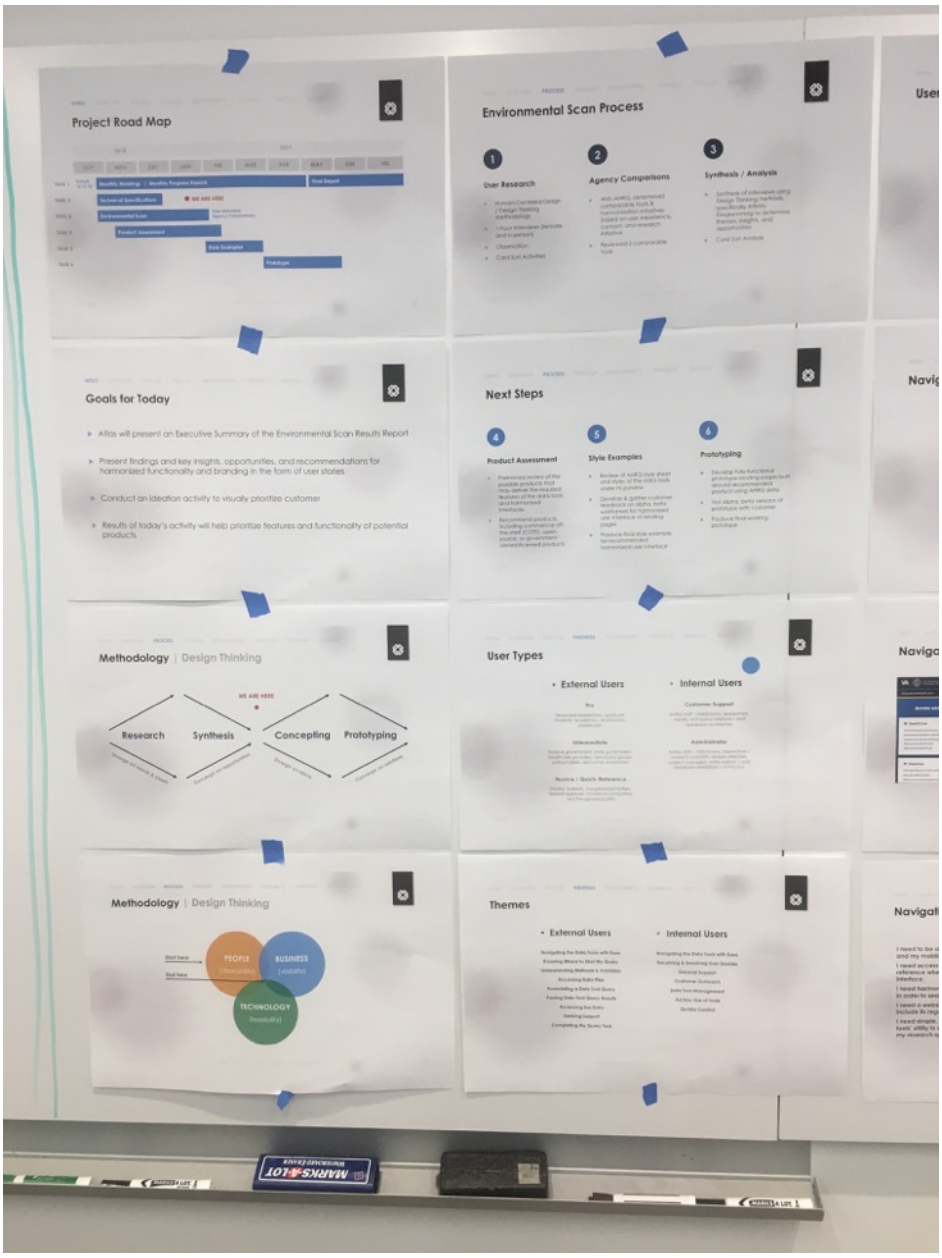

3. Starting / Formulating a Query
Data tools > MEPSnet IC > Start Button > Private, state or local, establishments. etc etc
To find trend for a specific item for a user, (1) I looked at the regular tables and (2) I went to MEPSnet IC and found that information for the last three years
To answer a user query, i usually reference MEPS IC ST first, then MEPSnet
Some of the tables are not available on the query tool. In terms of having all of the tables available on all of the query tools, that would be helpful for the IC -- we only have like a subset of the table on the query tool
4. Parsing Query Results
Not everything is in the MEPSnet IC, so i sometimes need to look at the summary tables
Labels are a little bit confusing: Do you want it establishments or employees? Confusing bc there are no examples of what that means -- one is % of establishments one is % of employees.
You have to read through every title & long descirption text of the Summary data tables to find what you need
Nice if there were drop-down menus to explore tables instead of having to read each description/title of data table
[Sidebar of summary table sections a high-level search through all section (see Kaiser Employer Health Benefits Survey image)]
1. Knowing Where to Start
Painstaking process of guiding them through the website - users know what they're looking for but don't know how to get to it
[Make data access user friendly: clearly demarcate public data files, and process for acquiring confidential microdata]
Census collects our data for us -- bc its confidential, people can't download microdata, you have to get clearance for data center
Multiple avenues to summary tables across MEPS site
ST: User interested in self-insured estimates > MEPS apge > Survey Components > Insurance/Employer > Summary Data Tables (Ntl Level) > Current Year > Private Sector Data > Scroll to Self Insured Data table
I look at ST (national level) for 2017 private sector self-insured table to, finding the rate for 500+ employees. THEN, I'd check query tools for what the trend looks like for last couple years (on graph). Then I'd copy/download that image and send it off
MEPSnet IC, which to me seems easy to use but to outside users have to walk them through it and I've heard that a lot
Federal Agency users' LOE and expertise vary widely, depending on whether agency is new, whether they've used our data heavily, or have an annual request
Insurance companies are similar to reporters -- they'll ask for 1-2 specific estimates (sometimes analyses) (eg # self-insured companies and firms in their state)
PhD students usually want some kind of analysis for their dissertation; they're trying to start by looking at the tables, see how much they can gather and determine if they want something more detailed
If PhD students want refinement we don't have – e.g., down to state or county level data – we usually steer them to the confidential data tables that are at the Census Bureau
Folks reach out with these queries both for convenience of a specific query or because they have navigation issues, or both
Users tend to have a hard time finding where the tables are
The [MEPS ST>Single YearEstimate>MEPSnet>TrendGraph] workflow is common for responding to queries from almost all user types, EXCEPT state government users
State govt user queries are unique in that they require MEPS IC State Tables: MEPS IC > Tables Search > State Tables > Enter Query > Download spreadsheet / copy link > send to user
Sometimes I use [MEPSnet IC] to generate numbers – look at some field over time – instead of having look at the tables
Use MEPSnet IC tools if I want to know a time trend, I want a nice graph – it's not nice enough, but it has a decent graph and shows me all of the numbers for all years in one place.
Sometimes I'll go to MEPSnet and generate just to make sure I'm generating the same thing I had before.
I go to summary data tables to look up say the percentage of employers offering health insurance by firm size for the most recent year available
MEPSnet IC use case: Get a request for information from some outside organization - eg a news organization – and they want something by the end of the day.It's best if I can do it off the tables

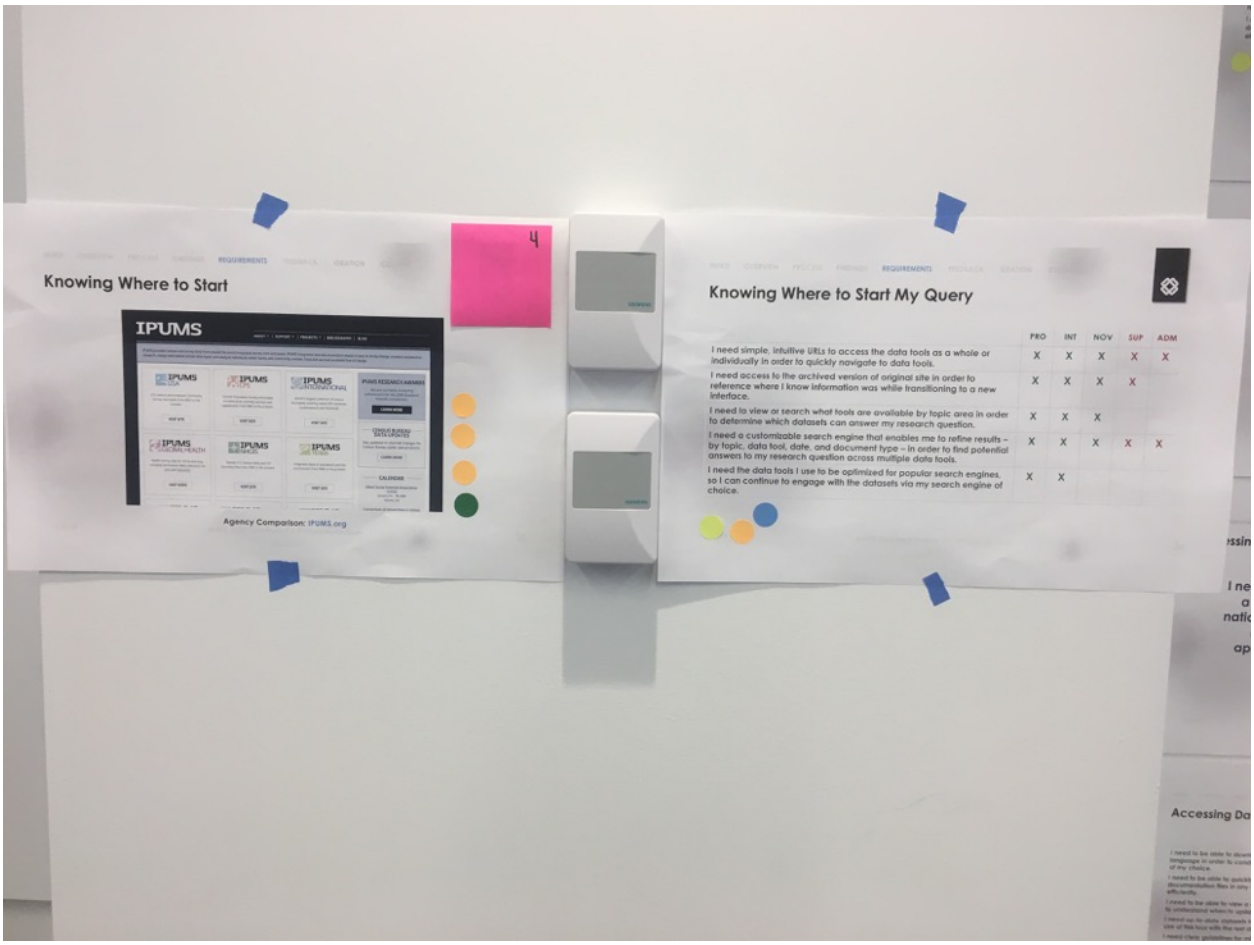

Context of Use / Behavior
Typically refer to MEPSnet IC first before MEPS ST IC because MEPSnet is easier to use
MEPS IC Chart Books - which i helped design - offer user-friendly way for users to see trends over time
Use MEPSnet IC tools if I want to know a time trend, I want a nice graph – it's not nice enough, but it has a decent graph and shows me all of the numbers for all years in one place.
The household data i download the raw data / download the SAS code they provide
Main competitor for reporter users: Kaiser survey
Kaiser does an annual survey of employers. It's it's inferior compared to MEPS (2000 vs 40000 sample)
Users often call the general number
Users usually not that sophisticated because otherwise they'd figure it out
Journalists usually want 1-2 estimates, (eg # establishments in Minnesota). Usually nothing complex.
Only some users are familiar with what's available on the tools
Insurance companies are similar to reporters -- they'll ask for 1-2 specific estimates (sometimes analyses) (eg # self-insured companies and firms in their state)
PhD students usually want some kind of analysis for their dissertation; they're trying to start by looking at the tables, see how much they can gather and determine if they want something more detailed
If PhD students want refinement we don't have – e.g., down to state or county level data – we usually steer them to the confidential data tables that are at the Census Bureau
I answer queries half a dozen to a dozen times a year, more frequent after I put out a report using the MEPS data from reporters
3. Starting / Formulating a Query
User was looking to write about hypertension, and was looking for the usage of different drugs to treat hypertension over the years
Typical TAR request: user doing research on ER Admissions and the cost of those admissions whether it be for a specific condition or just the average
It would be helpful, tho not necessary, to have more customizable age groups on MEPS ST HC (instead of just age ranges)
Some users mights find MEPS ST user friendly, requiring little guidance - but it doesn't give them everything they need
MEPS ST: I grab the tables for total expenditures by therapeutic class across a certain year span, then copy that link and send it to the user
Sometimes i'll go to the actual data files to search for a particular variable that's not in the MEPS ST, then send to user and tell them how to use it
Responding to more complicated queries, I can calculate it separately and do my own programming or download the data and run it. That data is not on the tool [MEPS ST] because there are so many different combinations that people may be looking, and we don't have any tool that allows them to quickly calculate those numbers
Ability to customize MEPS HC ST searches
You can't combine certain data (e.g. people 40+ diabetes, blood pressure what are the expenditures per year) and MEPSnet doesn't have conditions
4. Parsing Query Results
It's hard to get users to read the descriptive analysis; I'll have to paste it into an email with the link
[Better info hierarchy]
1. Knowing Where to Start
Most research scenarios for my use case are regarding total expenditures or medical condition expenditures
Complicated Queries
Not all questions are about the actual numbers - some questions are about analyses
Some users go to the papers we write to see a summary/analysis of the data we have
People who ask for more complicated stuff related to their research - eg people who have diabetes and blood pressure -
~50% of queries are unique/complicated – users have a specific project ie they're not looking

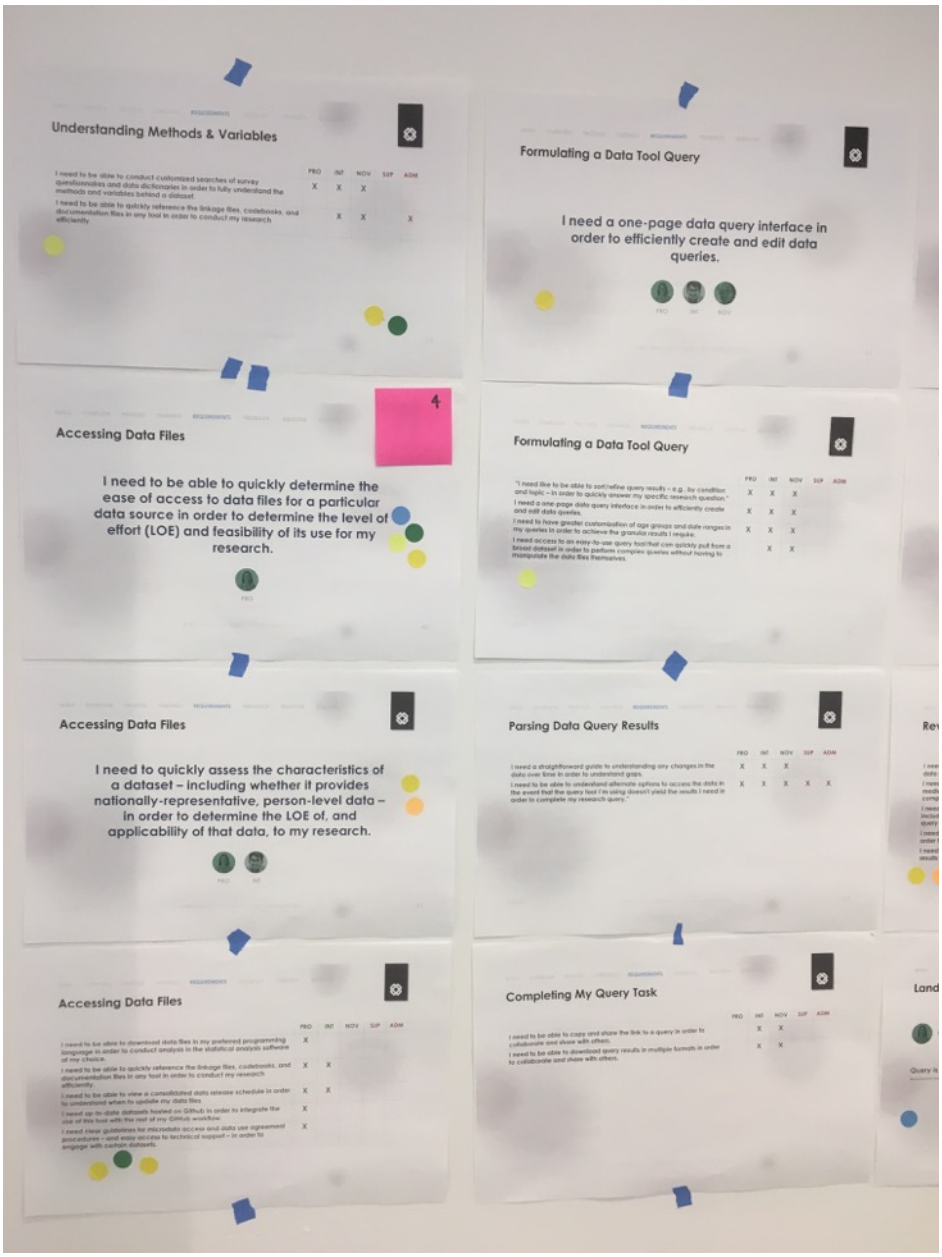
C. MID-PHASE WORKSHOP PHOTOS



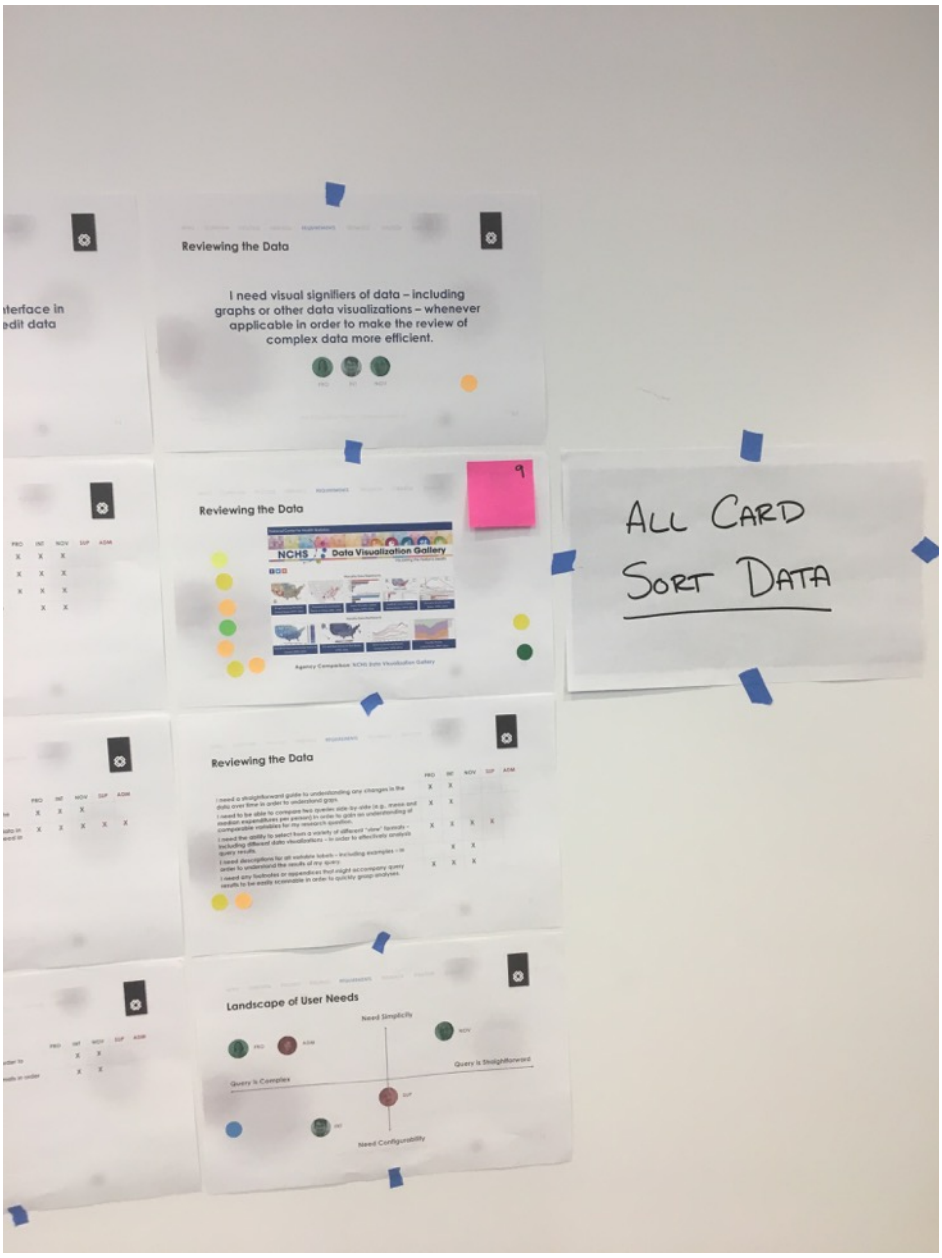

User Personas
External Users
Internal Users
Knowing Where to Start My Query
I need a simple, predictive search engine in order to find everything from estimates to survey questions.
Navigating with Ease
I need a user-friendly interface – with highly streamlined & visualized content – in order to reduce the cognitive burden of browsing complex information.
Knowing Where to Start My Query
I need a unified data portal from which to access all of AHRQ's data tools in order to seamlessly integrate multiple data sources in my research workflow.
Navigating with Ease
Knowing Where to Start My Query
data.HRSA.gov
1,935
500,000
4,000,000
Navigating with Ease
Knowing Where to Start My Query
I need all branding and sub-branding of my tool to remain prominent in a user interface in order to retain the well-established reputation and userbase of my dataset.


Project Road Map
Environmental Scan Process
User Research
Agency Comparisons
Synthesis / Analysis
Goals for Today
Atlas will present an Executive Summary of the Environmental Scan Results Report
Present findings and key insights, opportunities, and recommendations for harmonized functionality and branding in the form of user stories
Conduct an ideation activity to visually prioritize customer
Results of today's activity will help prioritize features and functionality of potential products
Next Steps
Product Assessment
Style Examples
Prototyping
Methodology | Design Thinking
Research
Synthesis
Concepting
Prototyping
User Types
External Users
Internal Users
Methodology | Design Thinking
PEOPLE
BUSINESS
TECHNOLOGY
Themes
External Users
Internal Users


Knowing Where to Start
IPUMS
Agency Comparison: IPUMS.org
Knowing Where to Start My Query
PRO INT NOV SUP ADM
I need simple, intuitive URLs to access the data tools as a whole or individually in order to quickly navigate to data tools.
I need access to the archived version of original site in order to reference where I know information was while transitioning to a new interface.
I need to view or search what tools are available by topic area in order to determine which datasets can answer my research question.
I need a customizable search engine that enables me to refine results – by topic, data tool, date, and document type – in order to find potential answers to my research question across multiple data tools.
I need the data tools I use to be optimized for popular search engines, so I can continue to engage with the datasets via my search engine of choice.



(photo removed)


Reviewing the Data
I need visual signifiers of data – including graphs or other data visualizations – whenever applicable in order to make the review of complex data more efficient.
Reviewing the Data
NCHS Data Visualization Gallery
ALL CARD SORT DATA
Reviewing the Data
Landscape of User Needs
Need Simplicity
Query is Complex
Query is Straightforward
Need Configurability

(photo removed)

RFTO #10: Developing Strategies for the Harmonization of AHRQ Data Tools (Data Tools)

Product Assessment Report

February 8, 2019

CONTENTS

REVISION HISTORY

Date	Version	Description	Author(s)
January 25, 2019	Draft	Initial draft prepared for AHRQ to review and provide feedback.	Atlas Team
February 1, 2019	AHRQ Review	AHRQ review, corrections, and feedback	Agency for Healthcare Research & Quality
February 8, 2019	Final	AHRQ feedback incorporated into the document	Atlas Team

1. EXECUTIVE SUMMARY

The Agency for Healthcare Research and Quality (AHRQ) provides several data tools and summary data tables that have dedicated user groups, but currently these tools are siloed and lack a common user experience (UX) and branding among them.

To better understand both the potential solutions available to support the harmonization of the AHRQ digital tools user interfaces (UI) and basic query functionality of these data tools, the Atlas Team conducted an assessment of commercial off-the-shelf (COTS) Data Visualization Products. The Atlas Team conferred with the AHRQ team concerning features and functionalities to consider as part of the assessment.

The Atlas Team, with guidance from the AHRQ team narrowed down the number of Data Visualization Products to assess based on priority:

High Priority	Low Priority
MicroStrategy	Birst
Qlik	Domo
Shiny	Looker
Tableau	Microsoft PowerBI
	Sisense

Given the general growth in cloud computing and data analysis tools in recent years, the external user experiences, or frontend design, tends to be similar across all Data Visualization Products. However, the Data Visualization Products tend to differ greatly when it comes to the internal UX, or backend design. These differences occur mostly with data deployment, data configuring and data visualization. These differences lead the Atlas team to recommend four (4) products based on AHRQ's internal and external users' priorities: MicroStrategy, Qlik, Shiny, and Tableau.

Product	Strengths	Weaknesses	Costs Estimates
MicroStrategy	• High touch implementation team when setting up MicroStrategy. • Advanced visualization engine available to backend users. • Ability to scale across all AHRQ products.	• Approximately 2x more expensive than current competitors' estimates.	**If Perpetual License** • Base Cost: $350,000 • Annual Maintenance Cost: $58,000 • Average Annual Cost for a 5-year period: $116,400 **If Term License** • Annual Cost: $150,000 MicroStrategy is willing to work with AHRQ team to accommodate budgets.
Shiny	• Approximately half the cost of Tableau, and 80% less than MicroStrategy.	• Steep learning curve for backend users with no R / Shiny experience.	**Total Annual Cost**: $38,300 • RStudio Shiny Server Pro: $10,000

	• Entirely customizable by the AHRQ team. • Lowest impact on AHRQ's existing environment.	• Architecture impacts ability to easily scale across all AHRQ data tools.	• RStudio Connect for 20 Users: $15,000 • RStudio Package manager: $10,000 • Shiny Apps Professional: $3,300
Tableau	• Product architecture can scale across all AHRQ products. • Intuitive graphics for creating visualizations.	• Reliance on self-service learning appeals to a limited set of administrators.	**Total Annual Cost**: $72,000
Qlik	• Experience with Federal clients. • Compelling frontend graphics.	• Less intuitive backend user experience.	**Total Annual Cost:** $80,300

After AHRQ has selected two products from the above recommendations, the Atlas Team intends to create prototypes to test the efficacy of each. These prototypes will take the form of two consecutive versions:

- **Alpha**: The AHRQ team will provide the Atlas Team 2-3 AHRQ datasets for each of the 2 products selected to build 4-6 prototypes. The AHRQ team will use the 4-6 prototypes to inform a decision on selecting a finalist data product.
- **Beta**: After selecting a final product, the Atlas team will develop a prototype for all 8 AHRQ tools and a sample dataset, totaling in 8 prototypes.
- **Final**: The final stage prototype will incorporate feedback for all 8 AHRQ tools and their sample datasets. The prototype code will be submitted to AHRQ for internal consideration.

1.1 Product Assessment at a Glance

The assessment of the Data Visualization Products did not differ by whether they were considered High or Low Priority. However, High Priority Data Visualization Products have a more detailed analysis in this report. Assessments for High Priority Data Visualization Products include:

- **A Product Overview**: A high level summary of the product offerings, including a table outlining strengths and weaknesses for quick digestion by the reader.
- **Hypothetical Product Implementation**: An estimated implementation for the product based on Atlas' research thus far to provide the reader with an understanding of the level of effort required to implement the product.
- **Federal Examples of the Product**: Easily accessible links, when available, to examples of the product being used in the Federal space.
- **Final Takeaways for the Product**: Concluding summary of the product's strengths and weaknesses as it relates to AHRQ's priorities to improve UIs, reduce costs, and simplify management.

Assessments of the Low Priority Data Visualization Products include a simple summary of the products' strengths and weaknesses along with the Atlas Team's final recommendations.

2. PRODUCT ASSESSMENT APPROACH

AHRQ requires a harmonized UI along with a data visualization and analysis COTS Data Visualization Product that meets the overlapping needs of backend and frontend users for its various data query tools.

To determine the appropriate COTS Data Visualization Product, the Atlas Team cast a wide net of potential solutions. These solutions were narrowed down to a list of 9 Data Visualization Products that were divided into 4 High Priority products and 5 Low Priority products. To understand the full capability of each tool, and its ability to satisfy AHRQ's unique needs, the Atlas Team conducted primary and secondary research.

The primary research centered around demonstrations of the Data Visualization Products with company representatives, followed by structured questions from a discussion guide to standardize the topics covered in each product demonstration; these discussions included a request for any Federal examples of the product in use, which are provided in this report. In a limited number of circumstances (e.g. Qlik and Domo), the Atlas Team was able to secure a demo version of the product for follow up research.

Part of the primary research with the High Priority products also involved fielding estimates for the cost of implementing the products. The standard approach the Atlas team took was asking for estimates based on embedding the tool in a public facing website that sees an average of 75,000 hits per month, and backend access for 10 super users as common baseline to all vendors. We used "75,000 users" in our requests for estimates based on the average of the estimated usages provided by the AHRQ team for the data tools.

All product teams provided us with estimates and stressed that these were ballpark estimates based on the limited information they have. Moreover, many of them expressed a willingness to 1) undergo capacity planning with the AHRQ team to determine the correct amount of infrastructure for the product to work and 2) work with the AHRQ team based on their budgetary requirements.

The Atlas Team also conducted secondary research involving online research to understand any potential pitfalls for implementing a Data Visualization Product. When possible, the Atlas Team utilized the same source of knowledge to ensure consistency in perspective, such as Gartner's "Critical Capabilities for Analytics and Business Intelligence Platforms". This provides the Atlas Team with sufficient knowledge of the space to give AHRQ the recommendation to implement either MicroStrategy, Shiny, or Tableau, based on AHRQ's internal priorities for the product.

To provide the reader with a fundamental understanding of how a large number of Data Visualization Products tend to work, an explainer is available in the next section (2.1). However, if the reader is already familiar with this technology, please continue on to the Incorporating Environmental Scan Results section (2.2).

2.1 Data Visualization Products' Approach Overview

The Data Visualization Products that we surveyed employ similar strategies to visualize data. The Software as a service (SaaS) model employed by a majority of products included in our research extract data from databases using standard connectors such as a connector to a Structured Query Language (SQL) database to connect to existing databases. In the rare

instances a database has a server or characteristics that prevent it from being connected with standard connectors, the Data Visualization Product team will use a more customizable option such as a Java Database Connector (JDBC) or Open Database Connectivity (ODBC) connector. See below for a visualization of a connector joining a client database to a cloud Data Visualization Product solution.

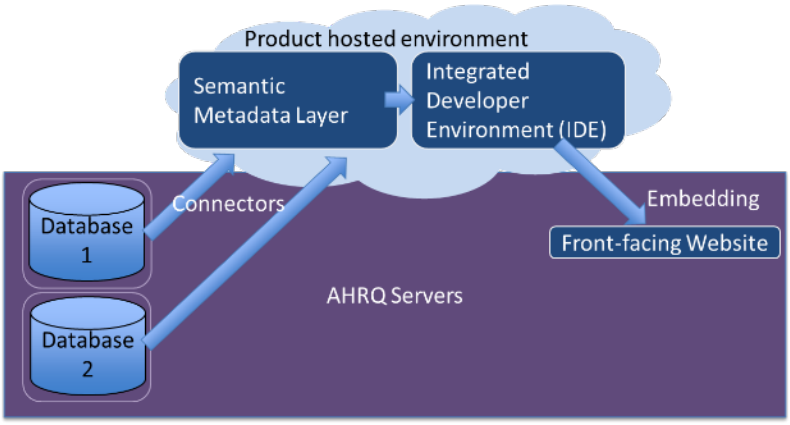


Figure 1. Visualization of SaaS strategy of connecting AHRQ's data to a SaaS provided semantic metadata layer for data modeling and then visualizing using an Integrated Development Environment (IDE). The visualizations are then embedded into AHRQ's front-facing website

Once the database is connected, the data is fed into a semantic metadata layer stored on the cloud. Storing on the cloud allows for much easier scalability for these Data Visualization Products, as more working memory and data storage can be added easily.

Only after a client's data is in the cloud database do the Data Visualization Products begin to differentiate from each other. The data is fed into the Data Visualization Product's software where visualizations and filtering can be quickly and easily applied giving the frontend user the feel of real time access to the client's data through the lens of a Data Visualization Product allowing them to create customized visualizations.

Each solution offers an approach of either using drag-and-drop to visualize how data is connected into the visualization sandbox. However, for savvier backend users, options are provided to incorporate SQL code to manipulate and join these fields.

Once a visualization is ready to go, the client team agrees on a schedule for updating the data. Data Visualization Product companies are able to accommodate very frequent data extractions into their tools, which is useful for quick fixes or recalls to incorrect data uploaded into tables. Nevertheless, all Data Visualization Products offer robust options for uploading data that far exceed AHRQ's needs. Moreover, publishing the data out into a public-facing environment is very simple for all products. All products recommend embedding the public version of the visualization through an iframe; which, in its simplest form, results in an embedded version of the product in a webpage. This allows the data to be uploaded as often as needed with minimal adjustments to the client's web servers. When a more robust connection is needed, most products offer an involved Java Script Application Programming Interface (API) or a Representational State Transfer (REST) API connection. The Atlas Team asked the representatives from the Data Visualization products if they had seen any impacts to 508 compliance when using iFrames to embed a product in a public facing site. The product

representatives have not encountered this issue but recommended embedding the tool in a site using Java Script API to minimize any risk in impacting 508 compliance.

2.2 Incorporating Environmental Scan Results

To ensure that the proposed products satisfy backend and frontend user requirements uncovered in the Environmental Scan, the Atlas Team selected those user requirements that are relevant to the Data Visualization Products' scope and those that are relevant to the site configuration. Below is a list of these user requirements that are considered requirements for the Data Visualization Products. Based on this assessment, these product requirements can be accommodated by the High Priority Products.

1.1.5. Formulating & Refining a Data Tool Query						Site Requirement	Product Requirement
User Story	**Pro**	**Int**	**Nov**	**Sup**	**Adm**		
"I need to be able to sort/refine query results – e.g., by condition and topic – in order to quickly answer my specific research question."	X	X	X				●
"I need a one-page data query interface in order to efficiently create and edit data queries."	X	X	X			●	●
"I need to have greater customization of age groups and date ranges in my queries in order to achieve the granular results I require."	X	X	X				●
"I need access to an easy-to-use query tool that can quickly pull from a broad dataset in order to perform complex queries without having to manipulate the data files themselves."		X	X				●
"I need to easily modify my query on one page – switching between variable groups, variable selection, and potential analyses – in order to answer a research question efficiently."	X	X	X	X	X		●
1.1.6. Reviewing the Data							
User Story	**Pro**	**Int**	**Nov**	**Sup**	**Adm**		
"I need visual signifiers of data – including graphs or other data visualizations – whenever applicable in order to make the review of complex data more efficient."	X	X	X				●
"I need to be able to compare two queries side-by-side (e.g., mean and median expenditures per person) in order to gain an understanding of comparable variables for my research question."	X	X				●	●
"I need the ability to select from a variety of different "view" formats – including different data visualizations – in order to effectively analysis query results."	X	X	X				●
"I need access to be able to select diagnostic and procedural variables in all the data tools in which its available."		X	X				●
"I need to be able to return to a query at a later time in order to complete my research unabated."		X	X			●	●
"I need to be able to download query results as a printable 1-page PDF in order to distribute salient data with others or reference it in the future."		X	X				●
1.1.8. Completing My Query Task							
User Story	**Pro**	**Int**	**Nov**	**Sup**	**Adm**		
"I need to be able to copy and share the link to a query in order to collaborate and share with others."		X	X			●	●
"I need to be able to download query results in multiple formats in order to collaborate and share with others."		X	X				●
1.1.9. Data Tool Management							
User Story	**Pro**	**Int**	**Nov**	**Sup**	**Adm**		
"I need backend management of the datasets to be user-friendly for both contractors and administrators in order to ensure that quality control efforts are carried out seamlessly every time the datasets are updated."					X	●	●
1.1.10. Quality Control							
User Story	**Pro**	**Int**	**Nov**	**Sup**	**Adm**		
"I need the query tool to be able to conduct statistical tests, so I can easily determine what's statistically significant for my research question."		X	X	X	X		●

3. HIGH PRIORITY PRODUCT ASSESSMENTS

Each High Priority Data Visualization Product assessment that follows includes:

- **A Product Overview**: A high-level summary of the product's offerings including a strengths and weaknesses table for quick digestion.
- **Hypothetical Product Implementation**: An estimated implementation for the product based on Atlas' research thus far to provide an understanding of the level of effort required in implementing the product.
- **Federal Examples of the Product**: Easily accessible links to examples of the product being used in the Federal space, when applicable. For public-facing sites, screenshots of their sample dashboards are provided.
- **Final Takeaways for the Product**: Concluding summary of the product's strengths and weaknesses with a final explanation of Atlas' recommendations.

3.1 MicroStrategy

3.1.1 MicroStrategy Overview

Strengths	Weaknesses
Established brand and company.	Estimated to cost twice as much to implement than Tableau.
Open platform for customized solutions.	Involved initial data model build out.
Supportive implementation staff.	
Innovative customizations built into the tool to support backend users.	
Experienced with Federal clients.	

MicroStrategy is a US commercial BI software that has over 20 years of experience in developing visualization and reporting tools for their clients. They are one of the older BI companies and are often regarded as pioneers in the space. This knowledge gives MicroStrategy a deeper well of insight compared to upstart tools, such as Looker or Sisense, making them a potentially stronger partner for AHRQ to collaborate with.

MicroStrategy can be deployed in different ways to display data. MicroStrategy offers COTS visualizations and templates but there are options for significant customization enhancements through MicroStrategy's open architecture. Some of the options in this architecture include:

- REST API: for bidirectional integration into custom web applications (apps)
- PUSH API: for sending data developed in MicroStrategy to 3rd Party vendors
- Data Connector Software Development Kit (SDK) to connect to non-traditional data sources
- Web services SDK for creating customized web apps
- Mobile SDK for creating customized mobile designs
- Visualization SDK for creating customized visualizations

- Embedding API for embedding data visualizations into a website
- Cloud API for extended cloud deployment

This open approach allows customers to use MicroStrategy both as a COTS product, but also to integrate with custom developed solutions. Moreover, the expansive offering allows for scaling up MicroStrategy throughout AHRQ's entire existing ecosystem while accommodating the varying needs of each digital tool.

The current version for MicroStrategy is Version 2019, which was released on 12/31/19. As of 2018, MicroStrategy adopted a policy of annual releases for Platform Releases that focus on production level security, stability and performance. These releases are applicable to all MicroStrategy customers. During the year, MicroStrategy has quarterly feature releases for customers with specific feature requirements.

3.1.2 Costs for Implementing MicroStrategy

A significant consideration for MicroStrategy is the cost. MicroStrategy offers two pricing models depending on AHRQ's needs: (1) Perpetual License or Term Licenses. A Perpetual License gives AHRQ the opportunity to own their license of MicroStrategy and pay annual maintenance for the product, such as for Standard Technical Support or for new versions of the software. Alternatively, the (2) Term License model involves leasing the software for use. In this model, AHRQ will pay for a lease of the software and for the support that accommodates the platform. If the AHRQ team intends to keep the software for more than a few years, it would make economic sense to pursue the Perpetual License.

3.1.3 Hypothetical MicroStrategy Implementation

An AHRQ implementation with MicroStrategy would first involve deciding to implement on-premises or on the cloud—MicroStrategy accommodates both styles of implementation. AHRQ's databases and the tables within each need to be uploaded into MicroStrategy's semantic layer so that the data can interact with MicroStrategy's applications. To connect the data to MicroStrategy, there are a variety of connectors that can be applied, ranging from standard connectors such as Oracle or SQL servers, to the less common such as Survey Monkey, which isn't a relational database. Building up the initial data model is an involved process for backend users, however the model allows administrators to create additional visualizations without having to undergo substantial revisions to the data model. Moreover, security layers can be provided for different frontend user types at the data model level, giving administrators the ability to determine which data frontend users are able to view.

Once a database is ready to load into the MicroStrategy semantic layer, an administrator can join data by using a graphical user interface (GUI) to drag and drop fields from different tables, by using SQL to type a query, or by selecting the tables to upload and joining the different fields manually.

When joining the data, a data preview option is available for administrators to double-check their joins to confirm the data extract is correct. Backend users can easily switch between the different ways to build out a data join. Data joins can also be made automatically if there's a name/object match with each table, however this match needs to be exact (e.g. if both tables had a "Provider" field, but one was stored as a text and the other was stored as a Binary Large Object (BLOB), then this join would not be automatic). The joining and loading process for MicroStrategy will be an involved process for the AHRQ team, but the Atlas Team believes that

because AHRQ's data structure remains consistent, this will be a one-time effort for a significant improvement in data availability and visualizations for all user types.

Visualizations within MicroStrategy are easily constructed through dragging and dropping variables. New variables can easily replace other variables and administrators can easily switch between their choices of visualizations to make everything match exactly what they are trying to achieve. Unique to MicroStrategy is a built-in natural language processer that gives backend users the ability to type out an idea for a visualization, and then builds out what it *thinks* you are trying to achieve. For example, an AHRQ administrator could type in "Visualize total medical expenditures for states" and the natural language processor would provide a recommended visualization which it thinks would display all states' total medical expenditures. This type of visualization management can accommodate a range of technological skills for administrators at AHRQ.

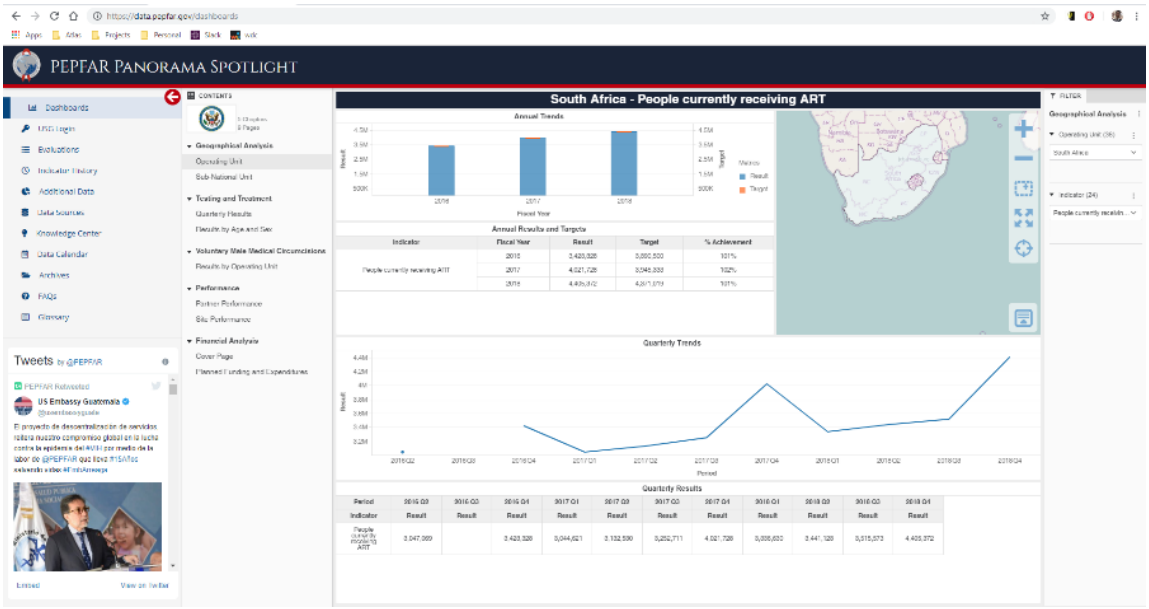
Visualizations for MicroStrategy are published through an application called "Dossiers", and Dossiers can be implemented through an iframe or a JavaScript Embedding API included with MicroStrategy. For more involved implementations that involve bidirectional integrations to a site, MicroStrategy recommends a JavaScript Embedding API as it allows for quick, real-time updates that are needed in the site. Nevertheless, an iframe will likely support all of AHRQ's needs for displaying data and easily recalling data without the need of waiting for data uploads.

One of MicroStrategy's weaknesses is that it's a more expensive tool relative to its peers. However, the additional expense funds options that will benefit the ARHQ team, which include an involved, supportive, and experienced implementation team, along with new and innovative customizations. The Atlas Team sees these additional offerings to set MicroStrategy apart from its competition. Moreover, the MicroStrategy implementation team will be willing to work with ARHQ's team to understand the full capacity needs of the tools to determine an optimal number of CPUs required to host AHRQs data in a MicroStrategy environment.

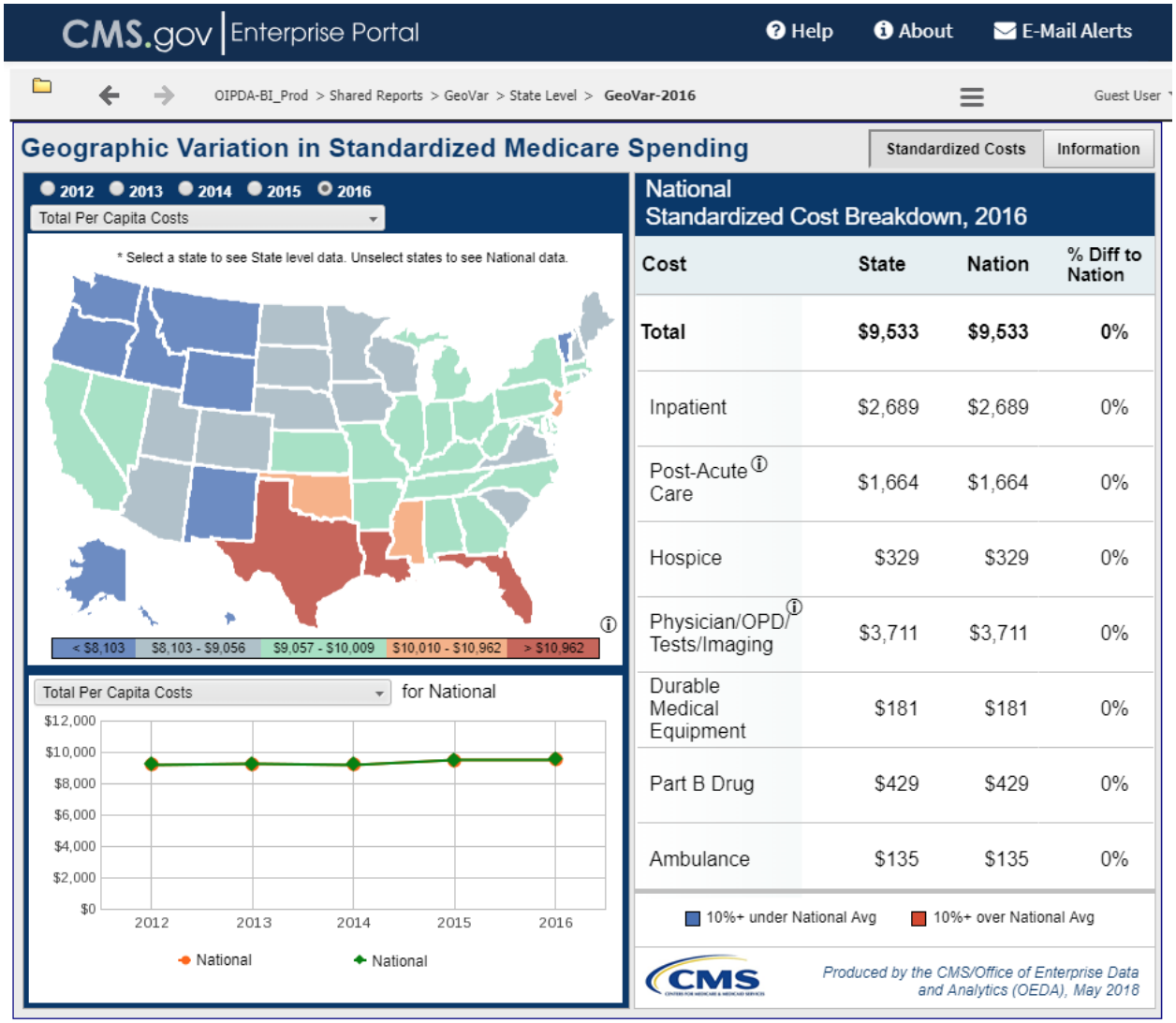
3.1.4 Federal Examples of MicroStrategy

A strength for MicroStrategy is its current footprint in the Federal space. MicroStrategy's presence in the Federal space makes the company adept at understanding the needs of Federal customer needs, such as 508 compliance and security provisioning. MicroStrategy is the platform for:

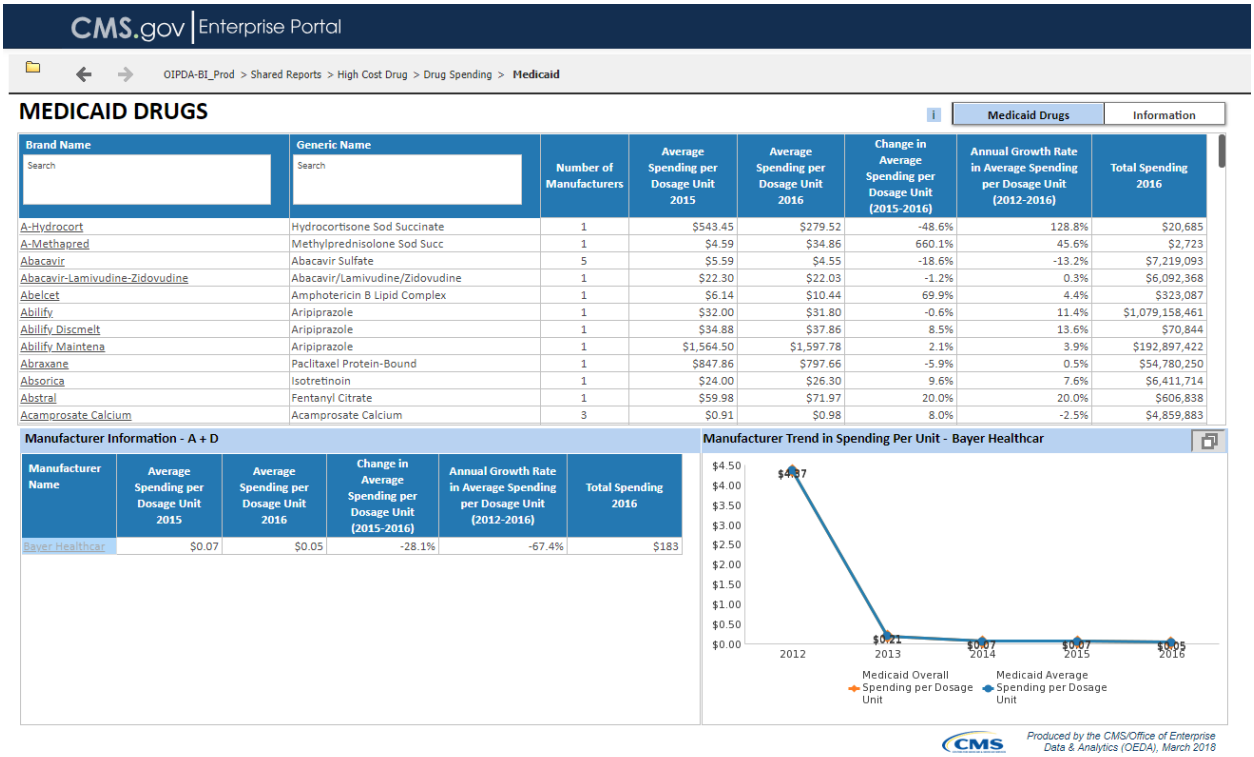
- The United States President's Emergency Plan for AIDS Relief (PEPFAR) data portal, which is a public-facing dashboard hosted by the United States State Department).



- Health and Human Services CMS.gov internal enterprise portal: this is a MicroStrategy hosted portal that allows users to review reports published by the Centers for Medicare & Medicaid Services (CMS) Office of Information Products and Data Analytics. When you first get to the portal page, you can click on the following links in each succeeding screen to view the Geographic Variation in Standardized Medicare Spending visualizations (sample screenshot below): *OIPDA-BI_PROD > Shared Reports > GeoVar > State Level > GeoVar-2016.*



- Below is a screenshot of another visualization page in the CMS.gov enterprise portal, this one for Spending for Medicaid Drugs. To get to this page, just click on the following links in each succeeding: *OIPDA-BI_PROD> Shared Reports > High Cost Drug > Drug Spending > Medicaid*.



CMS.gov | Enterprise Portal

OIPDA-BI_Prod > Shared Reports > High Cost Drug > Drug Spending > **Medicaid**

MEDICAID DRUGS

Medicaid Drugs | Information

Brand Name (Search)	Generic Name (Search)	Number of Manufacturers	Average Spending per Dosage Unit 2015	Average Spending per Dosage Unit 2016	Change in Average Spending per Dosage Unit (2015-2016)	Annual Growth Rate in Average Spending per Dosage Unit (2012-2016)	Total Spending 2016
A-Hydrocort	Hydrocortisone Sod Succinate	1	$543.45	$279.52	-48.6%	128.8%	$20,685
A-Methapred	Methylprednisolone Sod Succ	1	$4.59	$34.86	660.1%	45.6%	$2,723
Abacavir	Abacavir Sulfate	5	$5.59	$4.55	-18.6%	-13.2%	$7,219,093
Abacavir-Lamivudine-Zidovudine	Abacavir/Lamivudine/Zidovudine	1	$22.30	$22.03	-1.2%	0.3%	$6,092,368
Abelcet	Amphotericin B Lipid Complex	1	$6.14	$10.44	69.9%	4.4%	$323,087
Abilify	Aripiprazole	1	$32.00	$31.80	-0.6%	11.4%	$1,079,158,461
Abilify Discmelt	Aripiprazole	1	$34.88	$37.86	8.5%	13.6%	$70,844
Abilify Maintena	Aripiprazole	1	$1,564.50	$1,597.78	2.1%	3.9%	$192,897,422
Abraxane	Paclitaxel Protein-Bound	1	$847.86	$797.66	-5.9%	0.5%	$54,780,250
Absorica	Isotretinoin	1	$24.00	$26.30	9.6%	7.6%	$6,411,714
Abstral	Fentanyl Citrate	1	$59.98	$71.97	20.0%	20.0%	$606,838
Acamprosate Calcium	Acamprosate Calcium	3	$0.91	$0.98	8.0%	-2.5%	$4,859,883

Manufacturer Information - A + D

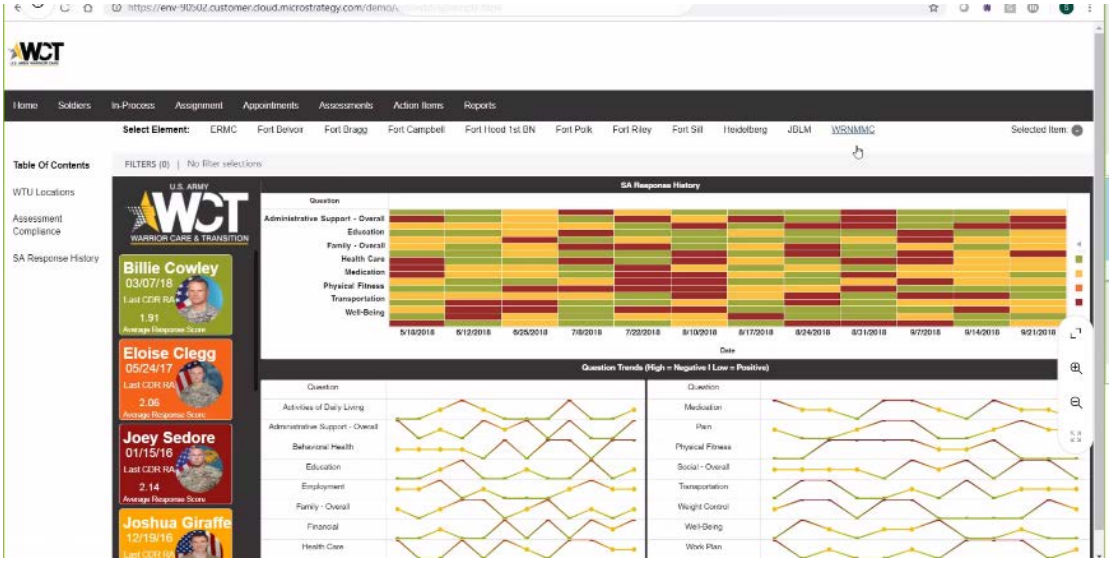
Manufacturer Name	Average Spending per Dosage Unit 2015	Average Spending per Dosage Unit 2016	Change in Average Spending per Dosage Unit (2015-2016)	Annual Growth Rate in Average Spending per Dosage Unit (2012-2016)	Total Spending 2016
Bayer Healthcar	$0.07	$0.05	-28.1%	-67.4%	$183

Manufacturer Trend in Spending Per Unit - Bayer Healthcar

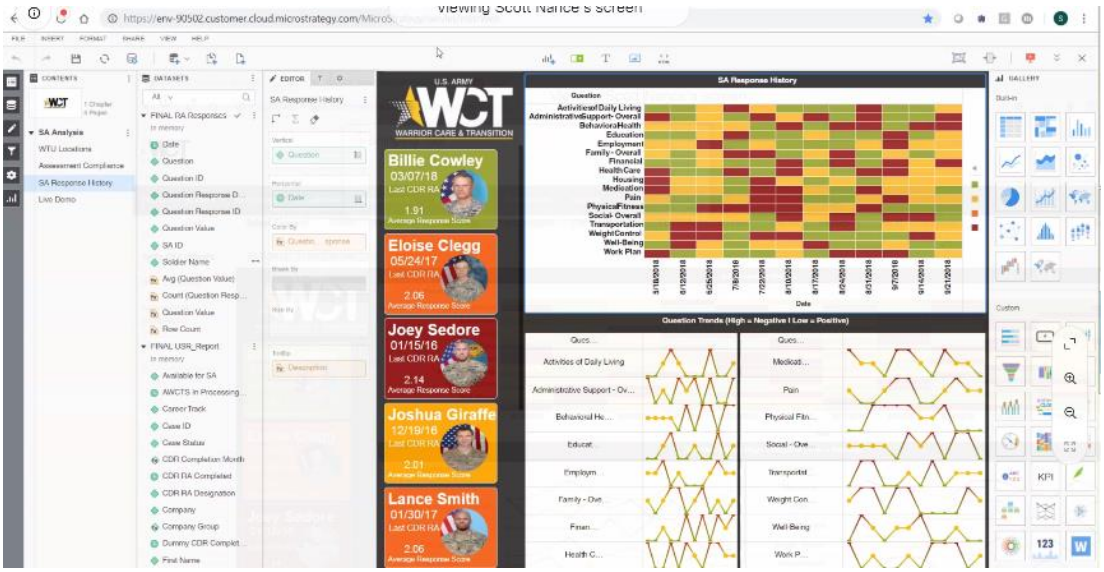
CMS

Produced by the CMS/Office of Enterprise Data & Analytics (OEDA), March 2018

- Department of Defense's Warrior Care and Transition internal facing dashboard (link to the actual MicroStrategy-based internal portal is not provided. However, screenshots are provided below). This site features bidirectional interactivity between the dashboard and the site hosting the dashboard, so adjustments made to the dashboard are reflected in the site.



  - Backend User's View



3.1.5 Final Takeaways for MicroStrategy

The improvement to the AHRQ's data query tools' UI will be immense if AHRQ selects MicroStrategy. Visualizations from AHRQ's data will be interactive for frontend users, and data visualization requests from frontend users can be more easily met by backend users. And MicroStrategy's centralized administrative page will give a simplified team of managers the

ability to determine data displays for frontend users. While MicroStrategy is a more expensive Data Visualization Product, it provides additional value, such as its language processing engine for visualizations. Each implementation also comes with a high touch implementation team who will support AHRQ's implementation team setting up MicroStrategy, which will lighten the load for AHRQ's team. Lastly, after the initial MicroStrategy setup, the product can be embedded throughout AHRQ's data query tools, allowing this product to scale across AHRQ's data query tools.

3.2 Qlik

3.2.1 Qlik Overview

Strengths	Weaknesses
Experienced with Federal clients.	Less intuitive administrative backend.
Compelling frontend graphics	Issues in scaling with large, disparate datasets.

Qlik's current analytics tool, Qlik Sense, is a SaaS Data Visualization Product and has a similar offering as MicroStrategy and Tableau. It can be deployed either on the cloud or on premises and uses a metadata layer to house data from various sources to display to frontend users. It employs a relatively intuitive backend UI but is less intuitive than Tableau or MicroStrategy. Qlik's strongest selling point is Qlik's experience in the Federal space and the compelling graphics it offers for Frontend users.

The most recent version of Qlik Sense is Version November 2018, which was released on 11/2/18. Qlik averages approximately 4 releases yearly, which include a mix of product enhancements, such as new visualizations and accessibility additions, and bug fixes.

3.2.2 Costs for Implementing Qlik

Qlik offers subscription licenses for Qlik Sense and estimated an annual cost of $83,000 based on our scenario. However, the Qlik team emphasized that a special enterprise license agreement (ELA) or blanket purchase agreement (BPA) might exist for Qlik with HHS or AHRQ. The quoted price is based on Government Services Administration (GSA) pricing. If AHRQ moves forward with Qlik, Qlik will work with AHRQ to accurately scope the capacity required and will review any ELAs or BPAs to provide a more favorable rate.

3.2.3 Hypothetical Qlik Implementation

Similar to MicroStrategy, Qlik can be deployed on premises or through the cloud to accommodate AHRQ's public facing datasets, the Qlik team will need to scope out the number of nodes (Qlik's infrastructure) required to support AHRQ's frontend userbase.

Once installed, Qlik will ingest AHRQ's data in a similar style to Tableau or MicroStrategy through a semantic metadata layer. Qlik's semantic layer is managed through a network of nodes that store the data, run analysis requested in the front facing app, store the data, and schedule uploads of the data. Like MicroStrategy or Tableau, Qlik is able to utilize out-of-the-box connectors and can support custom connectors where out-of-the-box options are not available, however AHRQ's Oracle and SQL servers are compatible with Qlik.

Visualizations are created through dragging and dropping data fields into Qlik's dashboard environment. This dashboard environment is slightly less intuitive than Tableau or MicroStrategy, however, and might lead to some frustration among administrators when creating variables to be visualized. It should be noted that Qlik requires an in-house application for geospatial data called "Qlik GeoAnalytics", which was introduced only in 2017. Qlik backend users, like Tableau backend users, have mentioned the platform's rigidity requires creating custom hacks to provide the frontend user with exactly what they're looking for.

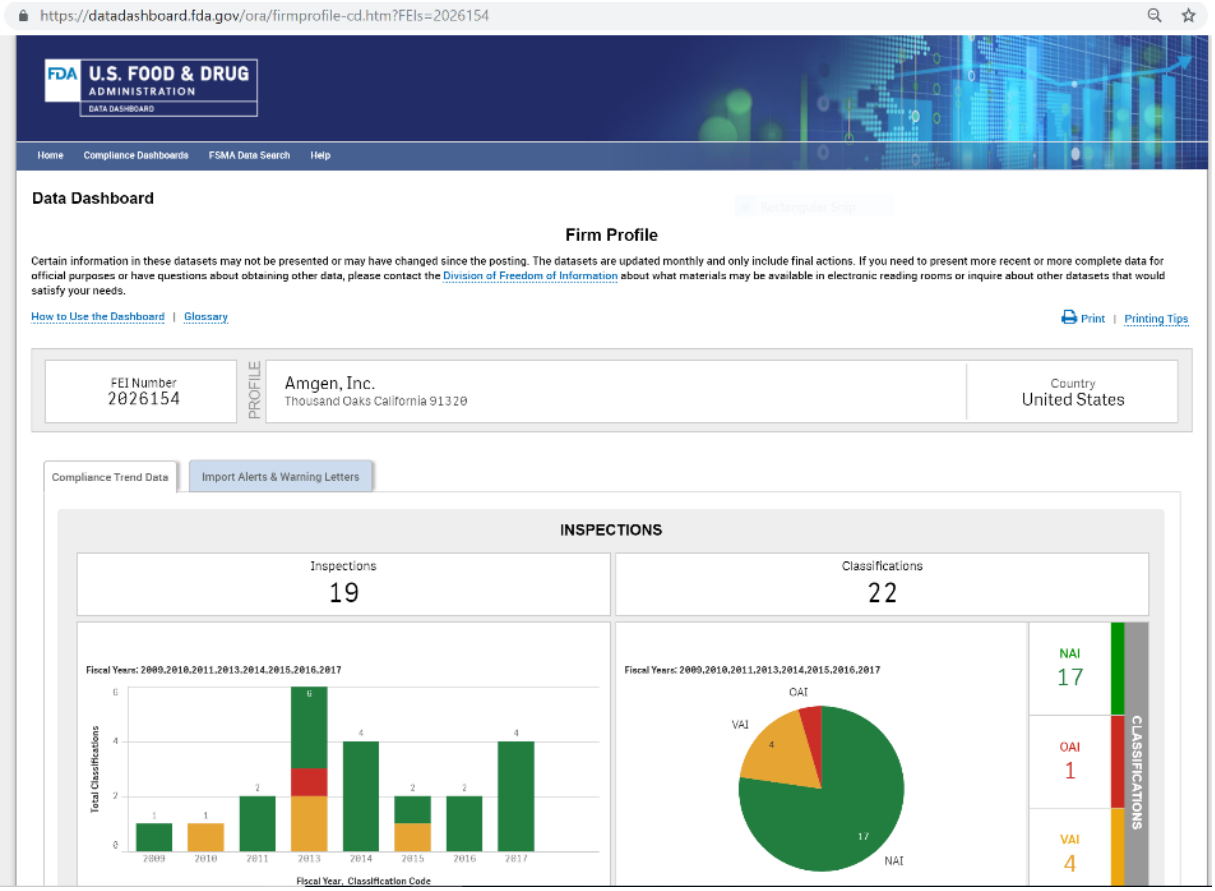
Like MicroStrategy or Tableau, Qlik offers up the ability to embed via iframe, REST API or through a JavaScript. For most implementations, iframe will be sufficient, but in instances where advanced integration is required, Qlik should be embedded with a REST API.

3.2.4 Federal Examples of Qlik

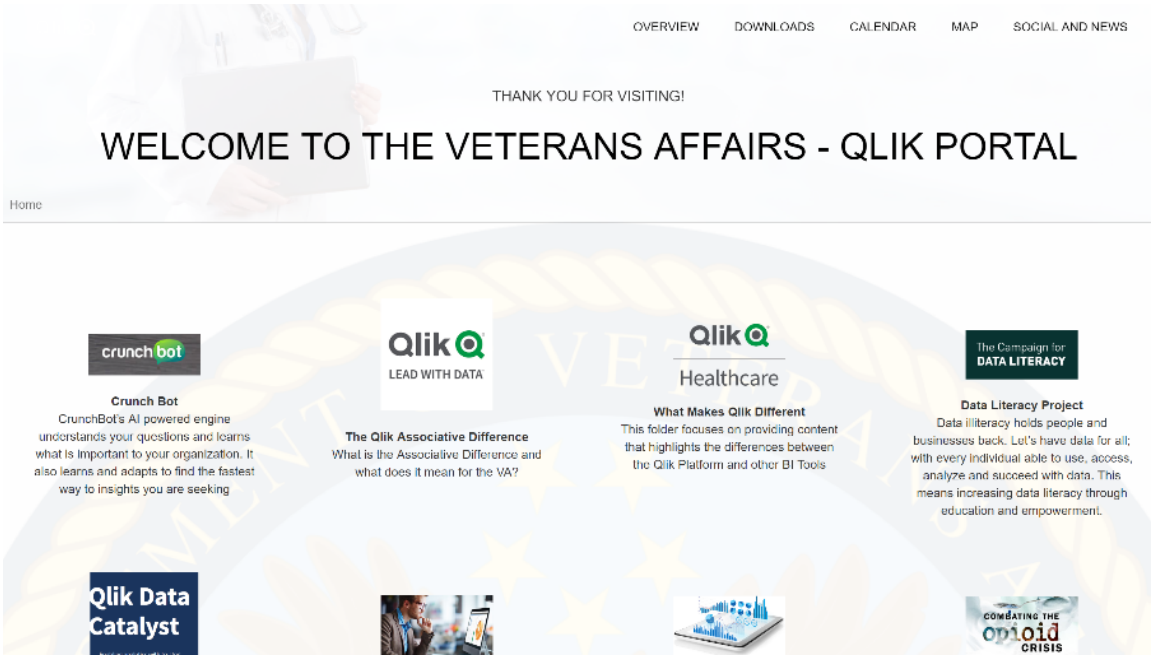
Setting Qlik apart is its presence in the Federal space. Qlik not only has interactive dashboards for uses to explore, but they also provide agency wide portals.

- The United States Food and Drug Administration's (FDA's) Compliance Dashboards for Inspections.

- FDA's Food Safety Modernization Act (FSMA) Data Search Portal provides a visualization page of FDA inspections. To see a sample of this page: from the FSMA Data Search portal, click on the Firm/Supplier Evaluation Resources link. Then, enter in a Firm Name in the search field (e.g., Amgen). In the resulting page, you can click on the firm's FDA Establish Identification (FEI) number (e.g., for Amgen, 2026154) to take you to a visualization page showing the firm's inspection results among other information.



- The United States Department of Veterans Affairs (VA) Qlik Portal which mainly hosts documents and videos. This site will require a sign-in process. After clicking on the link, click on “SIGN IN TO VIEW SECURED MATERIAL”. A popup window will appear with multiple sign-in options. Click on “LOGIN WITH YOUR EMAIL” for the simplest process. Once you enter an email address, you will be able to enter into the site.



This portal is supported by Qlik and although there aren’t any dashboards available for public consumption, it provides a feel for what’s possible with Qlik.

3.2.5 Final Takeaways for Qlik

Qlik can provide AHRQ with compelling visualizations for its frontend users, and Qlik’s extensive Federal experience could lead the Qlik to become a useful partner in supporting AHRQ to provide a more pleasant frontend and backend experience. And while the costs for Qlik is less than MicroStrategy, and comparative to Tableau, it offers a more built out solution than R Shiny. Qlik could prove to be a useful tool to simplify management of AHRQ’s multiple data products.

3.3 Shiny

3.3.1 Shiny Overview

Strengths	Weaknesses
Ability to upload and analyze many types of data (including survey data).	Coding learning Curve (R, Shiny UI, HTML, CSS).
Support from RStudio to host and publish content.	More involved build for AHRQ if this scales.
Open-source application with a large community.	

Shiny is not a SaaS product like Qlik, MicroStrategy or Tableau. Shiny is an easy to install package for the statistical programming language R. The package allows backend users to create interactive web apps using R along with HTML and CSS for a fully customized solution. Moreover, the full statistical ability for R is available to backend users providing them with a range of analysis and forecasting to display in Shiny.

The main benefit of Shiny is that every piece of every data tool can be customized. This provides an extraordinary ability to the central administrator to determine what the frontend user sees in AHRQ's digital tools. Moreover, the ability to modify visualizations is not dependent on a release schedule or a product roadmap.

However, the downside of this customizability is that all facets of the data tools need to be coded and developed. This means that a central administrator will need to be proficient in R, and any future builds or modifications will require an R subject matter expert (SME). Moreover, any additional modifications in HTML or CSS will require SMEs for the respective languages. This significantly impacts the ability to scale or transition the digital tools once they're developed. Fortunately, for this central administrator, the R userbase is active, passionate, extensive, and possible to learn from.

Moreover, AHRQ benefits from 2 additional products offered by RStudio to support this development: Shiny Server Pro and RStudio Connect. Shiny Server Pro provides an integrated development environment for the central administrator to code and develop web apps in R / Shiny. Moreover, RStudio Connect allows for push-button deployment and scheduled uploads to a web server to minimize the level of effort for a central administrator while reducing the probability of botched deployments.

The current version of R Shiny is version 1.2.0, which was released on 11/2/18. R Shiny currently averages approximately 4 releases per year that involve minor bug fixes. Product enhancements are released approximately once to twice a year.

3.3.2 Costs for Implementing Shiny

R Shiny offers a much lower setup cost compared to its competitors. The three tools that are needed to launch an R Shiny application are Shiny Server Pro, which costs approximately $10,000 annually for 20 concurrent backend users, RStudio Package Manage, which will be essential in centralizing and managing R packages across the AHRQ team, costs $10,000 annually. Lastly, RStudio Connect's base package for 20 named users is priced at $15,000.

This totals to $35,0000 per year for the number of backend users who will be required to do the development work for the different data tools.

3.3.3 Hypothetical Shiny Implementation

The first step to building a Shiny application for the AHRQ team will be installing RStudio Connect, RStudio Server Pro, and RStudio Package Manager on premises, as these products need to be installed behind AHRQ's security system. To enhance the ability to seamlessly integrate functioning Shiny code in AHRQ's tools will require using Shiny Server Pro as a code editor, code debugger and as a visualizer. Additionally, Shiny Server Pro allows for centralizing the development of Shiny code to a team of developers.

Shiny was designed specifically for interactive and compelling visualizations. An AHRQ development team will have no issues in selecting different types of visualizations for frontend users consuming AHRQ data currently. If a frontend user request comes in for a new visualization, the AHRQ development team will need to code this request, so it will be important for the team to standardize all visualization requests to fully understand the needs and requirements of the frontend user. Since visualizations will need to be coded by a development team, it will be important to source inspiration for data visualizations from the Shiny user community.

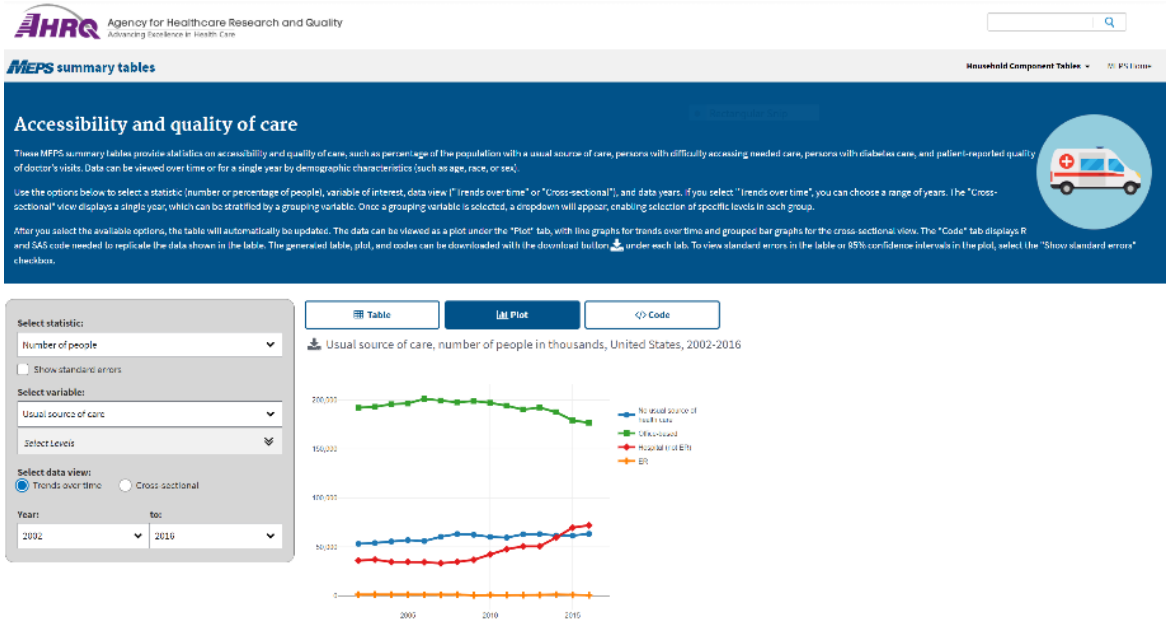
Once the developers connect up AHRQ's data sources to R / Shiny for development, they will need to develop the responsive application that they intend to display in the AHRQ hosted site. Once Shiny code is ready to be deployed into AHRQ source code, the AHRQ team can use RStudio Connect. RStudio Connect's push-button deployment allows for easy deployment into AHRQ's environment and provides a version compatibility check. RStudio Connect also has a scheduler functionality built into it which can allow for updates to be pushed on a regular basis without compromising the UX, which can be useful for deployments that need to be recalled.

If the AHRQ team wants to employ Shiny across all the data tools, then this process will need to be repeated for each data tool. However, if the AHRQ team would like to harmonize data sources, plans should be developed to determine which AHRQ data tool will host the different sources of data and how the data will be harmonized (i.e. stored in separate databases or stored in a new unified database). Given that Shiny doesn't provide a data modeling framework like other SaaS products, these important steps need to be considered.

3.3.4 Federal Examples of Shiny

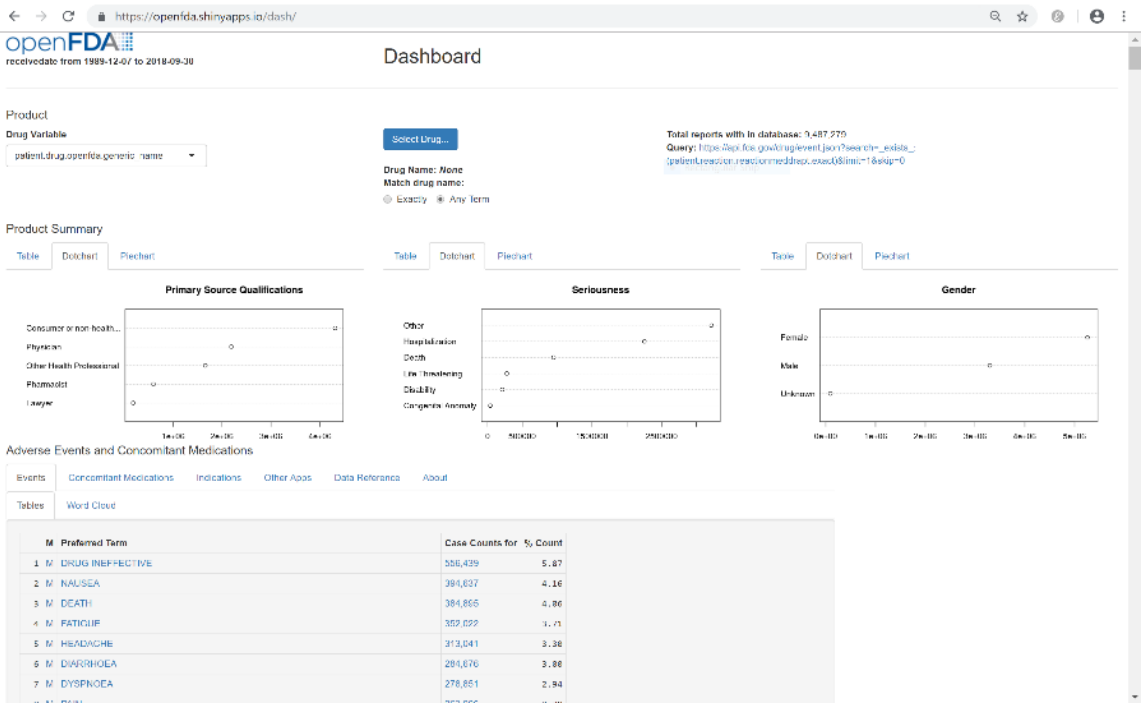
Since R and Shiny are open source, Federal Government agencies will use the tool for a variety of projects.

- AHRQ's [Medical Expenditure Panel Survey (MEPS) Household Component Summary Tables application].



- United States Department of Agriculture's (USDA's) [Region 11 Soil Quality Manager App].

- FDA's Open FDA Research Tools.



.

3.3.5 Final Takeaways for Shiny

Shiny's strongest selling point is that it's able to offer compelling, interactive frontend UIs for less than half the price of its competitors. However, the backend UI will be less elegant compared to offerings like Tableau or MicroStrategy and will involve a significant amount of coding for each visualization. Shiny would work well if AHRQ's ambitions are limited to updating the UI for their current data tools. While there are additional tools that help simplify management for Shiny code, new infrastructure will need to be considered if AHRQ intends to scale this solution to a new AHRQ portal or to fully harmonize different AHRQ data tools.

3.4 Tableau

3.4.1 Tableau Overview

Strengths	Weaknesses
Established brand and company.	Mild learning curve.
Extensive backend userbase with good resources for self-training.	Some hacks required for certain customizations.
Easy connection between metadata and data extracts.	
Experienced with Federal clients.	
Large visualization library.	

Tableau is a SaaS analytics product whose popularity has dramatically increased recently due to its self-service capabilities and its ability to quickly and easily provide visualizations for frontend users. Tableau benefits from having an extensive userbase while also having experience in the Federal space. The larger userbase provides a network for collaboration or idea sourcing, while the experience in the Federal space secures Tableau's brand as a trusted product.

Nevertheless, backend users of Tableau who are not involved in the nitty-gritty implementation details will experience a learning curve in getting started, albeit less of a learning curve than R / Shiny. Additionally, while Tableau has high regard as a quick-to-deploy SaaS solution, some users complain that the product can be rigid in delivering precise and unique customizations. For example, buttons outside of what's needed in a visualization require a unique hack. What compounds this issue is the support staff for Tableau is less helpful than a solution like MicroStrategy. Although the userbase can act as a hive mind to develop solutions for common Tableau problems, it will be up to the administrators to uncover solutions to Tableau problems. Nevertheless, for intuitive visualizations coupled with an easy-to-use intuitive interface, Tableau provides a compelling offering for AHRQ.

The most recent version of Tableau is version 2018.3.2. Out of all the products, Tableau has the most frequent release schedule, averaging a release once every 3 weeks. However, these frequent releases include bug fixes. Tableau includes product enhancements in releases approximately once or twice per year.

3.4.2 Costs for Implementing Tableau

Tableau offers an Unlimited user model for applications that will be embedded in public facing websites. To accommodate the number of users on a concurrent basis, the Tableau team recommends a Core model, which is hardware-based. The Tableau team estimated the Core model would cost $72,000 per year based on the assumption of roughly 75k monthly users. The Tableau team emphasized that part of the implementation process will involve scoping out a more accurate capacity than the current estimate.

3.4.3 Hypothetical Tableau Implementation

Like other SaaS products, Tableau offers both cloud and on-premises deployments. The Tableau team will work with the AHRQ team to scope out the number of CPUs required to accommodate the usual traffic to AHRQs data tools. Tableau is used to working with large enterprises that display their data for public consumption, and as such have an unlimited user model that support AHRQ's scenario.

Once a Tableau solution is deployed and administrator accounts have been provisioned, administrators determine which data are connected to a Tableau metadata semantic layer, allowing administrators to create visualizations. Administrators, with Tableau Creator licenses, use simple drag-and-drop, or develop SQL code, to select fields to visualize. Tableau can accommodate a variety of database servers to connect to this drag-and-drop solution by using out-of-the-box connectors for popular database servers such as Oracle or SQL. For database servers where an out-of-the-box connector is not available, the Tableau team will work to create an ODBC Connector.

Multiple data sources can be visualized simultaneously in this sandbox environment and published for public use. To determine the correct visualization, Tableau can recommend visualizations based on the types of data that are being used by the administrator. For further visualization inspiration, Tableau offers best practices on their website and have an extensive userbase where administrators could potentially source new ideas.

Similar to MicroStrategy, embedding can be done through an iframe or REST APIs or JavaScript APIs. An iframe is recommended for more simple implementations; however, if an implementation is more involved and requires bidirectional communication with the website hosting Tableau, then a JavaScript API would be recommended.

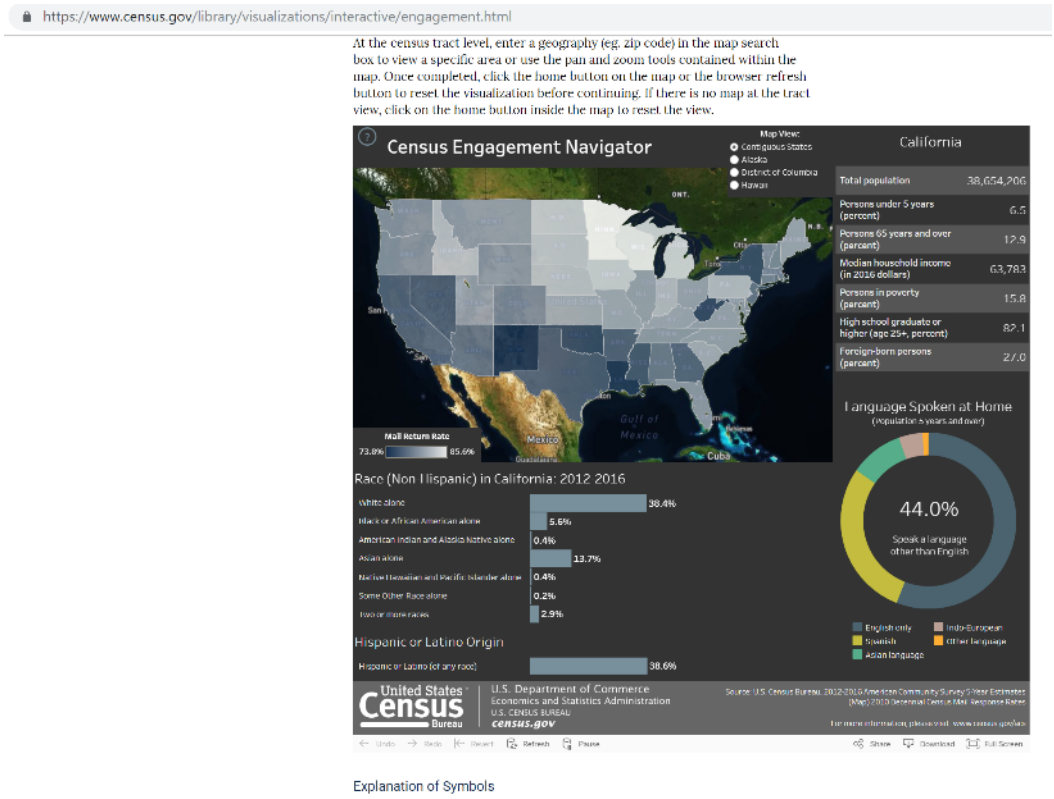
3.4.4 Federal Examples of Tableau

Tableau's presence in the Federal space is noticeable and aligns well with AHRQ's potential uses. For example, Health Resources and Services Administration (HRSA) is a featured Tableau client in multiple departments, as seen below:

- HRSA Data Warehouse (Currently a site considered by the Atlas Team in the Environmental Scan section for Agency Comparison).
    - For example, HRSA uses Tableau to visualize grants disbursement. See below for a mapping visualization of the recipient state funding level and the program area for funding.

- Federal Aviation Administration (FAA) internal Program Management Organization (PMO): Link not provided.

- The United States Census Bureau [Engagement Navigator](https://www.census.gov/library/visualizations/interactive/engagement.html).



Explanation of Symbols

3.4.5 Final Takeaways for Tableau

The ability for Tableau to be integrated into AHRQ's current data tools, combined with Tableau's offering of easy to create compelling visualizations, already makes the product a valuable solution. Considering Tableau's Federal Experience and its active userbase, the Atlas Team highly recommends Tableau as an easy-to-use solution for backend and frontend users of all skill levels that could visualize AHRQ data in a unique way for its userbase. Moreover, centralizing the administration of frontend UIs can help simplify the management of AHRQ's data query tools at a lower cost than a product such as MicroStrategy.

4. LOW PRIORITY PRODUCT ASSESSMENTS

4.1 Birst

When considering purchasing software, it's important to understand what problem the software is trying to solve. Compared to its competitors, Birst is unable to provide a convincing solution for AHRQ's needs and seems more appropriate for smaller organizations trying to make sense of their data rather than a large organization that already understands its data and what it wants to do with the data.

Birst is designed in a similar fashion to MicroStrategy or Tableau, where a 2-tier architecture incorporates a metadata semantic layer for data modeling, and then provides the ability for backend users to create visualizations. Birst provides automated machine learning for data uploaded into the semantic layer. However, the opportunity for AHRQ to take advantage of this is limited.

While Birst's offerings may be sufficient, there are not any compelling advantages that would dramatically increase the value for the AHRQ team compared to alternatives such as Tableau or MicroStrategy. The drag-and-drop method for creating visualizations or joining data are less intuitive than all other products. The visualizations offered by Birst are adequate, but not extraordinary and include standard mapping, plotting and data tables. One similarity between Birst and MicroStrategy is that there's a recommendation engine within Birst that suggests what type of data visualization would work best with selected data. However, this pales in comparison to MicroStrategy's ability to use natural language processing to provide the same outcome that's more intuitive for a lay backend user.

For additional statistical analysis, the AHRQ team will need to integrate Birst into either Tableau or R / Shiny, which at that point further degrades any selling points that this solution offers.

4.2 Domo

Domo provides a similar 2-tier architecture strategy as Birst, wherein data from the AHRQ databases would need to be loaded into a semantic layer before being able to visualize and manipulate. From there, Domo provides multiple in-house apps that, while impressive, do not provide any particular value for the AHRQ team, and could risk overriding existing processes. For example, Domo offers a chat app in their app store. If this is used along with an existing preferred application by AHRQ, then backend users might be confused which chat app to use to communicate with colleagues.

Domo offers standard data modeling through a drag-and-drop sandbox environment like other products. It can accommodate multiple sources of data if the time comes for the AHRQ team to harmonize certain data tools. Nevertheless, it offers these solutions at a more expensive rate than a similar alternative such as Birst given that it needs to support a slew of in-house apps that might not be valued by the AHRQ team. Moreover, Domo doesn't offer any on-premises deployment, which limits the flexibility for AHRQ's deployment.

4.3 Looker

While most BI solutions provide a cloud-based database to upload data into, Looker is unique in that it requires the client to setup their database to store that data. This means that the client

must not only have a database server setup for their own data but effectively duplicate their data to have the benefit of Looker.

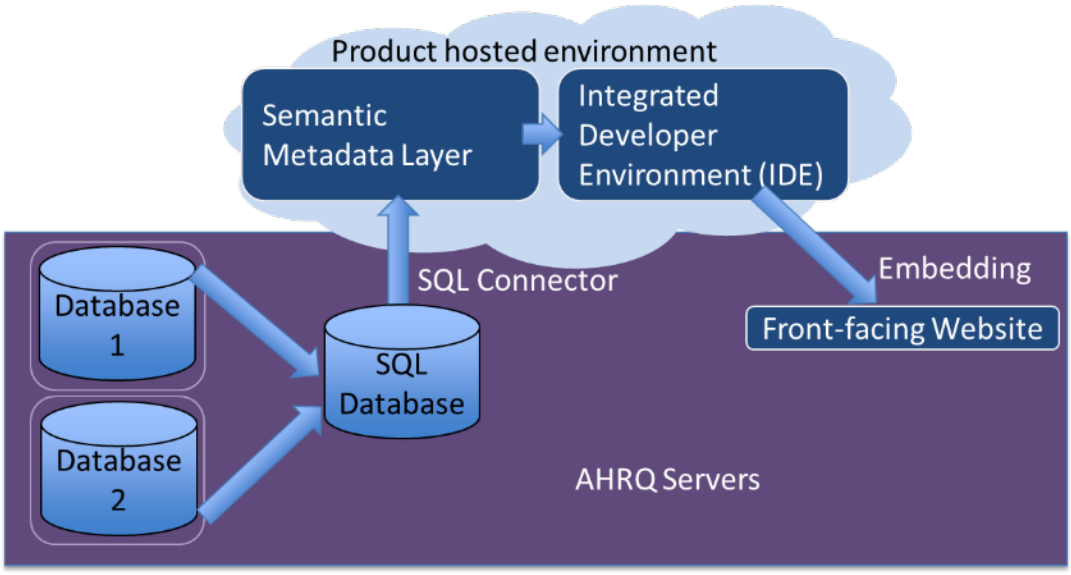


Figure 2. Example of a hypothetical Looker implementation

While Looker can provide useful visualizations, this unorthodox setup, coupled with a lack of experience with Federal clients, prevents the Atlas Team from recommending Looker as a viable solution.

4.4 Microsoft Power BI

Microsoft Power BI has a lot of promise. Power BI's backend is built off Excel, so backend users will be familiar with the layout. Moreover, it can intake data from several sources through its connectors, such as SQL and Oracle, but also from more unique connectors, such as Facebook or GitHub. Lastly, since Power BI has the backing of Microsoft, it already has a substantial userbase to rely on, even though it's only been in operation for 4 years.

Nevertheless, Power BI's backend UX is not as elegant as similar competitors, since it is forced to conform to a certain Windows standard. Furthermore, the formulas included in Power BI prevent it from doing more robust calculations that will be required for the AHRQ Query Tools (such as Z-tests).

4.5 Sisense

Sisense is a strong competitor for a Data Visualization Product. It offers a straightforward deployment for loading a variety of different AHRQ data into a Sisense hosted cloud environment and an intuitive UI to give backend users the ability to build out visualizations easily and quickly. Moreover, embedding the visualizations into AHRQ websites can be easily done using an iframe or an enhanced JavaScript coding for more involved implementations. This intuitive design will allow for quick scaling amongst all of AHRQ's digital tools, easily harmonizing a variety of datasets and providing compelling visuals to frontend users, giving AHRQ the chance to score quick wins.

What prevents Sisense from moving forward are its costs compared to the value that it can provide and its lack of experience with Federal clients. The cost of implementation for Sisense is similar to that of Domo's; however, Domo provides many more additional features for that price, thereby justifying the cost. The rationale for Sisense's costs is not immediately evident, but it prevents the Atlas Team from strongly endorsing this product as a viable alternative against current options. Moreover, Sisense is more nascent in building its company, brand and reputation, though it's further along than Looker. Sisense has not yet gathered the relevant experience to truly understand the needs of a Federal userbase, such as having 508 compliance available.

Sisense's simple functionality and easy-to-grasp interface nearly gave it an edge over competing Low Priority products; however, a lack of fully understanding Federal clients could risk AHRQ's reputation during a live deployment.

APPENDIX A: PRODUCT ASSESSMENT QUESTIONS

- Confirm Licensing Model for public consumption
- Scalability—How easy is it to:
    - Upload data
    - Increase the number of tables
    - Increase the number of connections between tables
    - Add backend users
    - Train backend users on using the product (i.e. is the display intuitive enough for backend users to start plugging away at the product, or do they have to learn something new to use the product)
    - Embed into a site
    - Connect to disparate data tables on different servers
- Speed of tool, are there any bottlenecks for speed?
- How do I update data in the tables?
- How do I create new tables?
- Examples of other Federal Agencies using this?
- Is it 508 Compliant? How do we get 508 compliance?
- What security provisions are in place for data? Is it hosted on their servers? What Firewall is present?
- Walk us through an example of using the table feature.
    - How intuitive is the navigation?
    - What are the steps needed to find a specific data point?
- Walk us through an example of using the mapping feature
    - How intuitive?
    - Is there anything more than just a map of the US?
- Walk us through an example of plotting
    - How intuitive?
    - How easy is to change variables to make the plot?
- Walk us through an example of data manipulation
    - How easy is this?
- Can you show us an example of an entity relationship diagram?
- What tables and plots are you able to download?
    - Table: xlsx, csv, json
    - Plot: png, jpeg
    - Maps: pdf
- How does this integrate with other statistical programs?
- How easy is that to integrate?
- What options are available to make this available for public consumption?

APPENDIX B: ACRONYMS AND DEFINITIONS

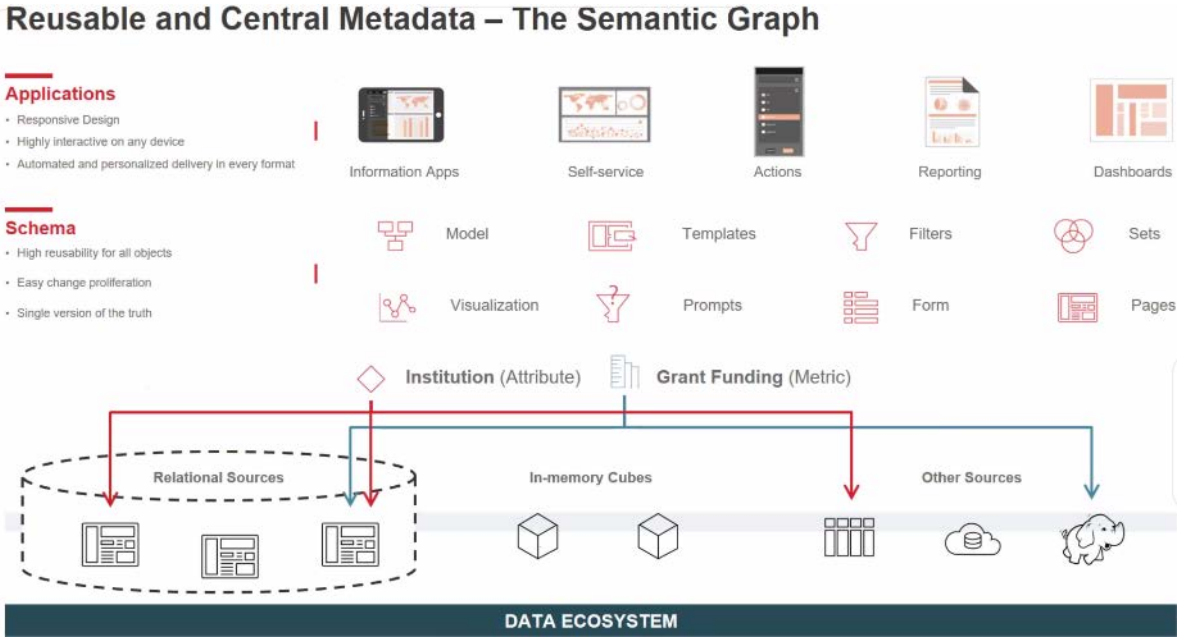
Table 1. List of Acronyms and Definitions

Acronym	Description
AHRQ	Agency for Healthcare Research and Quality
API	Application Program Interface
BLOB	Binary Large Object
BLS	Bureau of Labor Statistics
CMS	Centers for Medicare & Medicaid Services
COTS	Commercial off the-Shelf
CSS	Cascading Style Sheets
FAA	Federal Aviation Administration
FDA	Food and Drug Administration
GUI	Graphical User Interface
HPSA	Health Professional Shortage Area
HRSA	Health Resources and Services Administration
HTML	Hyper Text Markup Language
IDE	Integrated Development Environment
JDBC	Java Database Connector
MEPS	Medical Expenditure Panel Survey
NHSC	National Health Service Corps
ODBC	Open Database Connectivity
PEPFAR	President's Emergency Plan for AIDS Relief
PMO	Program Management Organization
REST	Representational State Transfer
SaaS	Software as a service
SDK	Software Development Kit
SME	Subject Matter Expert
SQL	Structured Query Language
UI	User Interface
USDA	United States Department of Agriculture
UX	User Experience
VA	Department of Veterans Affairs

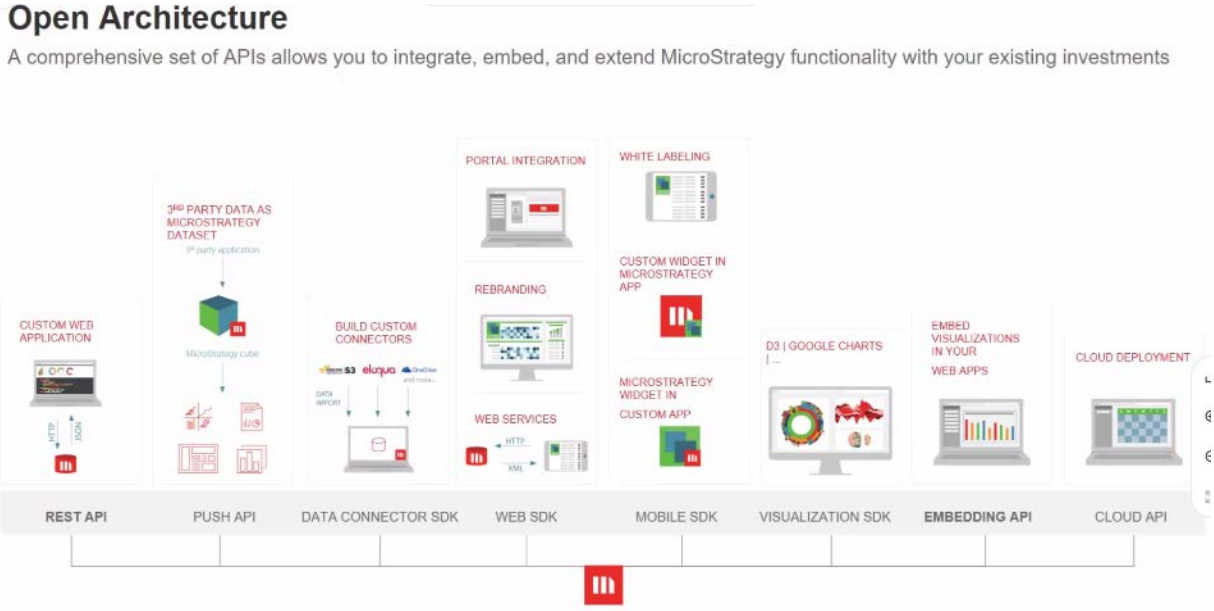
APPENDIX C: SCREENSHOTS

C.1 MICROSTRATEGY SCREENSHOTS

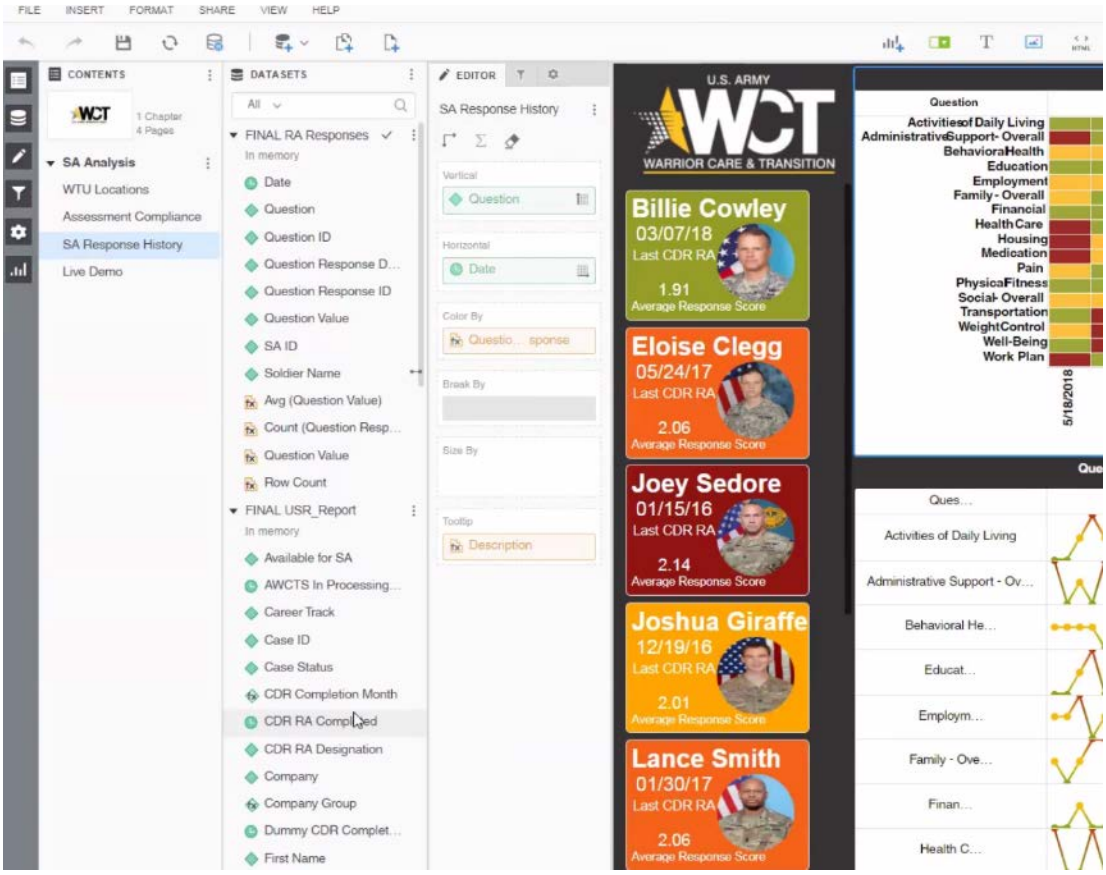
C.1.1 MicroStrategy Architecture



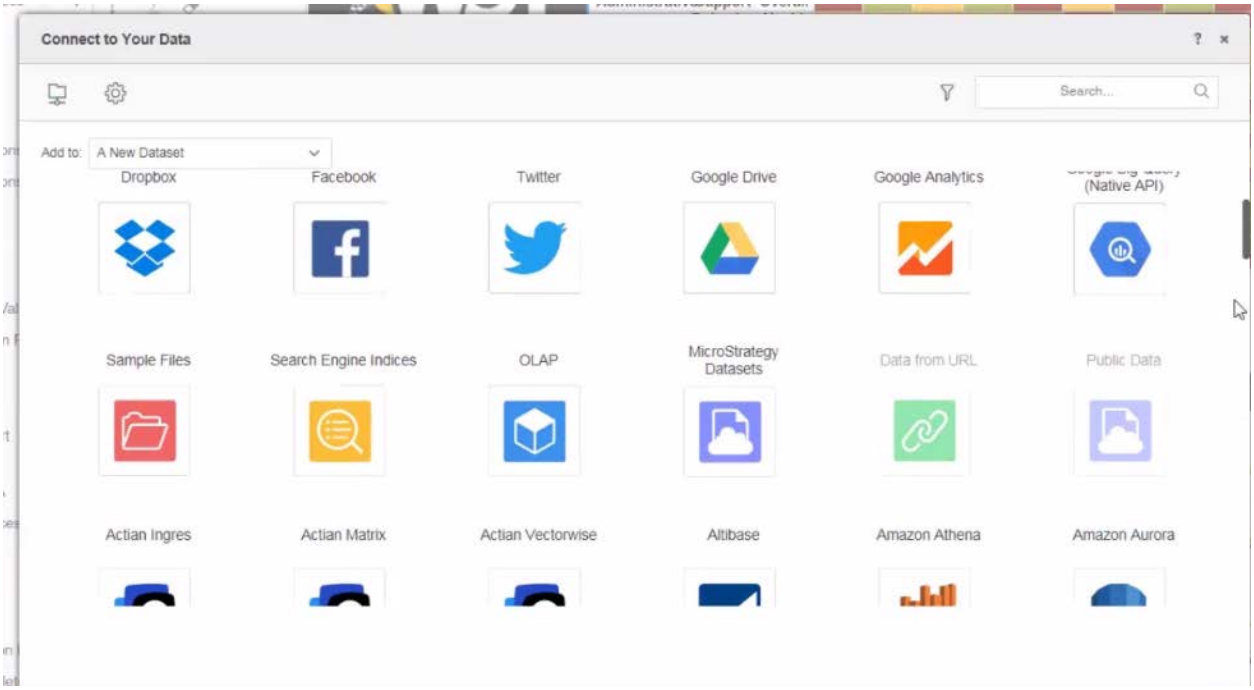
C.1.2 MicroStrategy Open Architecture



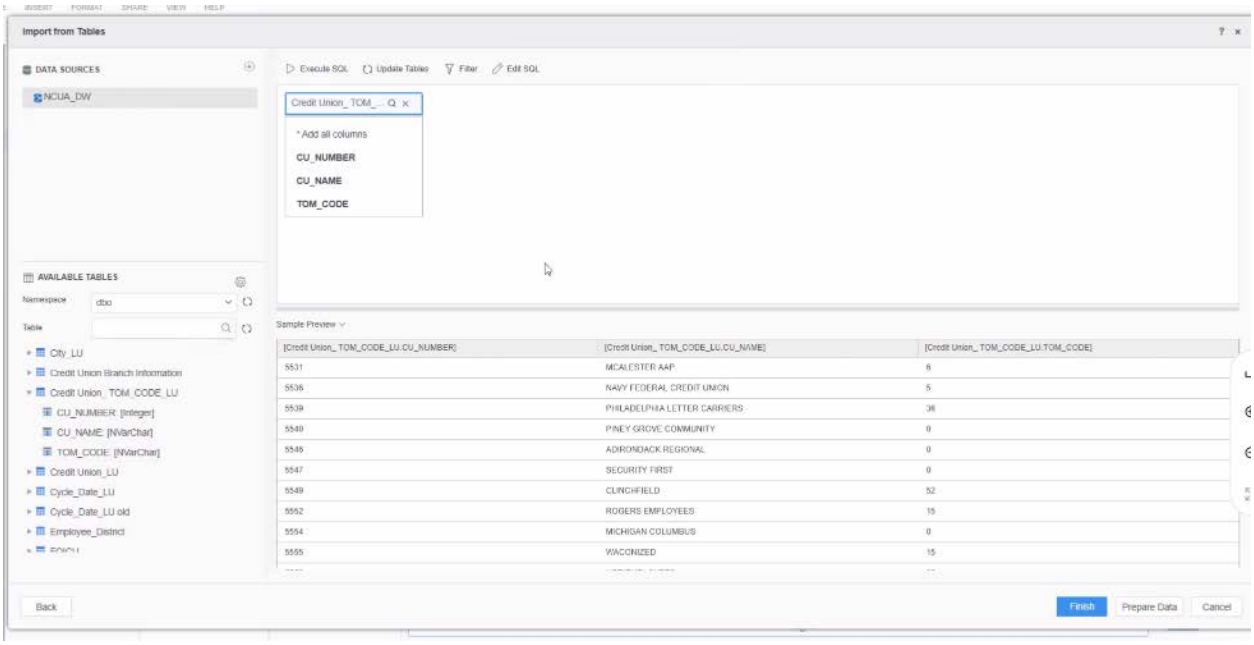
C.1.3 MicroStrategy Live Editing Visualization



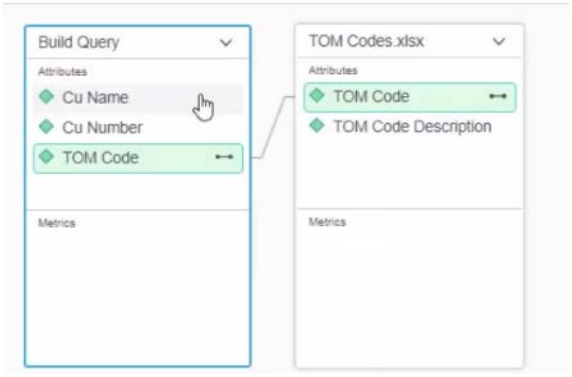
C.1.4 MicroStrategy Data Connectors



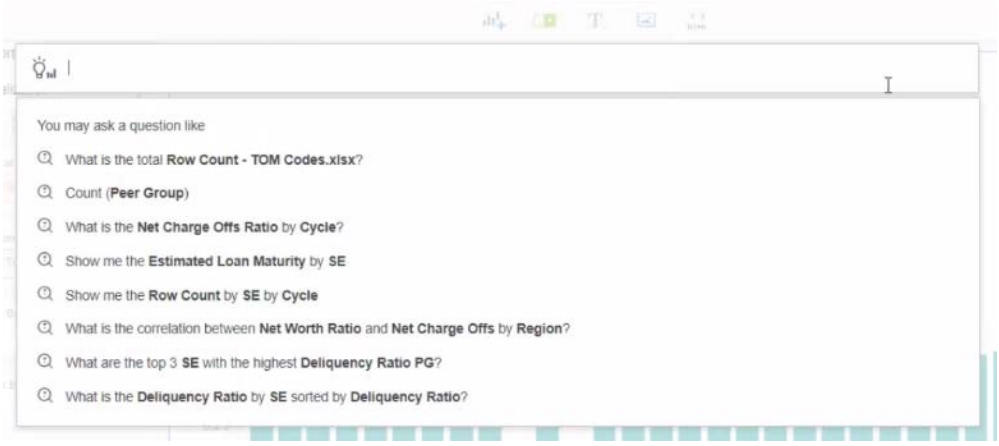
C.1.5 MicroStrategy Backend



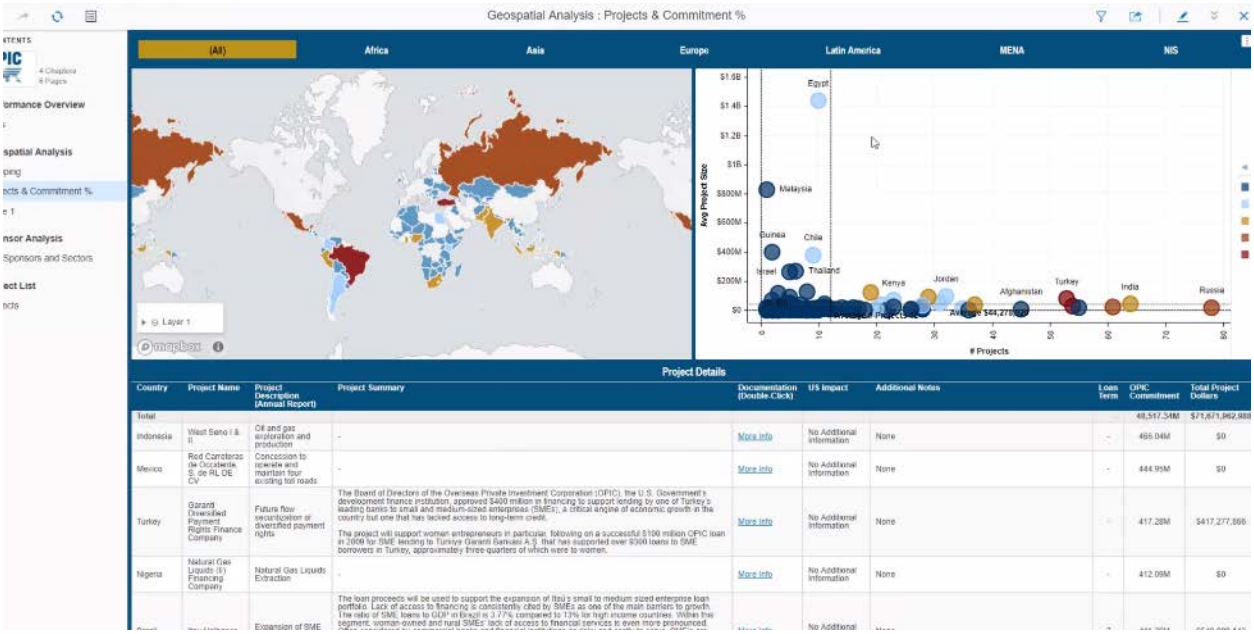
C.1.6 MicroStrategy Data Joins



C.1.7 MicroStrategy Natural Language Visualization Engine

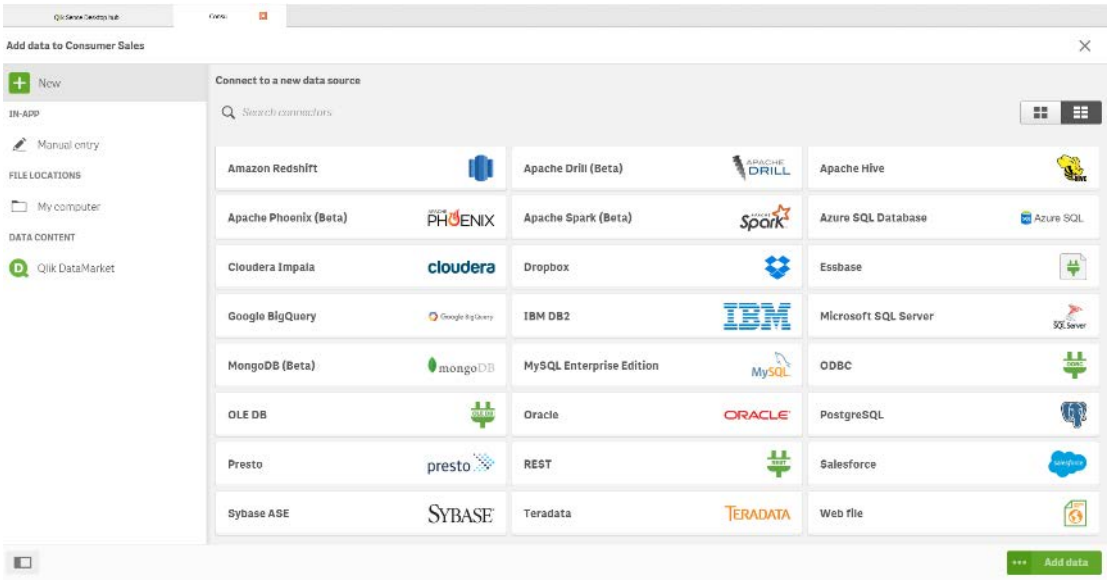


C.1.8 MicroStrategy Frontend



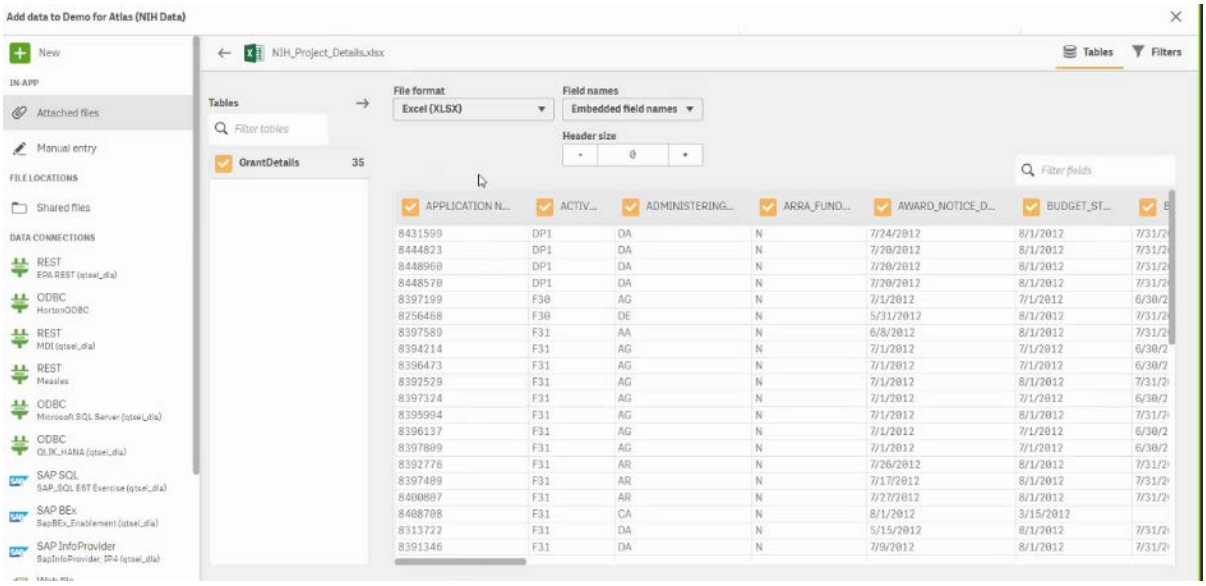
C.2 QLIK SCREENSHOTS

C.2.1 Qlik Data Connector

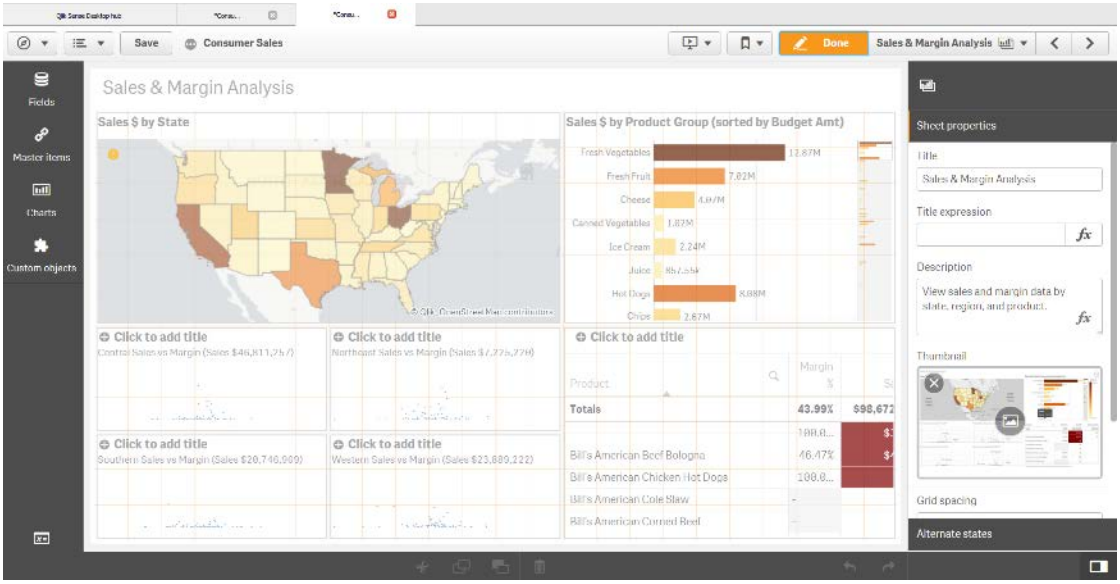


C.2.2 Qlik Backend Data Join

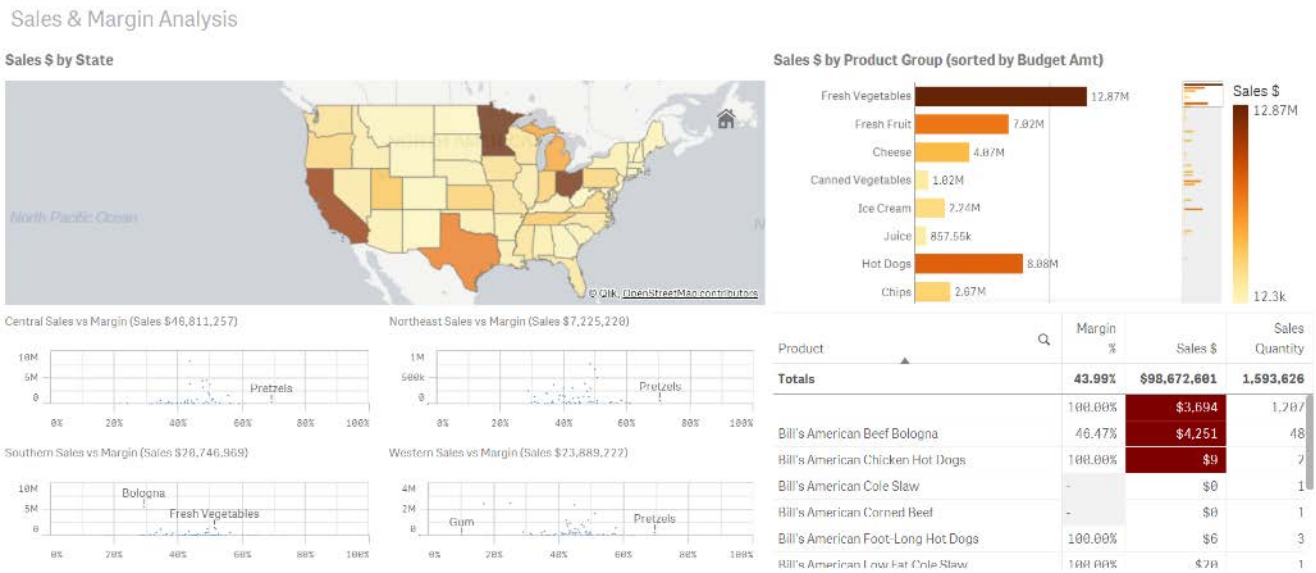
C.2.3 Qlik Backend Data View



C.2.4 Qlik Backend Editor



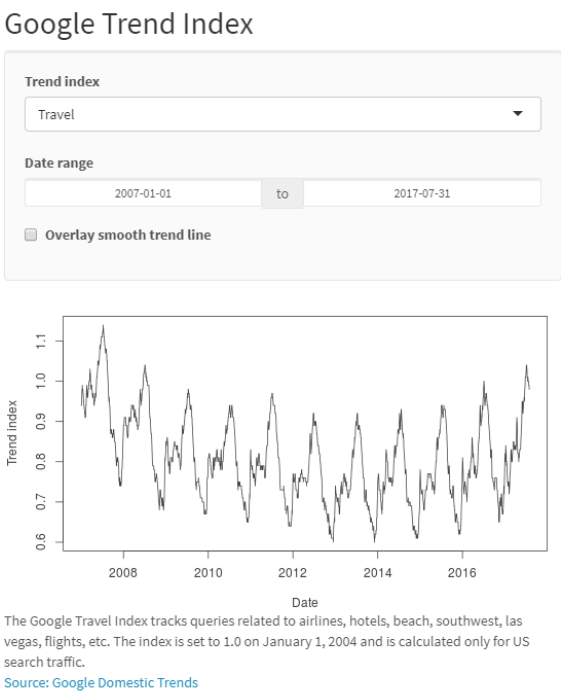
C.2.5 Qlik Frontend View



C.2.6 Qlik Visualizations

C.3 SHINY SCREENSHOTS

C.3.1 Shiny Application Frontend



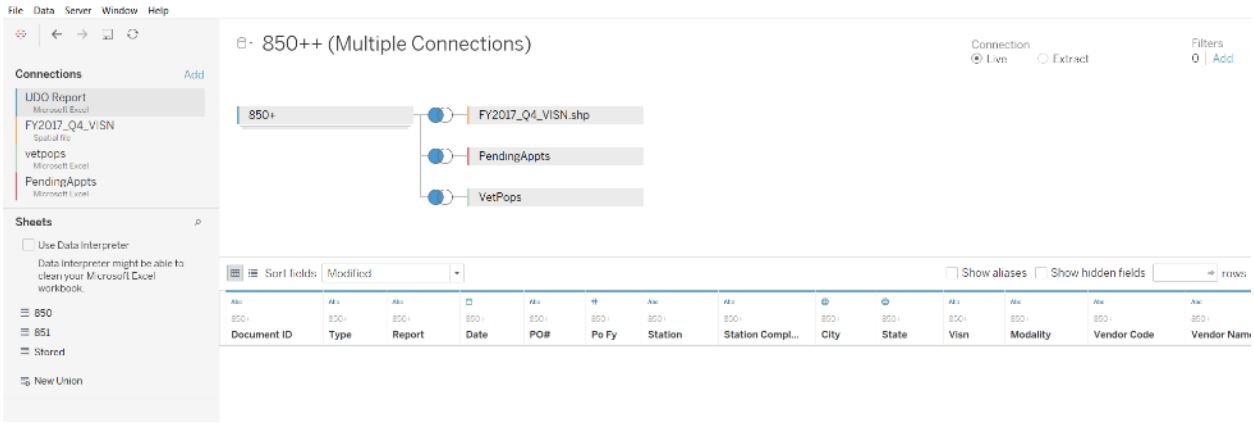
C.3.2 Shiny Application Backend

```
# Load packages
library(shiny)
library(shinythemes)
library(dplyr)
library(readr)

# Load data
trend_data <- read_csv("data/trend_data.csv")
trend_description <- read_csv("data/trend_description.csv")

# Define UI
ui <- fluidPage(theme = shinytheme("lumen"),
  titlePanel("Google Trend Index"),
  sidebarLayout(
    sidebarPanel(

      # Select type of trend to plot
      selectInput(inputId = "type", label = strong("Trend index"),
                  choices = unique(trend_data$type),
                  selected = "Travel"),

      # Select date range to be plotted
      dateRangeInput("date", strong("Date range"), start = "2007-01-01", end = "2017-07
                     min = "2007-01-01", max = "2017-07-31"),

      # Select whether to overlay smooth trend line
      checkboxInput(inputId = "smoother", label = strong("Overlay smooth trend line"),

      # Display only if the smoother is checked
      conditionalPanel(condition = "input.smoother == true",
                       sliderInput(inputId = "f", label = "Smoother span:",
                                   min = 0.01, max = 1, value = 0.67, step = 0.01,
                                   animate = animationOptions(interval = 100)),
                       HTML("Higher values give more smoothness.")
      )
    ),

    # Output: Description, lineplot, and reference
    mainPanel(
      plotOutput(outputId = "lineplot", height = "300px"),
      textOutput(outputId = "desc"),
      tags$a(href = "https://www.google.com/finance/domestic_trends", "Source: Google D
    )
  )
)

# Define server function
server <- function(input, output) {

  # Subset data
  selected_trends <- reactive({
    req(input$date)
    validate(need(!is.na(input$date[1]) & !is.na(input$date[2]), "Error: Please provide
    validate(need(input$date[1] < input$date[2], "Error: Start date should be earlier t
    trend_data %>%
      filter(
```

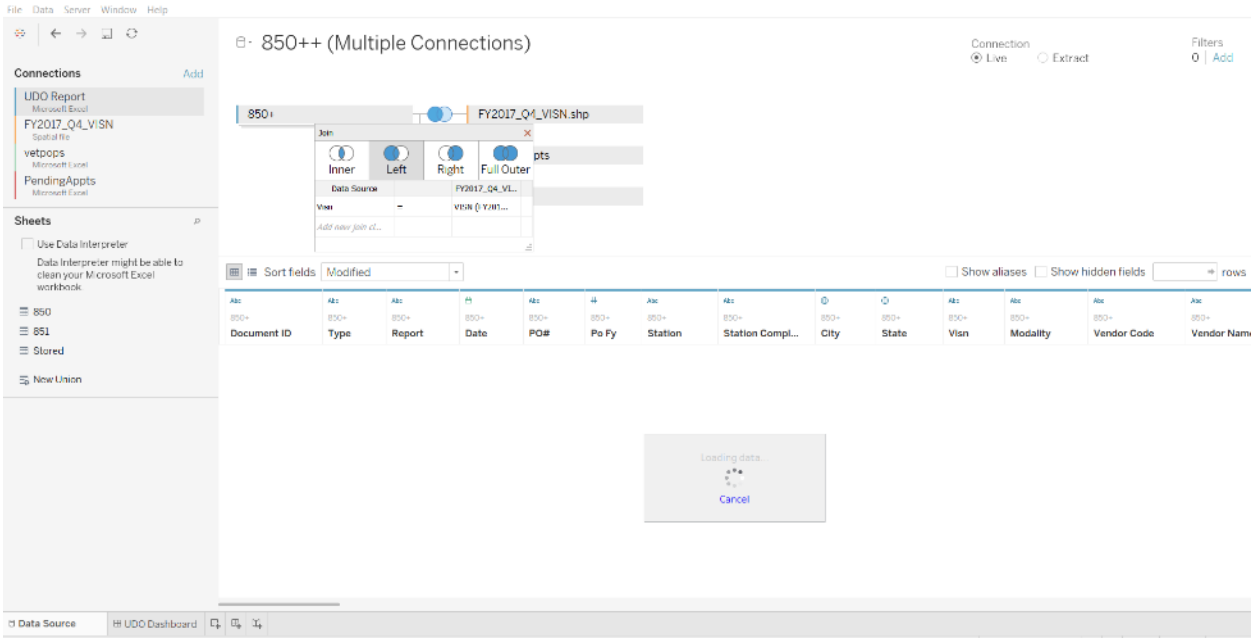
C.4 TABLEAU SCREENSHOTS

C.4.1 Tableau Front End Visualizations

C.4.2 Tableau Backend Data Join

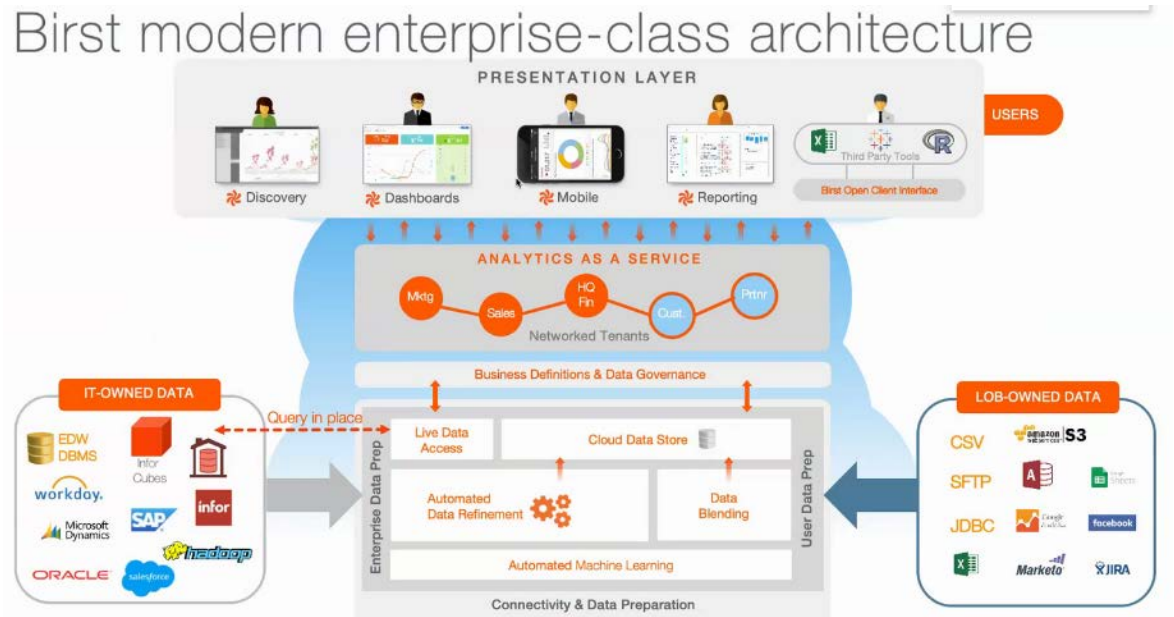


C.4.3 Tableau Backend Data Join Detail

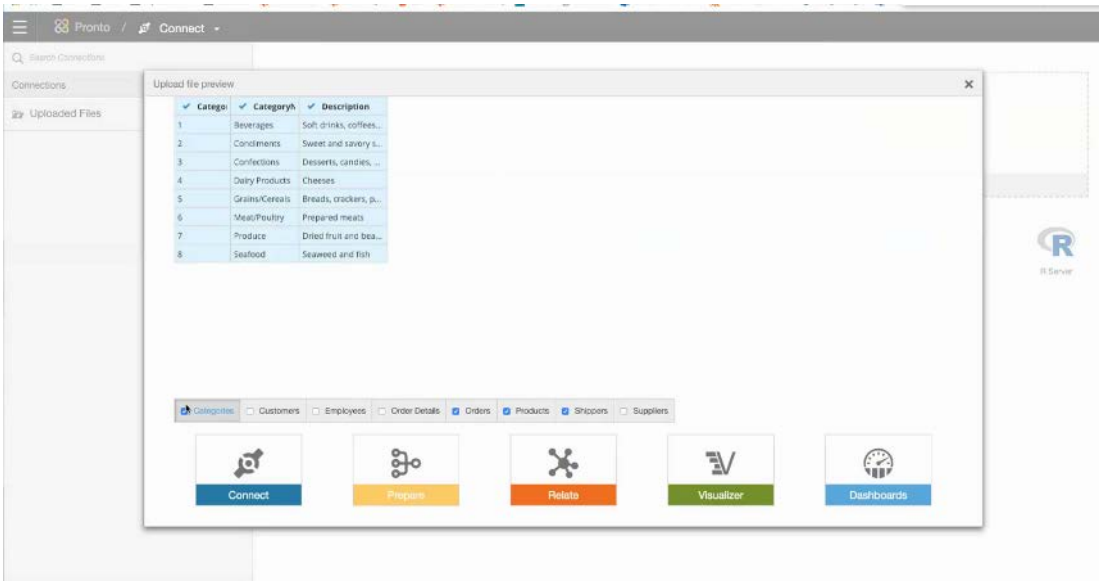


C.5 BIRST SCREENSHOTS

C.5.1 Birst Infrastructure View

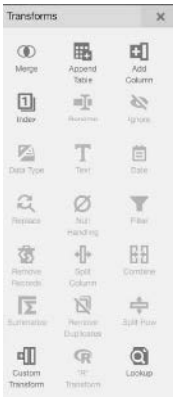


C.5.2 Birst Backend Data View



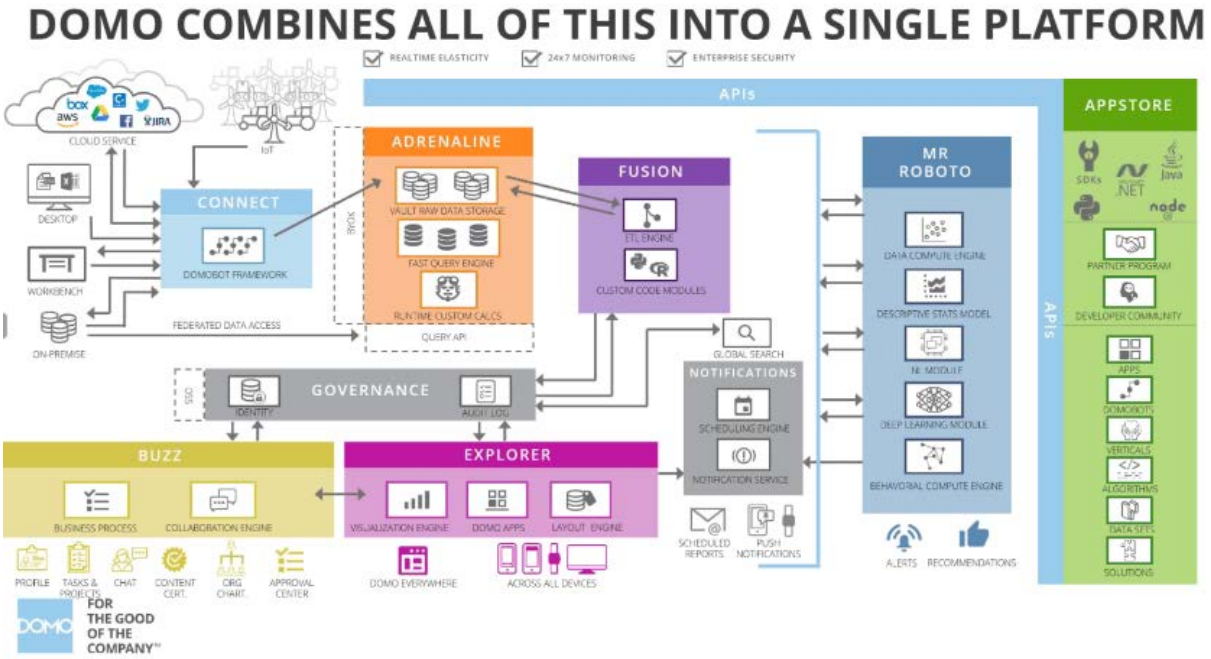
C.5.3 Birst Data Transformation

C.5.4 Birst Data Transformation Options

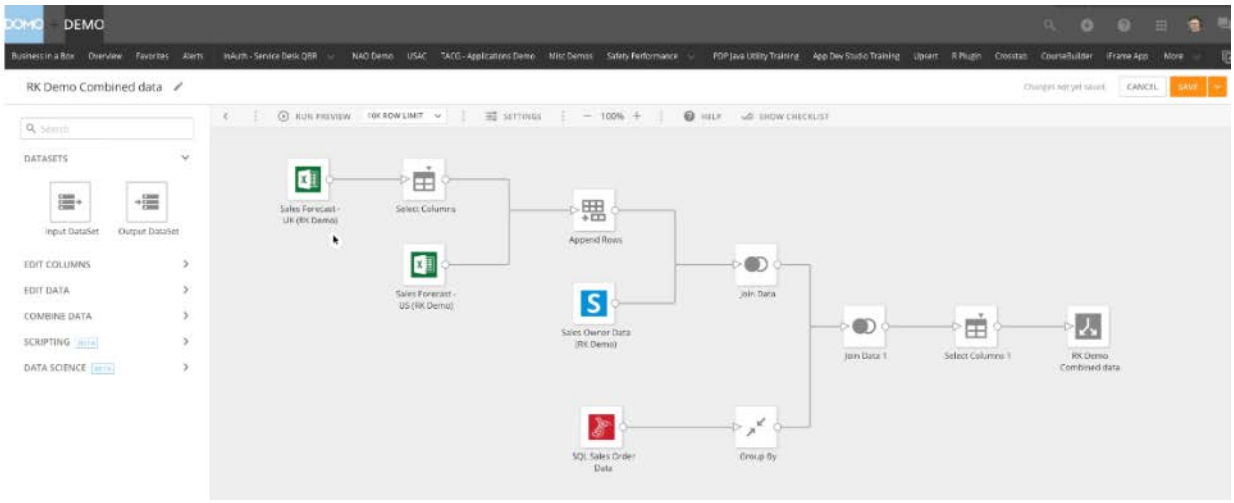


C.6 DOMO SCREENSHOTS

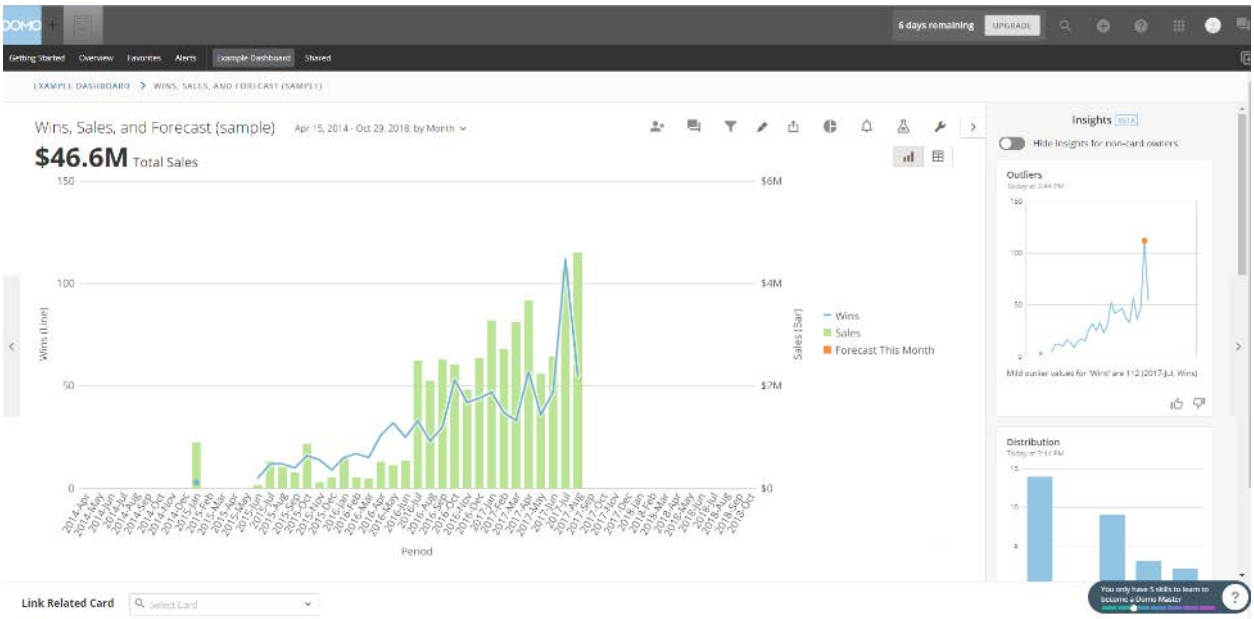
C.6.1 Domo Infrastructure



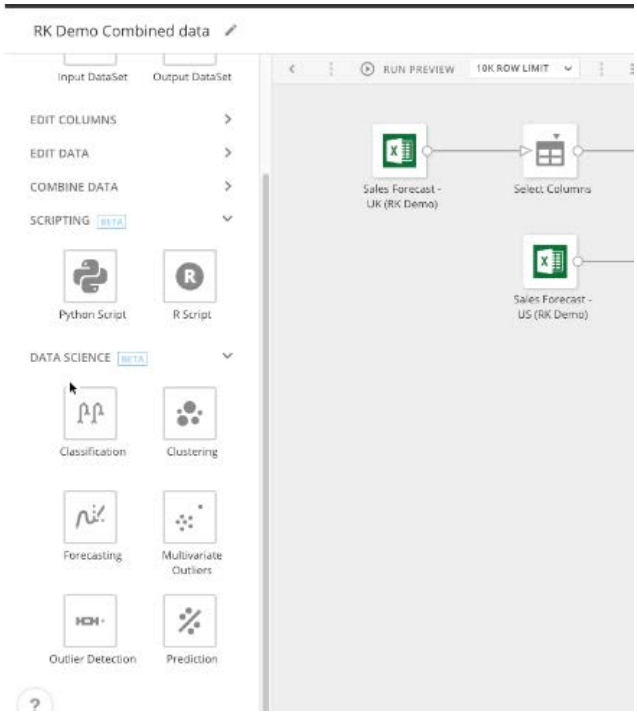
C.6.2 Domo Backend Data Modeling



C.6.3 Domo Frontend UI

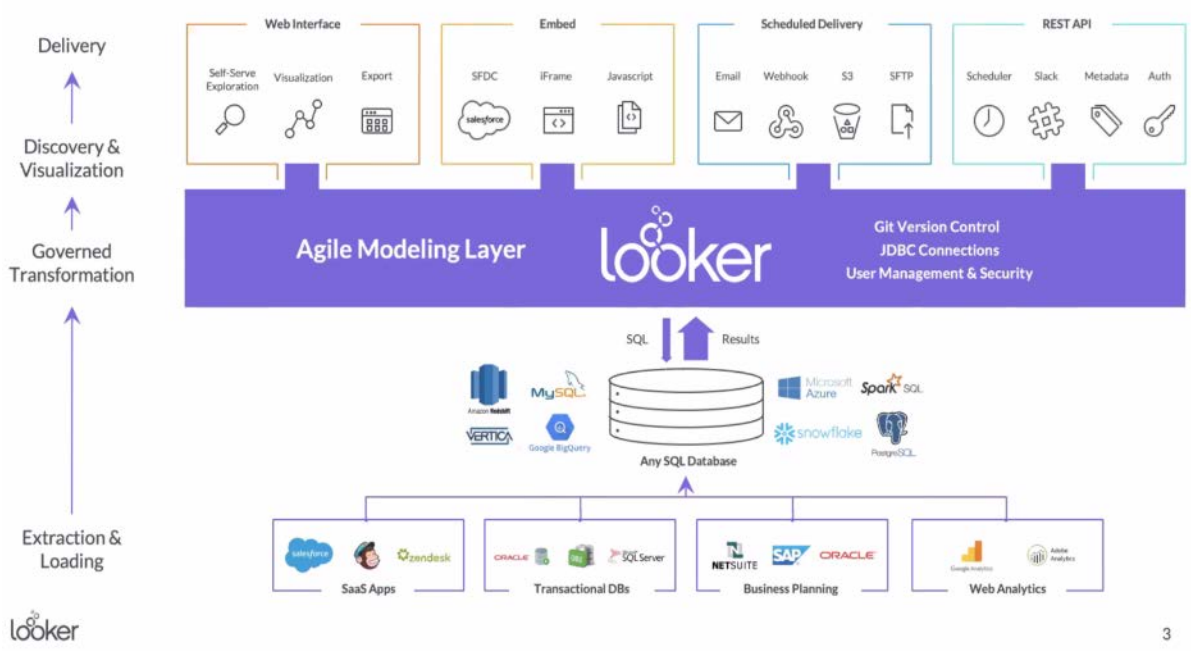


C.6.4 Domo Backend Data Modeling Options



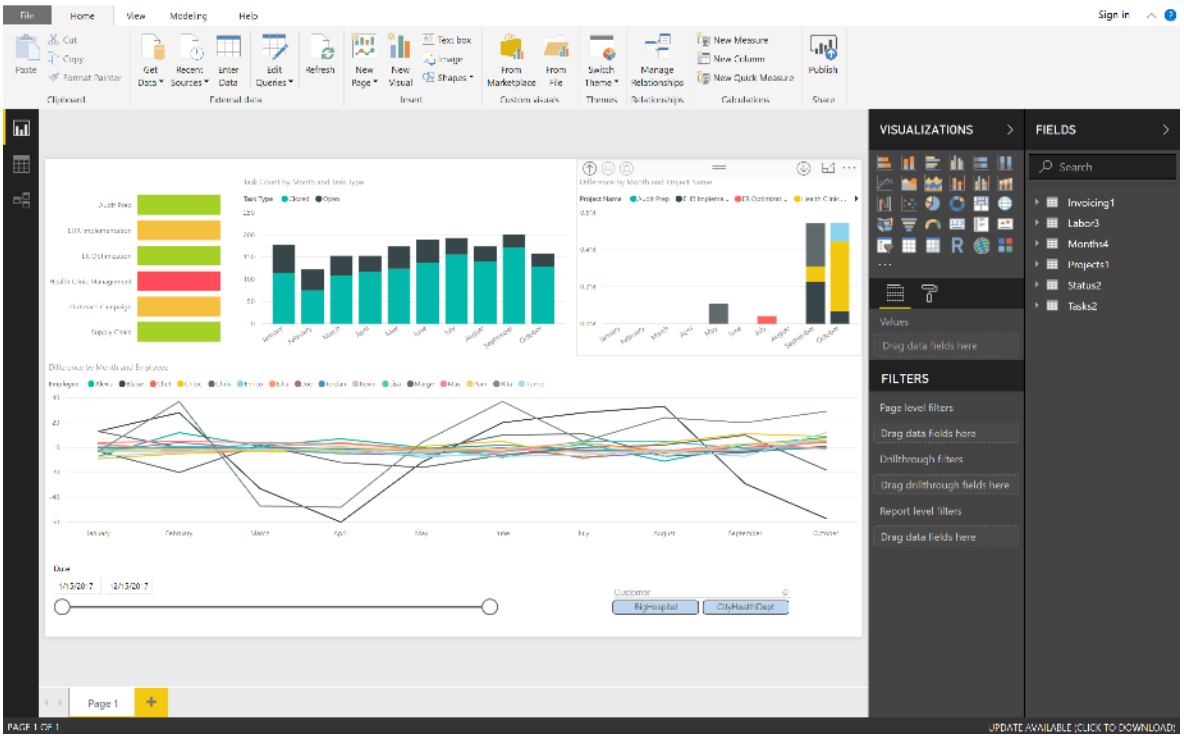
C.7 LOOKER SCREENSHOTS

C.7.1 Looker Infrastructure

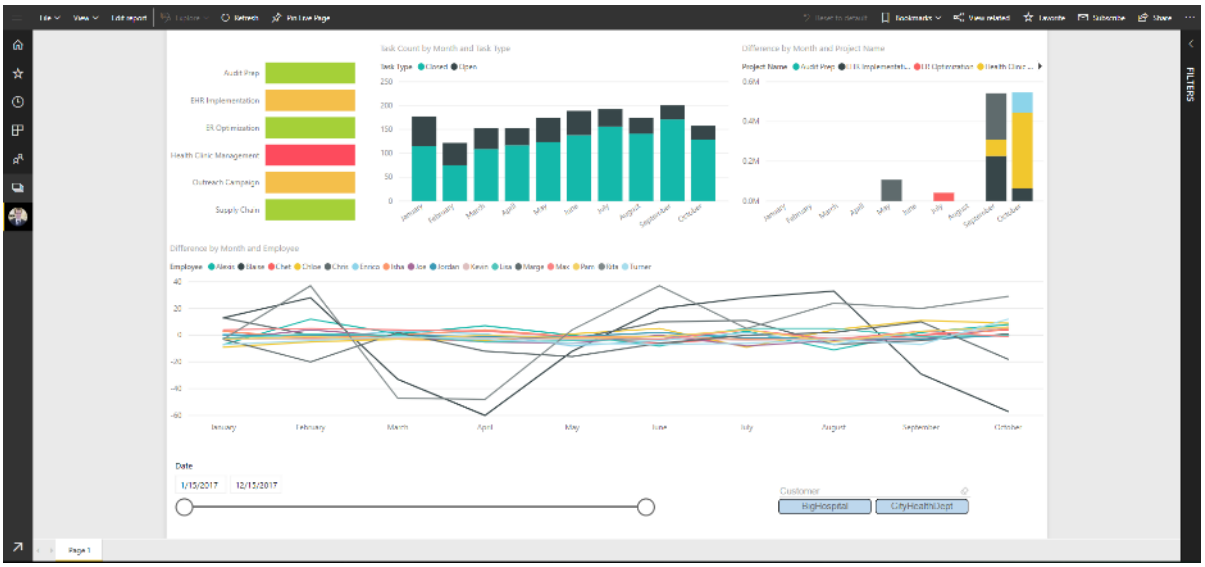


C.8 POWER BI SCREENSHOTS

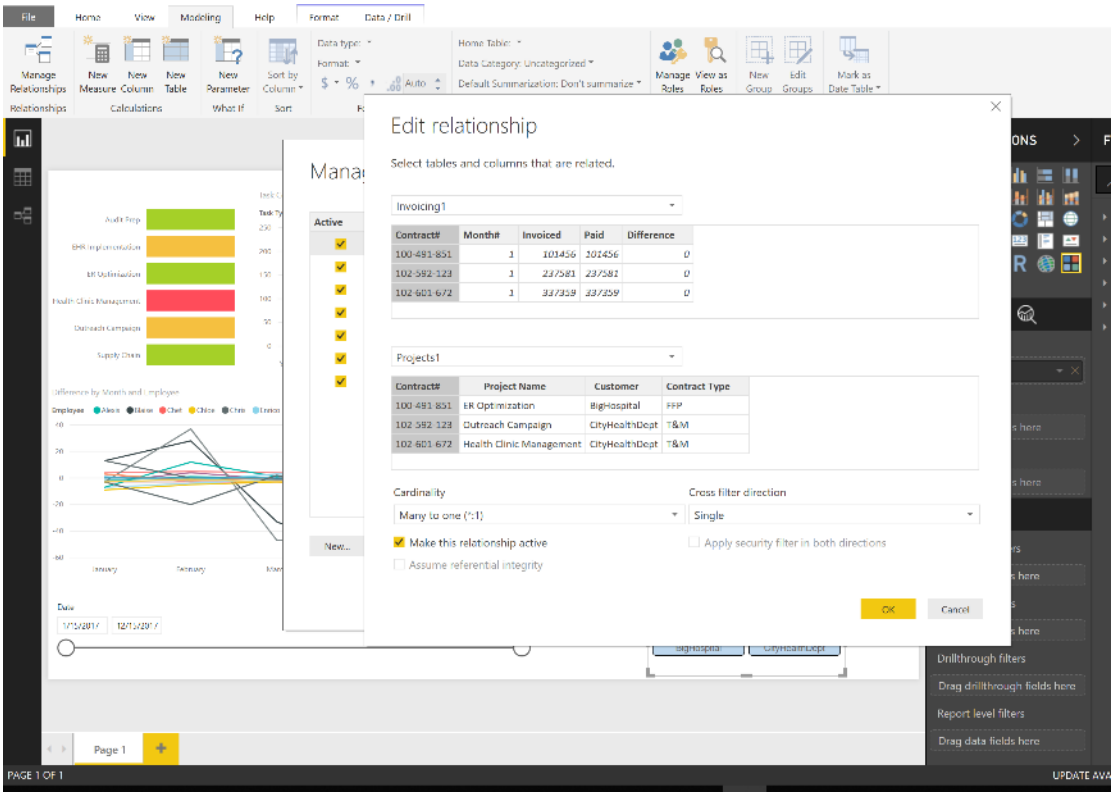
C.8.1 Power BI Backend Visualization Manager



C.8.2 Power BI Frontend Visualization

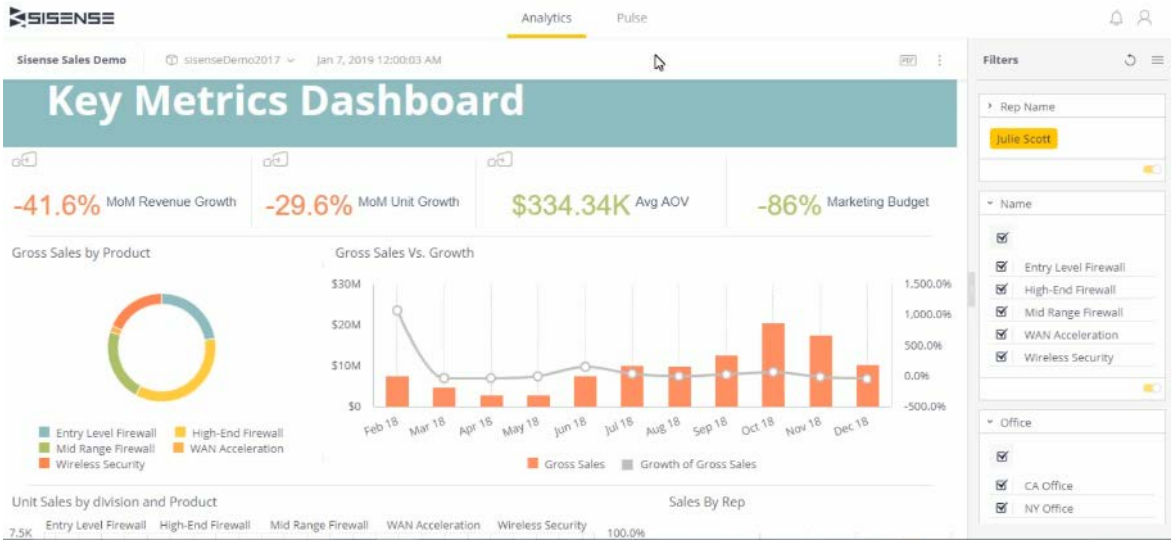


C.8.3 Power BI Relationship Manager

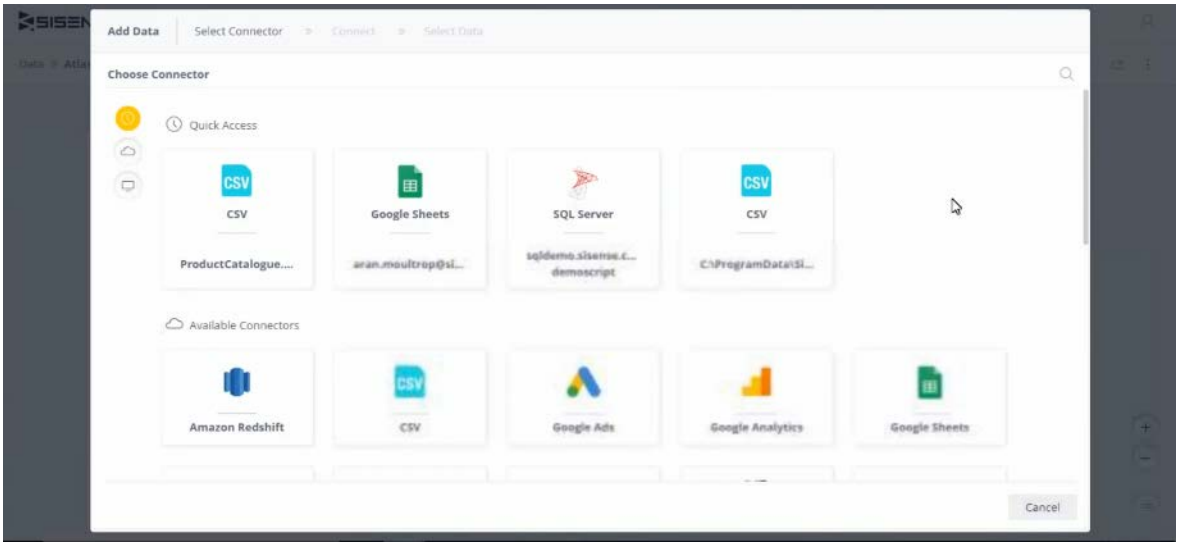


C.9 SISENSE SCREENSHOTS

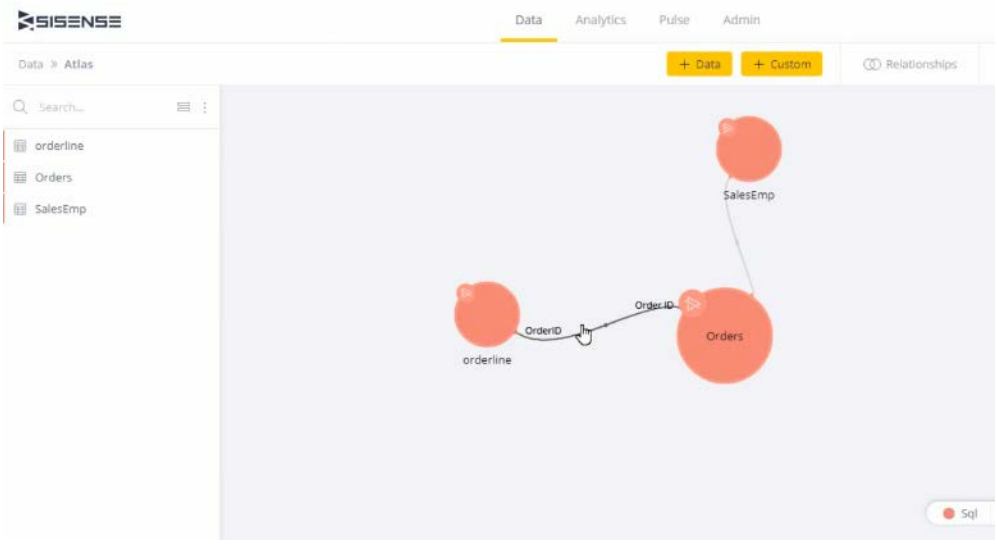
C.9.1 Sisense Frontend View



C.9.2 Sisense Data Connectors



C.9.3 Sisense Data Modeling



C.9.4 Sisense Data Modeling with Different Servers

RFTO #10: Developing Strategies for the Harmonization of AHRQ Data Tools (Data Tools)

Instructions for Reviewing the Final Style Examples

March 15, 2019

CONTENTS

Revision History

Date	Version	Description	Author(s)
February 8, 2019	Instructions for Reviewing the (Alpha) Style Examples	Report that provided guidance to the AHRQ review team on how to navigate through the Alpha version of the Style Examples	Atlas Team
February 22, 2019	Instructions for Reviewing the (Beta) Style Examples	Report that provided guidance to the AHRQ review team on how to navigate through the Beta version of the Style Examples	Atlas Team
March 15, 2019	Instructions and background for reviewing the Style Examples	Report that provides guidance to the AHRQ review team on how to navigate through the Style Examples and provides background on the development of Style Examples process	Atlas Team

NOTE: The following document was produced as part of AHRQ's Data Tools Harmonization Project Assessment Phase. The original document contained links to wire frames that are not available for viewing at this time. Those links have been replaced with the following highlighted text:

(link unavailable at this time)

1. INTRODUCTION

The Agency for Healthcare Research and Quality (AHRQ) provides several data tools and summary data tables that have dedicated user groups, but currently these tools are siloed and lack a common user experience (UX) and branding among them.

To better understand both the potential solutions available to support the harmonization of the AHRQ digital tools user interfaces (UI) and basic query functionality of these data tools, the Atlas Team conducted an environmental scan to gather and prioritize user needs and requirements. Based on the information that has been gathered, initial style layouts have been developed.

The intention of this document is to help you navigate and review the Style Examples. These Style Examples and the resulting prototypes are conceptual and exist to support AHRQ staff in evaluating the feasibility of a harmonized data tool portal and data visualization products. The prototypes will not be for public view or use.

1.1 Style Examples Development Process

Based on the results of the Environmental Scan phase, where high-priority requirements and desired improvements were identified by AHRQ, the Atlas Team created these Style Examples. The Style Examples underwent three iterations to make sure that we sufficiently incorporated key feedback from the main stakeholders and diverse users.

Each iteration allowed AHRQ data tool POCs and representatives to review and ensure that feedback items are communicated and, when appropriate, are implemented in the succeeding version of the Style Examples.

- **Alpha**: The wireframes presented during this version provided the initial concept for a harmonized data portal, user interface, and functionality of the data tools. The goal for the Alpha version was to establish a common look and feel so users can focus on understanding all the data tools AHRQ provides and how to find the right data for their research needs.

  Note: AHRQ provided feedback on the Alpha Style Examples on February 15, 2019.

- **Beta**: Beta designs incorporated AHRQ's feedback on the Alpha designs as well as additional designs for individual tool landing and query pages. The features and concepts from the Beta designs that were excluded from the final style examples can be found in the Appendix.

  From February 25 – March 4, AHRQ POCs and representatives provided feedback during one-hour sessions with the Atlas Team. In addition, during this same time period, feedback sessions were also conducted with several external customers of the data tools.

- **Final:** The final style design – of which this document is a part – incorporates all previous key feedback and will be the basis for building out the prototypes. Please note that while the design layout is established at this point, the Atlas Team welcomes any

suggested changes to content and minor aesthetic changes and will incorporate them into the prototype.

In working with the Office of Communications and the Contracting Officer Representative (COR), the Atlas Team also implemented those AHRQ style guidelines deemed relevant by the COR; this included implementing the appropriate header and footer and using the prescribed logos and style guidelines (if specified) for AHRQ, the Department of Health & Human Services (HHS), and each individual data tool. In addition, the Atlas Team was provided guidance on the fonts to use for the style examples and prototyping. The Atlas Team was given flexibility by the COR to design the site features to best support the creation of prototypes that will enable AHRQ to evaluate the feasibility of a common data visualization product and a harmonized user interface. The design concepts developed were based in part on the requirements gathered, the precedent established in the creation of existing user interfaces like the MEPS Summary Tables Household Component website, and the results of research on other agency data websites (e.g., https://data.hrsa.gov/). For more information on the results of this research and the process and result of gathering user requirements, please refer to the AHRQ Data Tools Environmental Scan Results Report (February 4, 2019).

1.2 Guide to Reviewing the Style Examples

The key content in this instructions document to support your review of the Final Style Examples are:

- **Style Example Link** –(link unavailable at this time)

  Note: InVision is the program that we use to allow designs to be interactive as well as to allow feedback to happen on the designs. More information about this link, the process of using it, and **recommended format for providing feedback on the Final Style Examples** is presented in Section 2.

- **Style Screenshots & Annotations**

  Section 2.1 contains screenshots of the style examples along with annotations to give background to our design decisions and the intended functionality. They will also indicate what is currently clickable on the InVision images.

- **Comparison with Existing Data Tool**

  Section 3 contains images that provide a comparison between each harmonized data tool's landing page design and the current version of the data tool site with key content highlighted to show how they may be represented in the proposed harmonized UI.

- **Style Examples + User Requirements**

  Section 4 contains annotated Style Example images that connect the user requirements detailed in the Environmental Scan Report to these designs. A complete list of the user requirements is also included.

2. STYLE EXAMPLES

To access and review the style examples, please go to this link: (link unavailable at this time) (it may take a few seconds to load).

Please note that, aside from the clickable sections in some of the screens that you'll be seeing (as noted by the hand pointer symbol), this version is solely made to provide the approximate look, feel, and basic flow of the proposed harmonized portal and data tools and should not be evaluated as a fully-functioning product.

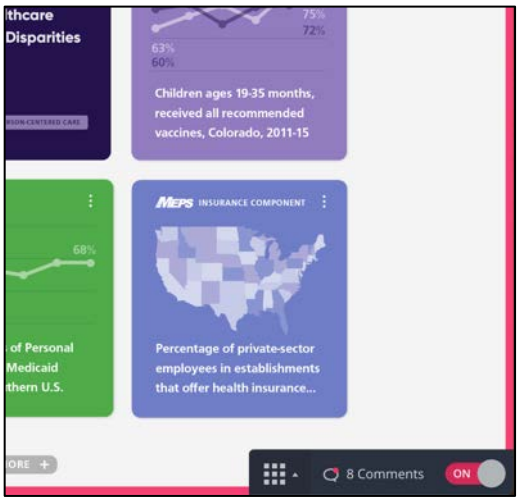
2.1 Screens in the Style Examples

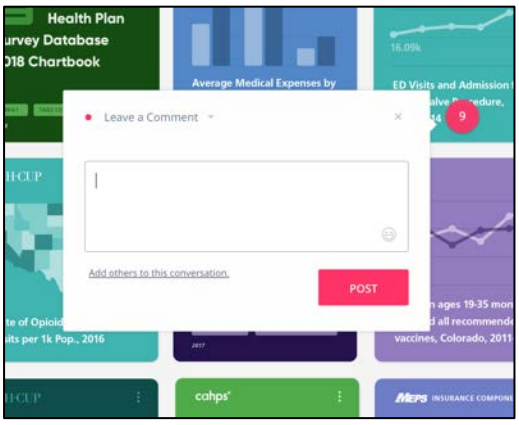
You will find several screenshots of the wireframes to assist you while navigating through the Style Examples in InVision. All clickable fields in InVision are highlighted in red below, followed by descriptions of each type of screen and relevant key sections.

2.2 Recommended Feedback Format

In order to consolidate feedback on specific features of the proposed harmonized user interface, we recommend that any comments be added directly to the InVision link provided above. To this end, please follow the below instructions:

1. Enter the prototype by clicking the InVision link provided above
2. Browse through the prototype until you find a feature for which you'd like to provide feedback
3. Move your cursor to the comments tab in the lower right-hand corner of the browser window (see 1st image at right)
4. Toggle the comments switch to the "on" position
5. Click on the area of the site for which you'd like to provide feedback, and type your comment (see 2nd image at right), then post
6. Toggle the comments switch (lower right-hand corner of browser window) to "off" to continue browsing





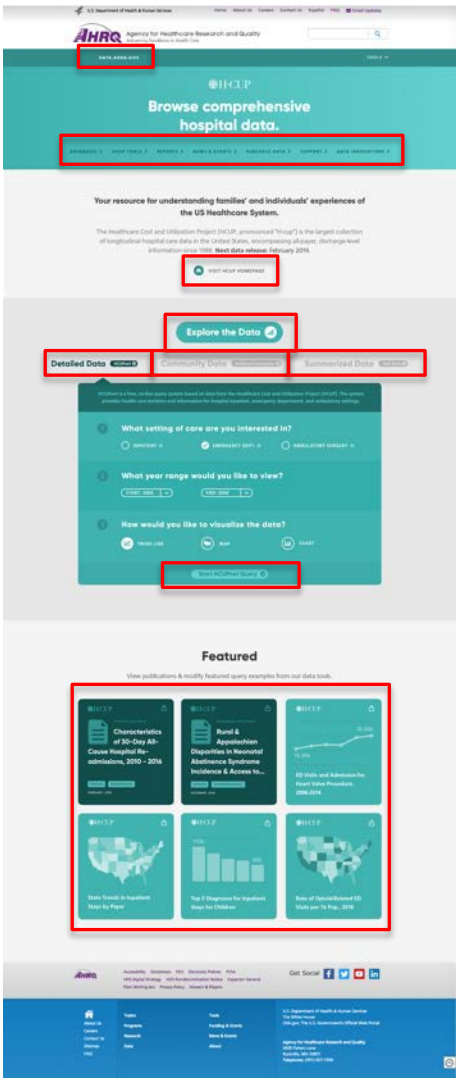
AHRQ Data Tool Main Portal

Upon entering the Style Examples via the link provided above, you will see the Data Tool Portal page. Named "data.AHRQ.gov", this unified interface creates an exploratory, conversational tone that enables users to consider potential constraints to their research question while providing them with several ways of understanding and engaging with each of AHRQ's tools:

Data Tools: Clicking on any of the data tool tiles will take you to the landing page of the data tool that you selected. Each retains the branding and style of the data tool, while also including a brief call to action of its utility and key characteristics.

Featured: A list of data visualizations that AHRQ will wish to feature will help to illustrate the potential queries available by the data tool; clicking on any one of these featured data visualizations will send them to a pre-populated query page, allowing the user to customize that specific query. In addition, this section contains featured reports, which enables users to see and access publications from across all the data tools. In the upper right corner of each of the query and report tiles is a share button that provides a "copy link" to clipboard, a "share" via email, and a "download" option for quick use.

Note: In the Style Examples, while each tile under the Featured section will bring the user to a sample Query page UI, that page might not represent the information on the original tile that was clicked.



Data Tool Landing Page

This is a general representation of a landing page for a data tool. The layout and functionality are largely the same as that of the portal page for the purpose of maintaining a consistent user experience; however, there are two exceptions on the landing pages: a description section at the top describes the tool in detail, and an "Explore the Data" section allows users to enter the data tool either directly or via an introductory query form. Please note that some features previously featured in the Beta Style Examples (e.g., data download buttons) were deprioritized for the Final Style Examples and, correspondingly, the prototype; these deprioritized features can be found in the Appendix.

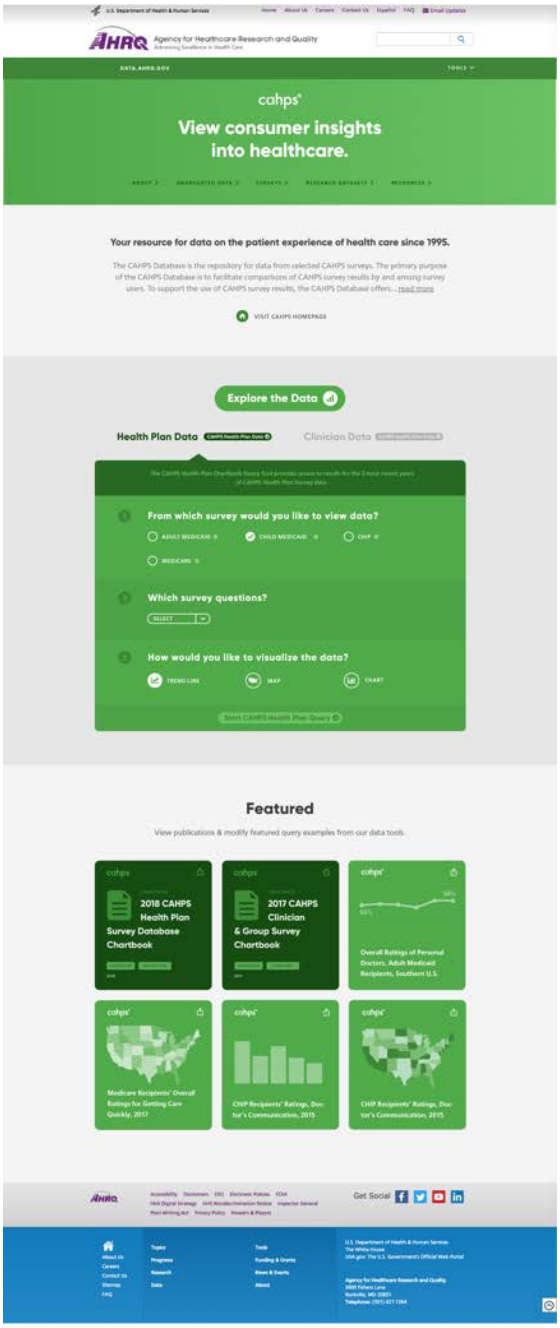
Description Section: this section contains a longer description of the tool and its offerings as well as a link to "Visit Homepage", which will direct the user to the original version of the site.

Explore the Data: This data section is intended to provide users of all types with a gateway to the tool(s) available for this dataset. Pro and intermediate users with a firm understanding of the tools available can enter via the "Explore the Data" button at the top, or any button below for specific data tools (e.g., HCUPnet, HCUPnet Community data, or HCUP Fast Stats). For novice users, each of the data tools is accompanied by pithy, plain language descriptions (e.g., the HCUPnet is described as "Detailed Data," Fast Stats as "Summarized Data") and a query form; from top to bottom, this query form first provides a description of that particular tool's function, and a 3-4 step process for beginning a query, each written as a question. Users can select the variables they are interested in and then send their selections to the Query page by clicking the "Start Query" button.

Featured: When on a tool landing page, these featured reports and visualizations would only be from that specific tool. The share buttons in the right corner of each of these tiles allow users to share, copy the link to, or download the report or query. For reports, when users click on the tile they are sent directly to the current page for that publication.

Note: In the Style Examples, only a select few of these tiles are clickable. This is just to provide a representation of what a subsequent screen could be.

The following are screenshots of all the data tools' landing pages.



Harmonized CAHPS Landing Page (with Health Plan Data Query selected)

Harmonized CAHPS Landing Page (with Clinician Data Query selected)

Please note that during the Beta Style Examples feedback session for CAHPS, the POC requested that a query version of the CAHPS chartbooks for Clinician and Health Plan surveys be developed instead of the existing Online Reporting System for each. We are noting this request here so that it can be a considered as part of a full implementation when AHRQ makes the decision to develop the actual harmonized site.

Harmonized HCUP Landing Page (with HCUPnet Query selected)

Harmonized HCUP Landing Page (HCUPnet Community Query)

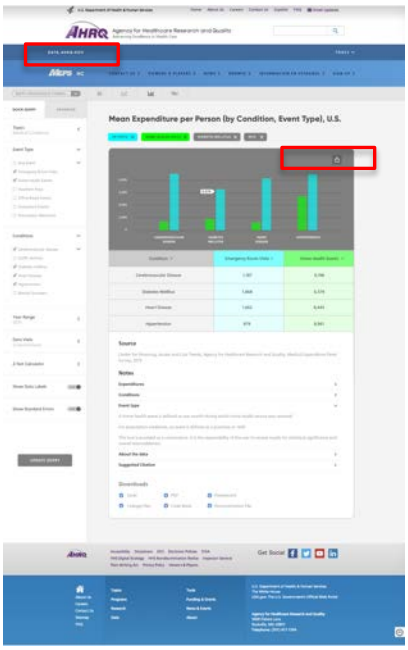
Harmonized HCUP Landing Page (HCUP Fast Stats selected)

Harmonized MEPS HC Landing Page

Harmonized MEPS IC Landing Page

Harmonized NHQDR Landing Page

UI Query Interface Screen



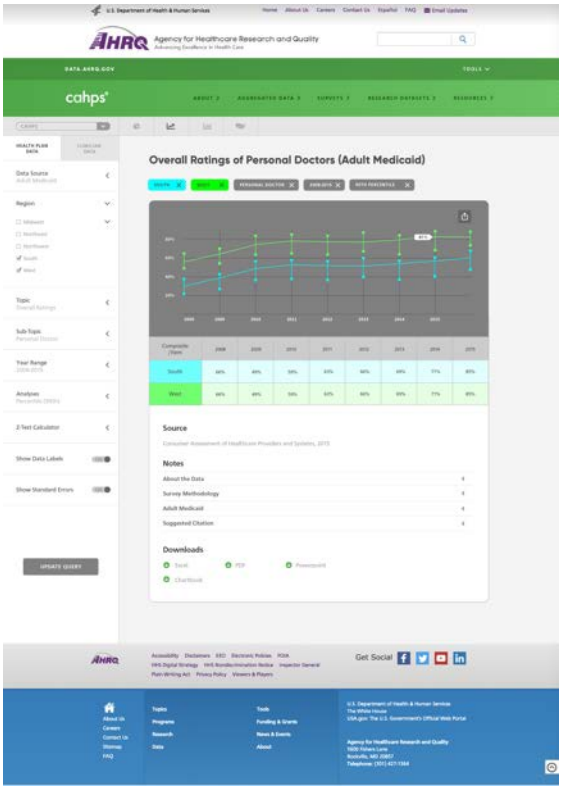
Tool & Visualization Selection: Below the navigation bar is a new row of buttons including “Tool” and the icons for data visualization types. The purpose of the “Tool” dropdown menu is twofold: it allows users to quickly switch between the query interface of any of the data.AHRQ.gov tools; it also allows users to enter the query page directly from any landing page via the “Explore the Data” button, after which they need to specify which specific tool (e.g., HCUPnet or HCUP Fast Stats) they’d like to use in the Variable Selection section.

Tool & Variable Selection: The left-side panel is envisioned for tool & variable selection. At the top are tabs, each representing a different tool. In MEPS HC, for example, these tabs (represented at left) are described as (1) “Quick Query” for the Summary Tables tool, and (2) “Advanced” for the MEPSnet tool; the tabs for other tools, however, are named both explicitly (e.g., HCUPnet, HCUP Fast Stats) and with a high-level description (e.g., Detailed Data, Summarized Data).

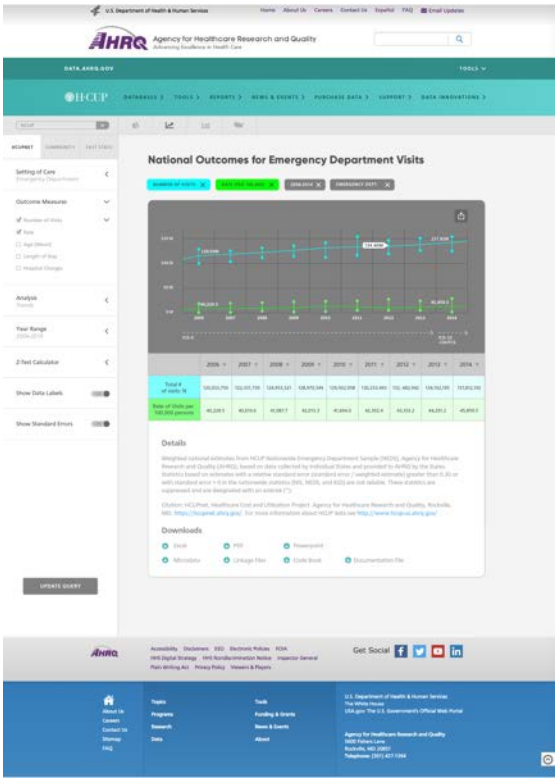
Visualization: The primary space for this page is reserved for the visualization. Above the visualization, the selected variables will appear. Additionally, users can opt to build a comparison, share, or download the visualization from the navigation above.

Below the visualization will be the applicable data table or tables and the relevant footer information.

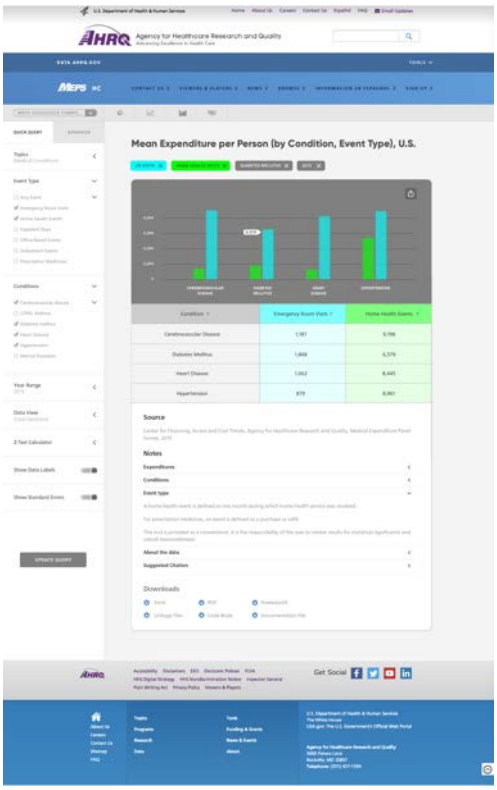
The following are screenshots of each data tool's UI Query Interface Screen.



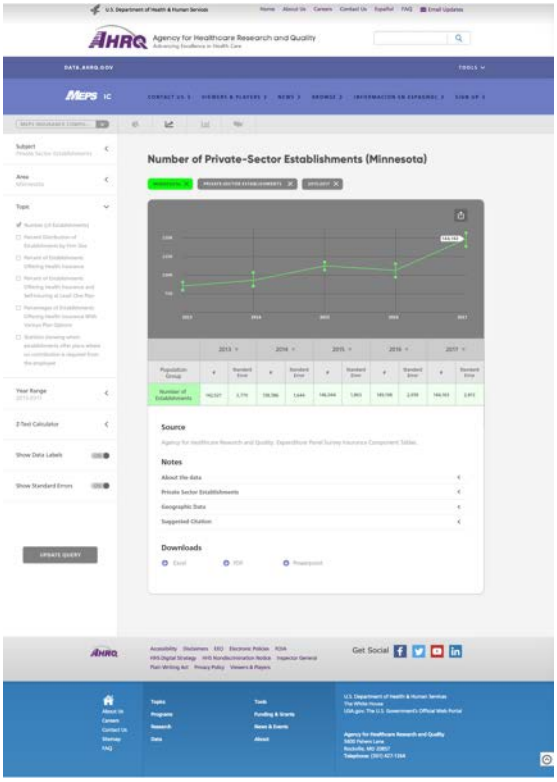
Harmonized CAHPS Query Page (with Health Plan Data Query selected)



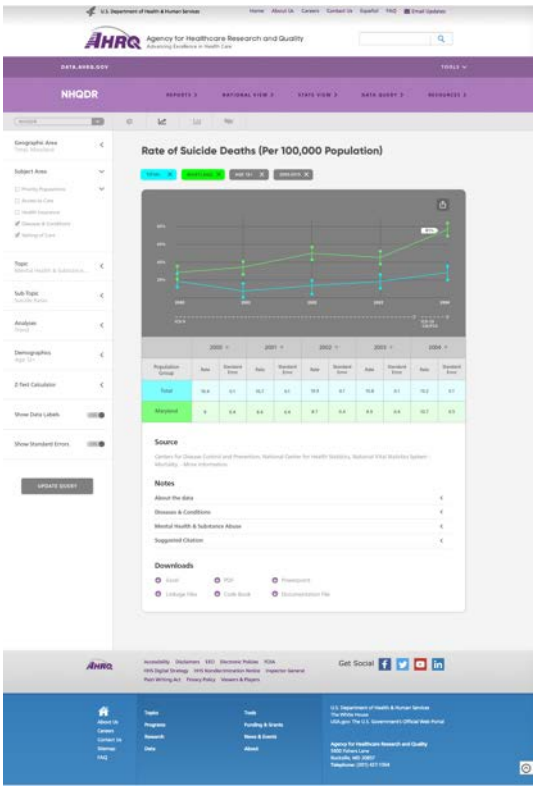
Harmonized HCUP Query Page (with HCUPnet Query selected)



Harmonized MEPS HC Query Page (MEPS HC ST or "Quick Query)



Harmonized MEPS IC Query Page (with MEPSnet IC Query selected)



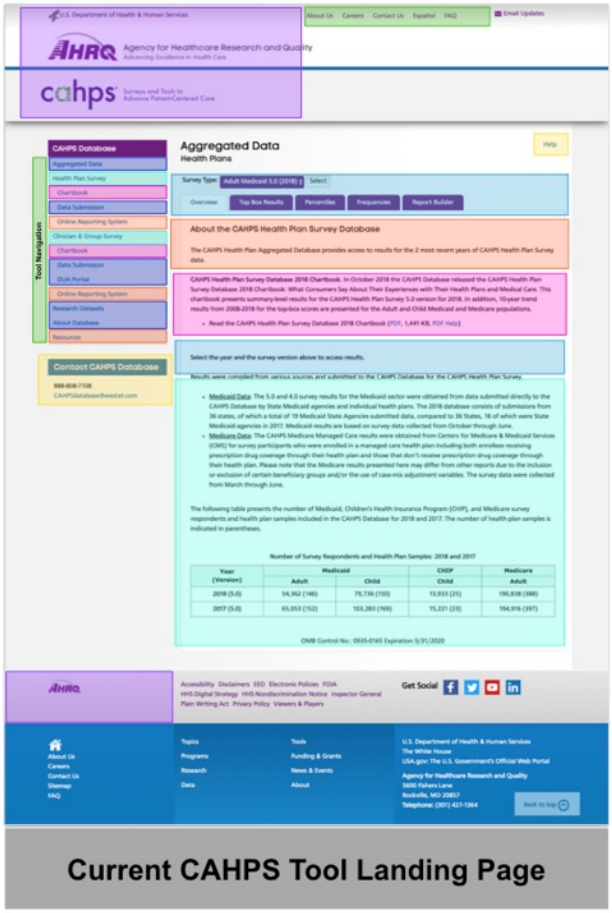
Harmonized NHQDRnet Query Page

3. COMPARISON WITH EXISTING DATA TOOL

The following graphics provide a general comparison between the current version of the data tool and the Style Example of that harmonized data tool design; key content is highlighted to show how they will be represented in the harmonized version of the data tool's landing page in the future. **Please note that the Style Examples represented here do not represent all of the gathered requirements, but rather only those that were selected by AHRQ during the Beta Style Examples feedback sessions for a minimum viable product as needed to develop the prototypes.**

Below is the content key to use for all of the following content comparisons between the current AHRQ tool landing pages and the proposed harmonized versions.

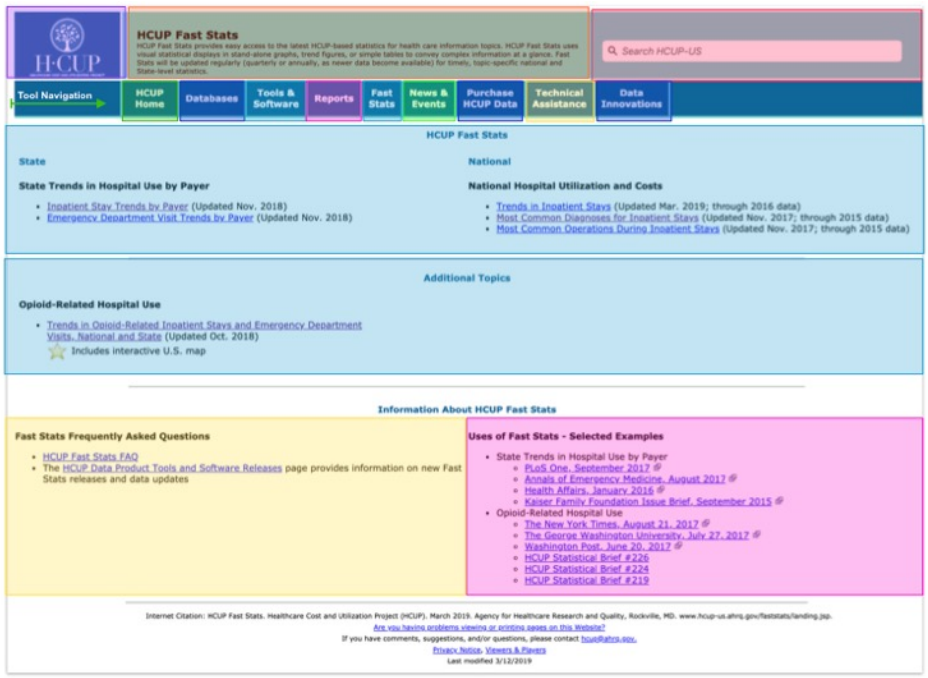
3.1 CAHPS



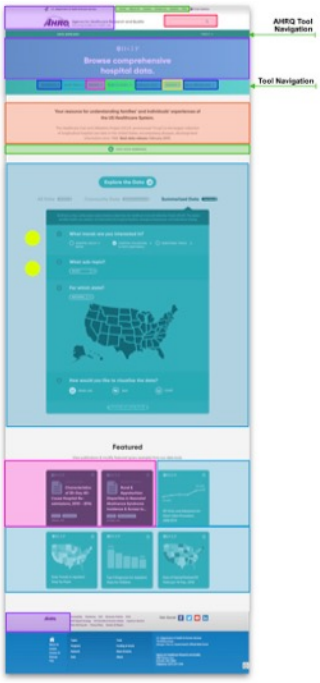
3.2 HCUP Fast Stats

Content Key:

Branding	Query Tool	Search	News	Methods	Data
Navigation	Support	Publications	Topics	About	Customization



Current HCUP Fast Stat Tool Landing Page



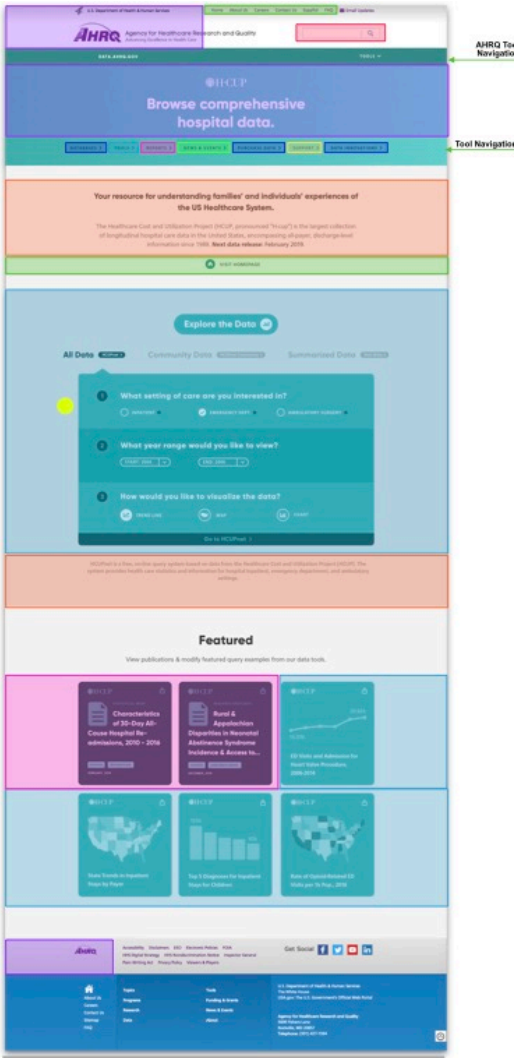
Harmonized HCUP FS Tool Landing Page

3.3 HCUPnet

Content Key:

Branding	Query Tool	Search	News	Methods	Data
Navigation	Support	Publications	Topics	About	Customization

Current HCUPnet Tool Landing Page

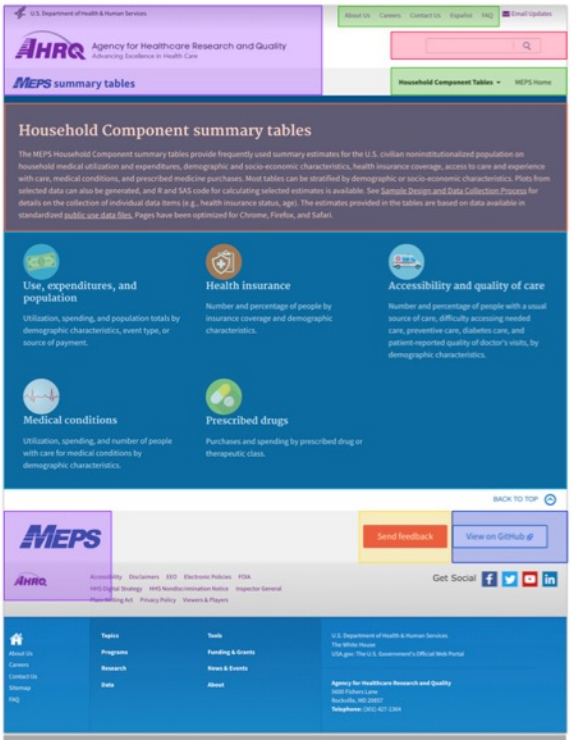


Harmonized HCUPnet Tool Landing Page

3.4 MEPS HC Summary Tables

Content Key:

Branding	Query Tool	Search	News	Methods	Data
Navigation	Support	Publications	Topics	About	Customization



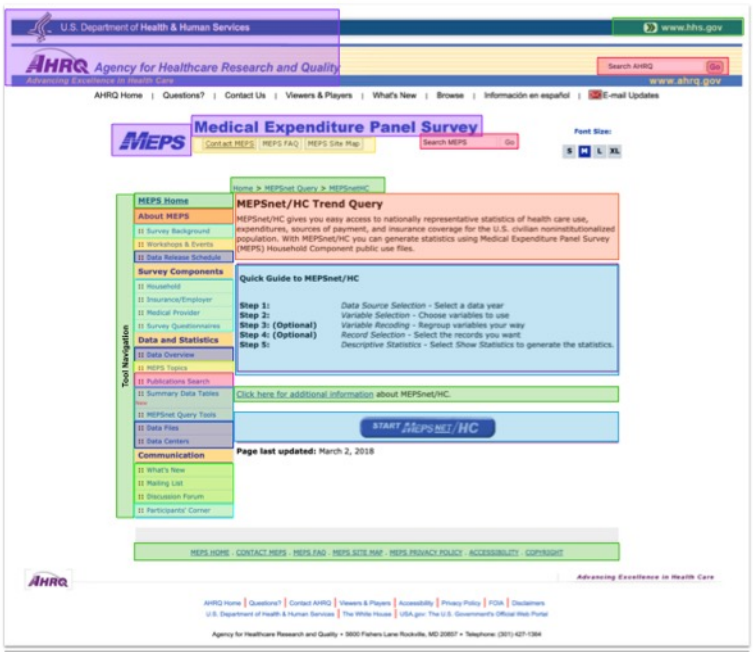
Current MEPS HC Summary Tables Tool Landing Page

Harmonized MEPS HC Summary Tables Tool Landing Page

3.5 MEPSnet (HC)

Content Key:

Branding	Query Tool	Search	News	Methods	Data
Navigation	Support	Publications	Topics	About	Customization



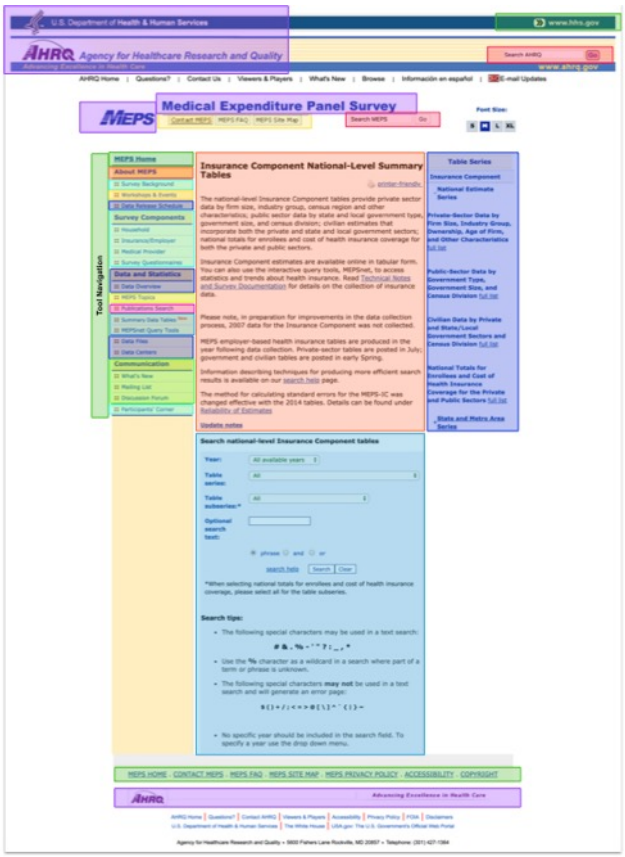
Current MEPSnet HC Tool Landing Page

Harmonized MEPSnet HC Tool Landing Page

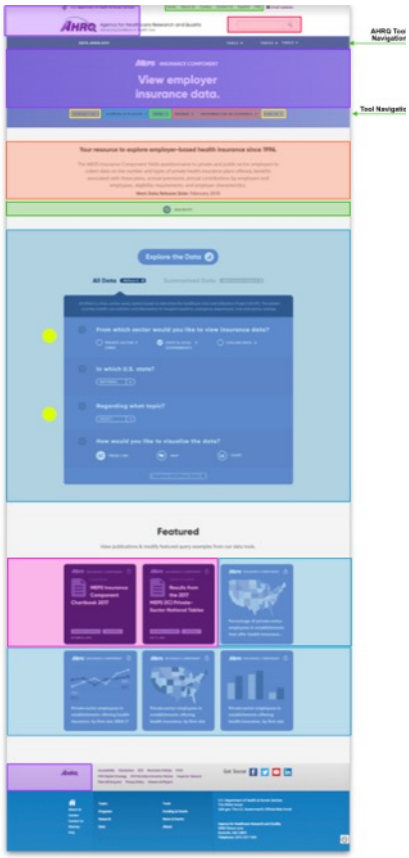
3.6 MEPSnet (IC)

Content Key:

Branding	Query Tool	Search	News	Methods	Data
Navigation	Support	Publications	Topics	About	Customization



Current MEPSnet IC Tool Landing Page

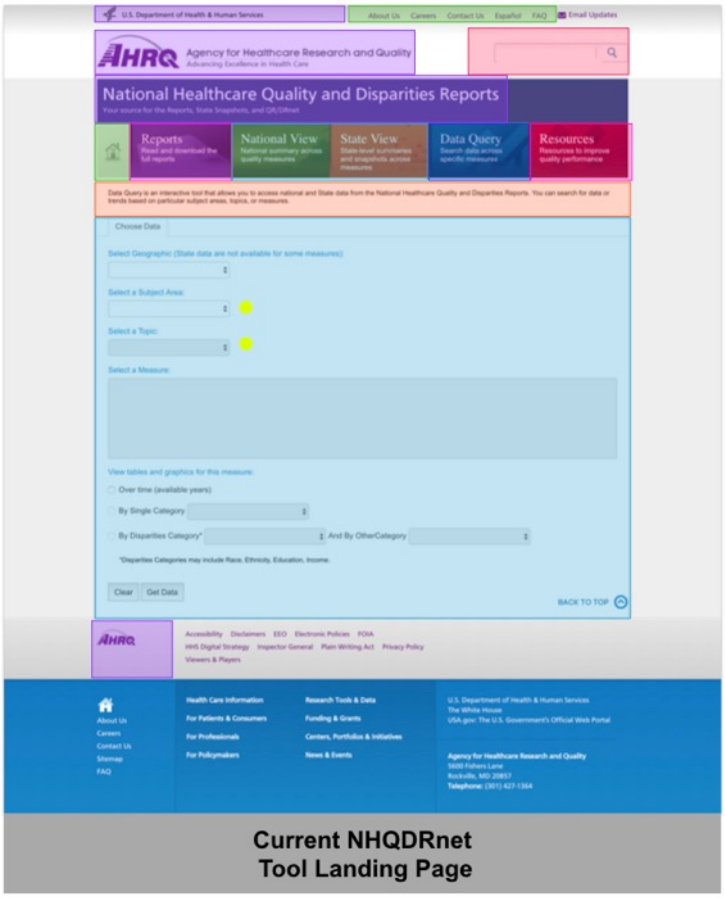


Harmonized MEPSnet IC Tool Landing Page

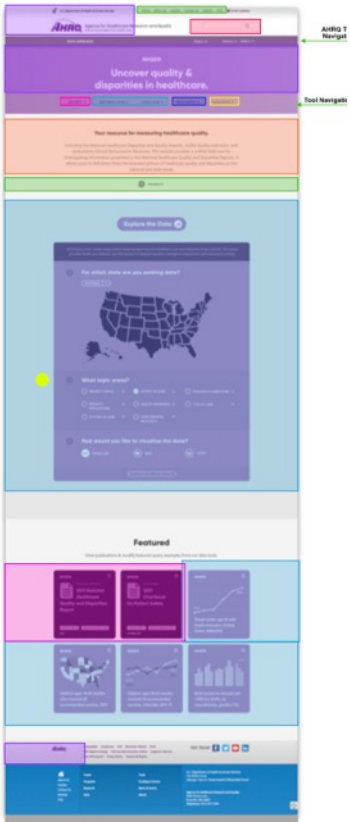
3.7 NHQDRnet

Content Key:

Branding	Query Tool	Search	News	Methods	Data
Navigation	Support	Publications	Topics	About	Customization



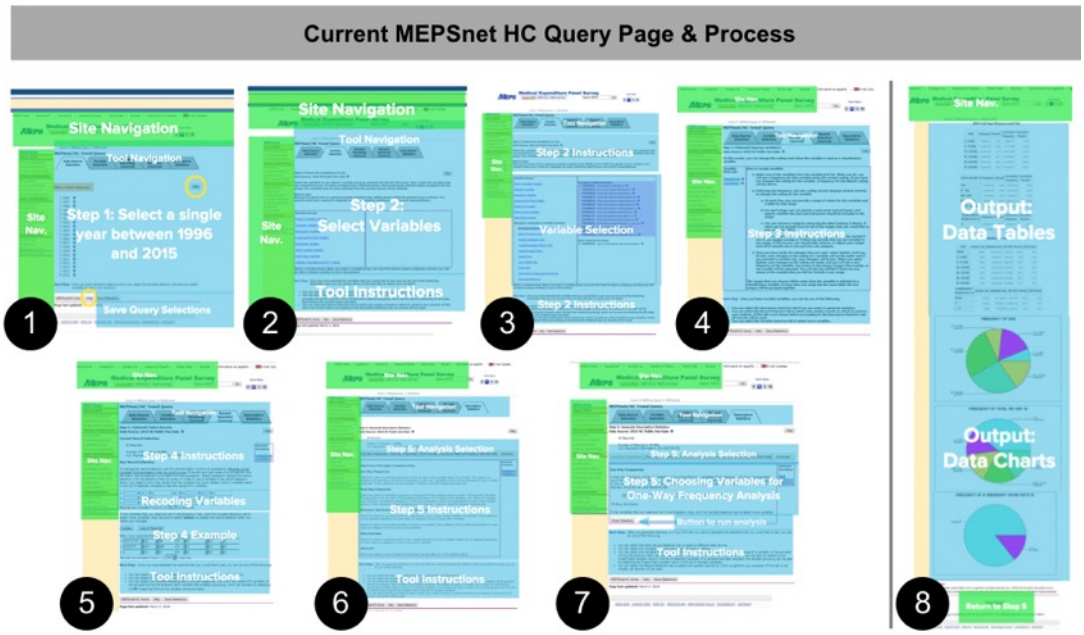
Current NHQDRnet Tool Landing Page

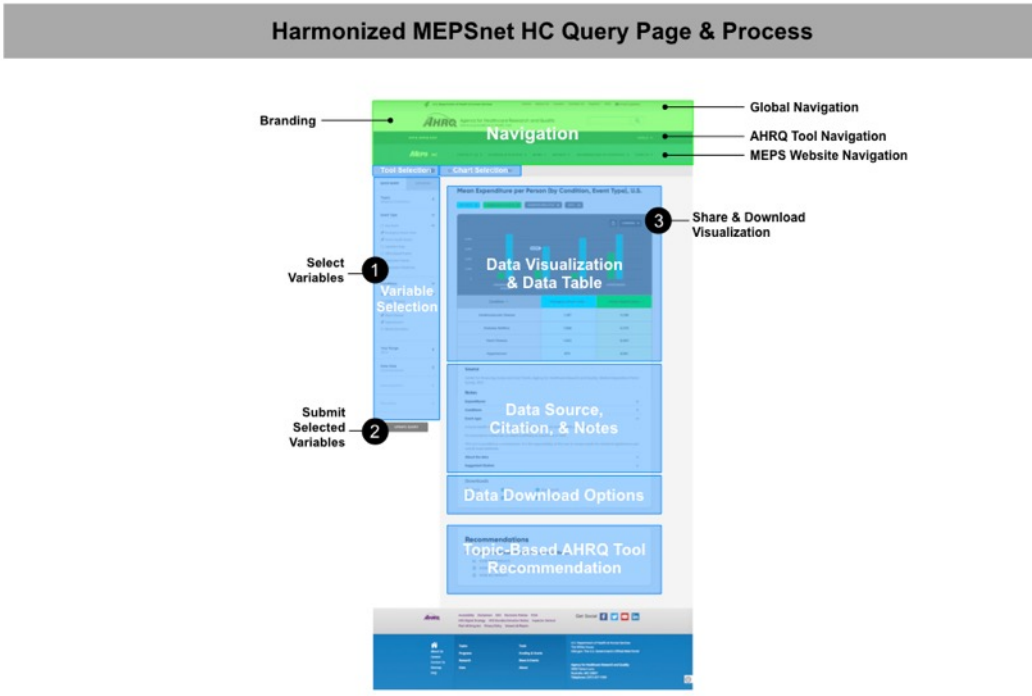


Harmonized NHQDRnet Tool Landing Page

3.8 Query Page Comparison

We selected MEPSnet HC to demonstrate how the data query user experience has been translated from the current data tools to the proposed harmonized design. The numbers in the black circles indicate consecutive user actions (e.g., buttons clicked or variable selection).

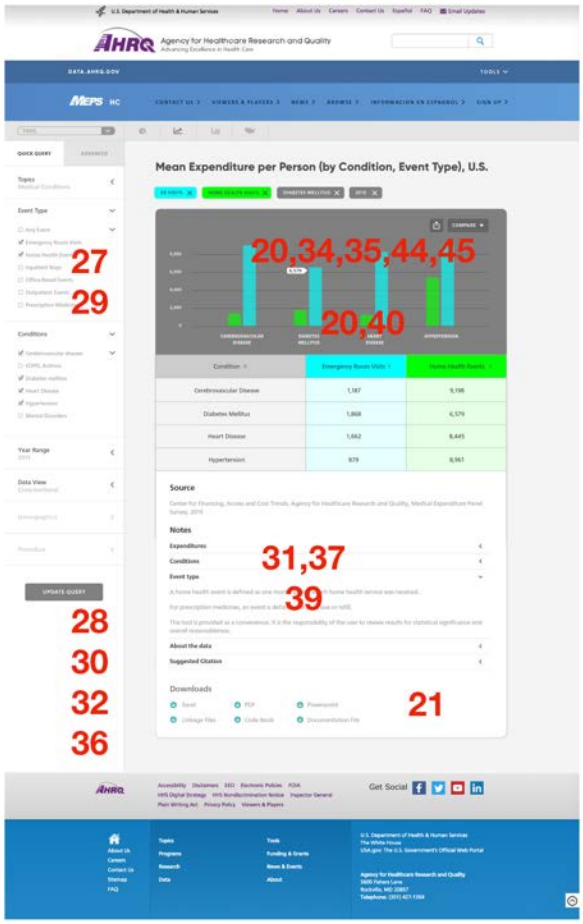




4. STYLE EXAMPLES AND THE USER REQUIREMENTS

The user requirements were used as basis for designing the style examples. In the wireframes below, we've mapped the user requirements, represented by numbers, to the design decisions that fulfill this requirement. The table following the wireframes contains a numbered list of the requirements included.

Please note that the purpose of the Style Examples is to represent a harmonized user interface and, as a result, some requirements are not included due either to their backend nature or their inclusion could be in the actual prototypes or in the application that will be actually developed and launched. Some other requirements and features were deprioritized for this specific phase of work and from the prototypes during the Beta Style Examples feedback sessions. Those deprioritized features are included in the appendix.


Agency for Healthcare Research and Quality
Mean Expenditure per Person (by Condition, Event Type), U.S.
20,34,35,44,45
20,40
27
29
28
30
32
36
31,37
39
21
Source
Notes
Downloads
Get Social

#	Category / Theme	User Requirement	Featured in Final	Not Featured / Applies only to the Backend
1	**Navigating the Data Tools with Ease**	"I need to be able to access basic query functionality on both my laptop and my mobile device in order to spontaneously reference information."		X
2	**Navigating the Data Tools with Ease**	"I need a user-friendly interface with highly streamlined & visualized content – preferably that can be viewed with minimal scrolling and scanning of extraneous text – in order to reduce the cognitive burden of browsing complex information."	X	
3	**Navigating the Data Tools with Ease**	"I need access to the archived version of original site in order to reference where I know information was while transitioning to a new interface."	X	
4	**Navigating the Data Tools with Ease**	"I need a similar site map and terminology to the original site in order to continue my use of the site relatively unabated."	X	
5	**Navigating the Data Tools with Ease**	"I need harmonized UIs that have consistent look, feel, and terminology in order to seamlessly transition between AHRQ's various data tools."	X	
6	**Navigating the Data Tools with Ease**	"I need a website that's optimized for all browser types in order to include its regular usage in my normal workflow."		X
7	**Navigating the Data Tools with Ease**	"I need simple, easy-to-understand terminology and descriptions of the tools' utility in order to quickly grasp which tool can be used to answer my research question."	X	
8	**Knowing Where to Start My Query**	"I need a unified data portal from which to access all of AHRQ's data tools in order to seamlessly integrate multiple data sources in my research workflow."	X	
9	**Knowing Where to Start My Query**	"I need to view or search what tools are available by topic area in order to determine which datasets can answer my research question."		X
10	**Knowing Where to Start My Query**	"I need a customizable search engine that enables me to sort/refine results – including by topic, data tool, date, and document type – in order to find potential answers to my research questions across multiple data tools and summary data tables."		X
11	**Knowing Where to Start My Query**	"I need the data tools I use to be optimized for popular search engines, so I can continue to engage with the datasets via my search engine of choice."		X
12	**Knowing Where to Start My Query**	"I need simple, intuitive URLs to access the data tools as a whole or individually in order to quickly navigate to."	X	
13	**Knowing Where to Start My Query**	"I need a query portal that allows me to engage by either clicking on prescribed topic areas or by filling a customizable search/query field in order to conduct a query of my choice."		X

#	Category / Theme	User Requirement	Featured in Final	Not Featured / Applies only to the Backend
14	**Knowing Where to Start My Query**	"I need the data query tools or summary tables and their function to be prominently featured in the UI in order to efficiently answer my research question."	X	
15	**Knowing Where to Start My Query**	"I need a simple, predictive search engine in order to find everything from estimates to survey questions."		X
16	**Knowing Where to Start My Query**	I need all branding and sub-branding of my tool to remain prominent in a user interface in order to retain the well-established reputation and userbase of my dataset.	X	
17	**Knowing Where to Start My Query**	I need a cover page that displays both program descriptions and available data in order to properly demonstrate the utility of my tool to users.	X	
18	**Understanding Methods & Variables**	"I need to be able to quickly reference the linkage files, codebooks, and documentation files in any tool in order to conduct my research efficiently."		X
19	**Understanding Methods & Variables**	"I need to be able to conduct customized searches of survey questionnaires and data dictionaries in order to fully understand the methods and variables behind a dataset."		X
20	**Understanding Methods & Variables**	"I need the relationships between CCS, ICD-9, and ICD-10 codes to be designated whenever applicable in order to understand gaps, discrepancies, or shifts in the data."	X	
21	**Accessing Data Files**	"I need to be able to download raw data files in order to conduct analysis in the statistical analysis software of my choice."	X	
22	**Accessing Data Files**	"I need access to code on GitHub that will allow me to access up-to-date public-use files that live on the AHRQ website."		X
23	**Accessing Data Files**	"I need clear guidelines for microdata access and data use agreement procedures – and easy access to technical support – in order to engage with certain datasets."		X
24	**Accessing Data Files**	"I need to be able to view a consolidated data release schedule in order to understand when to update my data files."		X
25	**Accessing Data Files**	"I need to quickly assess the characteristics of a data source – particularly if it provides nationally-representative, person-level data – in order to determine the LOE of, and applicability of that data, to my research."	X	
26	**Accessing Data Files**	"I need to be able to quickly determine the ease of access to data files for a particular data source in order to determine the LOE and feasibility of its use for my research."	X	

#	Category / Theme	User Requirement	Featured in Final	Not Featured / Applies only to the Backend
27	**Formulating & Refining a Data Tool Query**	"I need to be able to sort/refine query results – e.g., by condition and topic – in order to quickly answer my specific research question."	X	
28	**Formulating & Refining a Data Tool Query**	"I need a one-page data query interface in order to efficiently create and edit data queries."	X	
29	**Formulating & Refining a Data Tool Query**	"I need to have greater customization of age groups and date ranges in my queries in order to achieve the granular results I require."	X	
30	**Formulating & Refining a Data Tool Query**	"I need access to an easy-to-use query tool that can quickly pull from a broad dataset in order to perform complex queries without having to manipulate the data files themselves."	X	
31	**Formulating & Refining a Data Tool Query**	"I need to be able to understand alternate options to access the data in the event that the query tool I'm using doesn't yield the results I need in order to complete my research query."	X	
32	**Formulating & Refining a Data Tool Query**	"I need to easily modify my query on one page – switching between variable groups, variable selection, and potential analyses – in order to answer a research question efficiently."	X	
33	**Reviewing the Data**	"I need visual signifiers of data – including graphs or other data visualizations – whenever applicable in order to make the review of complex data more efficient."	X	
34	**Reviewing the Data**	"I need to be able to compare two queries side-by-side (e.g., mean and median expenditures per person) in order to gain an understanding of comparable variables for my research question."	X	
35	**Reviewing the Data**	"I need the ability to select from a variety of different "view" formats – including different data visualizations – in order to effectively analysis query results."	X	
36	**Reviewing the Data**	"I need to be able to select diagnostic and procedural variables in all the data tools in which its available."	X	
37	**Reviewing the Data**	"I need any footnotes or appendices that might accompany query results to be easily scannable in order to quickly grasp analyses."	X	
38	**Reviewing the Data**	"I need to be able to return to a query at a later time in order to complete my research unabated."		X
39	**Reviewing the Data**	"I need to be able to download query results as a printable 1-page PDF in order to distribute salient data with others or reference it in the future."	X	

#	Category / Theme	User Requirement	Featured in Final	Not Featured / Applies only to the Backend
40	**Reviewing the Data**	“I need descriptions for all variable labels – including examples – in order to understand the results of my query.”	X	
41	**Reviewing the Data**	"I need the option to view map visualizations for any relevant data in order to illustrate the richness of my dataset to users."	X	
42	**Seeking Support**	“I need to be able to easily submit a technical assistance request or a request for a special run of data in order to resolve my research question when my query results aren’t sufficient.”		X
43	**Seeking Support**	“I need user queries to be easily and efficiently routed through the UI to the right Support Users in order to provide substantive customer support.”		X
44	**Completing My Query Task**	“I need to be able to copy and share the link to a query in order to collaborate and share with others.”	X	
45	**Completing My Query Task**	“I need to be able to download query results in multiple formats in order to collaborate and share with others.”	X	
46	**Data Tool Management**	“I need backend management of the datasets to be user-friendly for both contractors and administrators in order to ensure that quality control efforts are carried out seamlessly every time the datasets are updated.”		X
47	**Quality Control**	“I need the query tool to be able to conduct statistical tests, so I can easily determine what's statistically significant for my research question.”		X

APPENDIX A: ACRONYMS AND DEFINITIONS

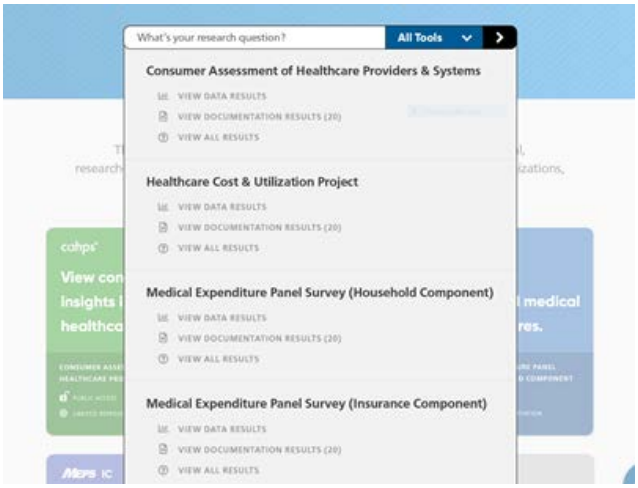
Table 1. List of Acronyms and Definitions

Acronym	Description
AHRQ	Agency for Healthcare Research and Quality
COR	Contracting Officer Representative
CAHPS	Consumer Assessment of Healthcare Providers and Systems
HC	Household Component
HCUP	Health Cost and Utilization Project
HHS	Department of Health & Human Services
IC	Insurance Component
MEPS	Medical Expenditure Panel Survey
NHQDRnet	National Healthcare Quality and Disparities Reports Data Query Tool
UI	User Interface
UX	User Experience

APPENDIX B: DISCONTINUED DESIGNS FOR PROTOTYPES

The following are design concepts that were presented as part of the Alpha or Beta version of the Style Examples that are not part of the prototypes development but are potential features for future consideration.

Search Dropdown



This screenshot shows how search results could appear upon entering a query in the search bar. Results reveal results by tool in three categories: "View Data Results", as envisioned, will send the user to the query interface, "View Documentation Results" will reveal results by document type, and "View All Results" reveals a consolidated search results field for that tool.

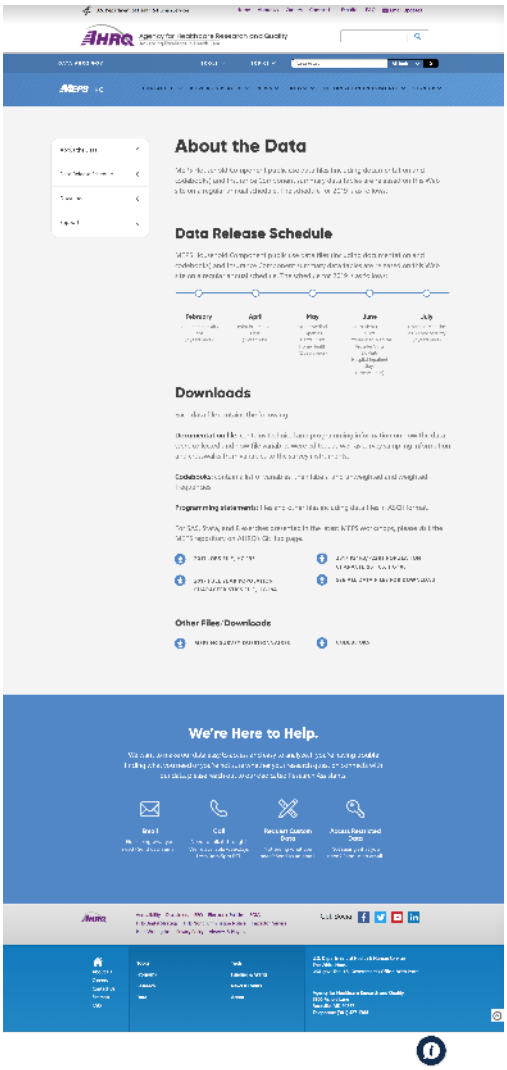
Note: In the Style Examples, when you click on a category in the dropdown, you will only be taken to the landing page of the appropriate data tool.

Topic Page

The Topic page intent is to address the user needs for an easy way to see all the data and resources pertaining to a healthcare topic. User requirements 8 and 9 directly request this experience.

Upon selecting a topic on either the portal or on one of the landing pages, the user's options are constrained to the tools that are relevant to that topic, and featured data queries and reports within that topic area.

Data Content Page



AHRQ Navigation: Standardized navigation to the Topic and Tool Landing Pages

Tool Navigation: Navigation to content related to the particular tool.

Static Sidebar (Page Navigation) - Closed State: Allows users to jump to a specific content section.

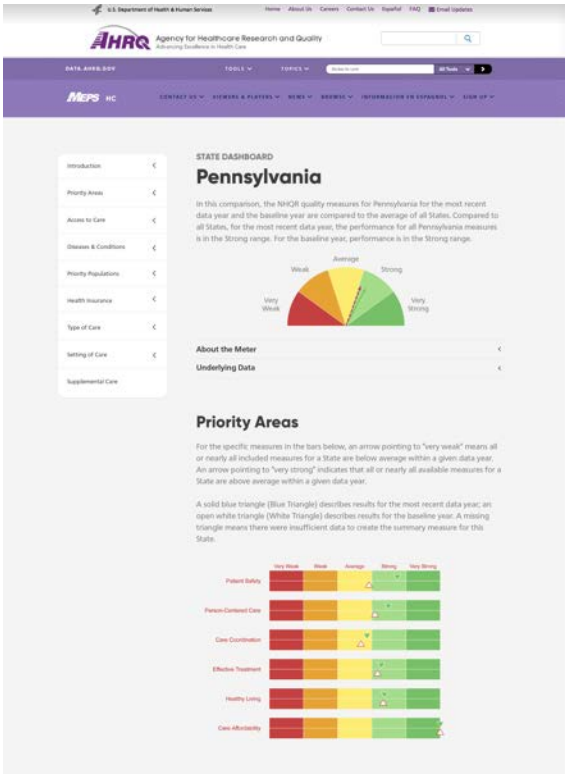
About the Data: Copy orienting visitors to the data page and defining the different data files.

Data Release Schedule: Provides quick read of the data sets recently released and scheduled to be released.

Downloads: Visitors can download the most recent data files and access all data files available for download.

Other Files / Downloads: We're thinking through how to make reference files easy to access should a visitor want to look up a variable code or a specific survey question. If possible, we intend for reference materials to be available here for easy access and download.

Here is a variation on the Data content page, made for NHQDRnet State & National Dashboard pages. Here, clicking on the dashboard data visualizations could send the user to the query portal to manipulate the data represented in more detail.



Deferred Concepts on the AHRQ Data Tool Portal

Some sections that were presented in the Beta version of the Style Examples for the Data Tool Main Portal (a screenshot of which is provided on this page) have been deferred for future consideration. These sections are the following:

Search Bar: due to the reliance of all interviewed users on search engines, the search bar has been implemented as a priority feature for engaging with the tool. This search bar connects users search terms to relevant data and publications. If users select a data tool result, they go to the specified data tool query page with a pre-populated query based on the search terms provided.

Topics: this section offers another way of narrowing down the tools at the users' disposal by topic area. Users selecting a topic would go to a topic page that will allow them to see all the relevant tools with data on those topics as well as featured queries and reports.

Support: a floating "support" button (bottom right corner above the support section) and a dedicated customer support section serve to both reinforce the availability of AHRQ's dedicated technical assistance staff while providing those staff with efficient routing of customer outreach. A Support section will exist in all screens that are part of the harmonized design.

Deferred Concepts on the Data Tool Landing Page

The following are sections and features that were presented in the Beta version of the Style Examples but have been deferred for future consideration.

Description Section: In this section, the following links have been removed from the final version of the Style Examples and will be considered for possible future consideration:

- "About" describes the dataset and methodologies in detail.
- "Data Files" leads to a page containing data files in various formats, microdata access guidelines, codebooks, linkage files and other relevant information for pro users.
- "Publications" contains chart books, reports, and other relevant information.
- Note: A link to the original version of the site is still part of the Style Examples ("Visit Homepage").

Topics: As in the Data Portal Main Page, the Topics section has been removed and will be considered for future consideration.

RFTO #10: Developing Strategies for the Harmonization of AHRQ Data Tools (Data Tools)

Prototype Report

June 14, 2019

Contents

REVISION HISTORY

Date	Version	Description	Author(s)
April 9, 2019	Guide to the Alpha Prototypes	Preliminary report accompanying the Alpha Prototypes deliverable, documenting access to the prototypes, initial findings and preliminary assessment of several features	Atlas Team
May 3, 2019	Guide to the Beta Prototype	Report accompanying the Beta Prototype deliverable, documenting access, included data tools, and list of features	Atlas Team
May 31, 2019	Prototype Report	Report accompanying the prototype deliverables documenting access, scope, and list of features	Atlas Team
June 14, 2019	Prototype Report (Final)	Incorporating AHRQ's feedback and updates on the Prototype Report that Atlas previously submitted on May 31, 2019 (sections updated: 2.1, 2.2, and 2.5)	Atlas Team

Note: The following document was produced as part of AHRQ's Data Tools Harmonization Project Assessment Phase. The original document contained links to prototypes that are not available for viewing at this time. Those links have been replaced with the following highlighted text:
(link unavailable at this time)

Usernames and passwords associated with the prototypes have been removed and replaced with the following highlighted text:
(removed)

1. INTRODUCTION

The Agency for Healthcare Research and Quality (AHRQ) provides several data tools and summary data tables that have dedicated user groups, but currently these tools are siloed and lack a common user experience (UX) and branding among them. The AHRQ data tools included in this project are:

- Consumer Assessment of Healthcare Providers and Systems (CAHPS)
- The Health Cost and Utilization Project's (HCUP) HCUPnet
- HCUP Fast Stats
- Medical Expenditure Panel Survey (MEPS) Summary Tables: Household Component (HC)/MEPSnet (HC)
- MEPS Summary Tables: Insurance Component (IC)/MEPSnet (IC)
- National Healthcare Quality and Disparities Reports Data Query Tool (NHQDR)

After evaluating the potential for harmonizing these AHRQ statistical data tools through previous tasks (General Specifications, Environmental Scan, Product Assessment, and Style Examples), the project team is developing prototypes to demonstrate how the selected commercial off-the-shelf (COTS) data visualization product could deliver the important features of these data tools to the public.

During the previous prototyping phases, the project team developed initial data visualizations for all the data tools and built out the visualizations in Tableau through iterative work sessions with the different points-of-contact (POCs). From the feedback through the work sessions, the project team has developed these visualizations in the current prototype. Please note that:

- **The Prototype is conceptual and exists to support AHRQ staff in evaluating the feasibility of a harmonized data tool portal using Tableau.**

- **The Prototype is for demonstration purposes only and will not be for public view or use.**

- **The Prototype will eventually be transferred and hosted at AHRQ so that it can be viewed and used internally for an indefinite period of time.**

- **The website portal landing page is now simplified to encourage focus on the actual data visualizations that are shown in the prototype.**

- **This report may be updated based on feedback and findings resulting from AHRQ's review of the prototype. The review period will be from June 3 – 7, 2019 and will include a demo of the prototype on June 6, 2019.**

1.1 Guide to Reviewing the Prototype

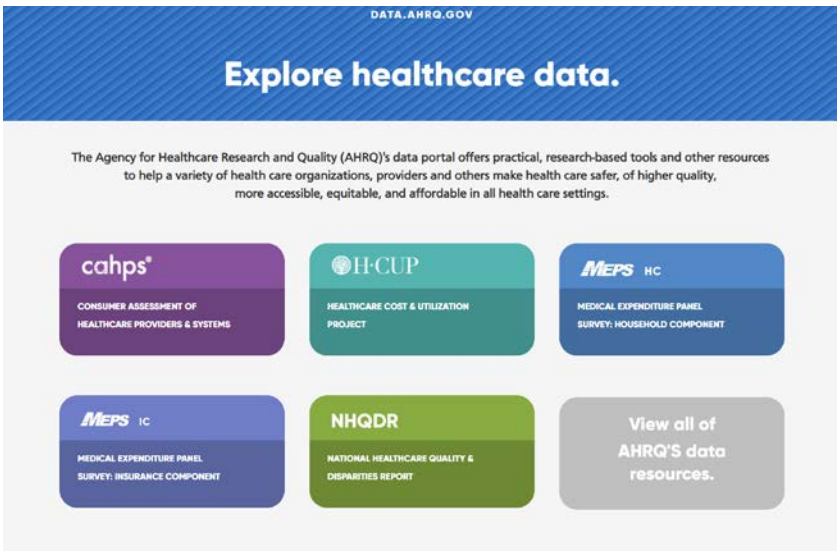
If you're accessing this within the AHRQ network, please use **Internet Explorer (IE)** or **Mozilla Firefox** as your browser to access the Prototype.

Use this link to go to the Prototype: (link unavailable at this time). This link will take you to the Prototype Introduction Page. Click on the "Portal" link on the upper right-hand side of the screen.

How to Get to the Data Visualizations

On the portal's main page, click on any of the data tool tiles. Each tile will direct you to that specific data tool's UI Query Page where you can see the Tableau data visualizations. Once on that data tool's UI Query Page, you will be asked to sign-in to Tableau to see the data visualizations. When asked to **sign-in to Tableau**, use the following login information:

Username: **(removed)**
Password: **(removed)**



2. DATA VISUALIZATIONS

The Prototype includes data visualizations for all the data tools. The project team had work sessions with POCs of the data tools in order to clarify and expand the data visualizations that were presented during the Alpha and Beta Prototypes. For each of these data tools, the data visualizations that have been developed are noted.

2.1 CAHPS

For CAHPS, the project team developed initial data visualizations for:

- Clinician & Group Survey
- Health Plan Survey

For the two CAHPS surveys, the types of the data visualizations and key features available are indicated (x) in the table below.

Data Tool: CAHPS					
Data Visualization	Graph & Tabular Data		Map	Standard Errors Display Option	Download/Share
	Line/Trend	Bar/ Cross-Section			
Clinician & Group Survey	N/A	x	N/A	N/A	x
Health Plan Survey	N/A	x	N/A	N/A	x

In the prototype, the tabular data tables are shown at the bottom of the graph within the same visualization page.

The visualizations include options to view top box scores and percentiles.

The current prototype has an estimated **20%** of the actual survey data that is currently available in CAHPS.

Please note that the POC has expressed interest to include older CAHPS survey data (not just the survey data that's currently available on the CAHPS website) into the actual harmonized data portal once that's in development.

2.2 HCUPnet

For HCUP, the prototype includes data visualizations for HCUPnet. You will be able to generate visualizations of the following:

- Inpatient Stays
- Emergency Department

The following are not part of the prototype:

- Ambulatory Surgery
- Community

The table below contains a list of the data visualizations developed for HCUPnet along with indications (x) of the visualization types and key features that are available.

Data Tool: HCUPnet					
Data Visualization	**Graph & Tabular Data**		**Map**	**Standard Errors Display Option**	**Download/Share**
	Line/Trend	**Bar/ Cross-Section**			
Inpatient	x	N/A	x	x	x
Emergency Department	x	N/A	x	x	x
Ambulatory Surgery	N/A	N/A	N/A	N/A	N/A
Community	N/A	N/A	N/A	N/A	N/A

In the prototype, the tabular data tables are shown at the bottom of the graph within the same visualization page.

The current prototype for HCUPnet only includes **less than 1%** of the metadata in HCUPnet, according to the POC. This is due to the inclusion of detailed ICD9 and ICD10 condition codes in the current HCUPnet tool, which are not included in the Tableau prototype. However, the POC also noted that the current loaded data in the prototype accounts for handling an estimated **70%** of the inquiries and data manipulations by users.

2.3 HCUP Fast Stats

Data visualizations in the prototype for Fast Stats were developed for the following:

Under “State Trends in Hospital Use by Payer”

- Inpatient Stay
- Emergency Department Visit

Under “National Hospital Use and Costs”

- Trends in Inpatient Stays
- Most Common Diagnoses for Inpatient Stays
- Most Common Operations During Inpatient Stays

Under “Opioid-Related Hospital Use”

- Trends in Inpatient Stays and Emergency Department Visits

The types of the data visualizations and key features that are currently available in the prototype are indicated (x) in the table below.

Data Tool: HCUP Fast Stats					
Data Visualization	**Graph & Tabular Data**		**Map**	**Standard Errors Display Option**	**Download/Share**
	Line/Trend	**Bar/ Cross-Section**			
State Trends in Hospital Use by Payer:					
Inpatient Stay	x	N/A	x	N/A	x
Emergency Department Visit	x	N/A	x	N/A	x
National Hospital Use and Costs:					
Trends in Inpatient Stays	x	x	N/A	N/A	x
Most Common Diagnoses for Inpatient Stays	x	x	N/A	N/A	x
Most Common Operations During Inpatient Stays	x	x	N/A	N/A	x
Opioid-Related Hospital Use:					
Trends in Inpatient Stays and Emergency Department Visits	x	N/A	x	N/A	x

In the prototype, the tabular data tables are shown at the bottom of the graph within the same visualization page.

The prototype is using an estimated **95%** of the actual HCUP Fast Stats data available on the current AHRQ website.

2.4 MEPS HC

For MEPS HC in the prototype, you will be able to generate visualizations of data from the following HC summary tables:

- Use, Expenditures, and Population
- Health Insurance
- Accessibility and Quality of Care
- Medical Conditions
- Prescribed Drugs

Below is a list of the data visualizations developed for MEPS HC along with indications (x) of the visualization types and key features that are available.

Data Tool: MEPS HC					
Data Visualization	**Graph & Tabular Data**		**Map**	**Standard Errors Display Option**	**Download/Share**
	Line/Trend	**Bar/ Cross-Section**			
Use, Expenditures, & Population	x	x	N/A	x	x
Health Insurance	x	x	N/A	x	x
Accessibility & Quality of Care	x	x	N/A	x	x
Medical Conditions	x	x	N/A	x	x
Prescribed Drugs	x	N/A	N/A	x	x

In the prototype, the tabular data tables are shown at the bottom of the graph within the same visualization page.

The project team received an estimated **99%** of the actual available MEPS HC data to upload and use in the prototype.

Please note that in working and discussions with the POC, it was decided that the MEPSnet/HC Trend Query tool in the current MEPS website will not be included in any of the prototypes.

2.5 MEPS IC

For MEPS IC in the prototype, you will be able to generate visualizations of the following:

- Private Sector - National
- Private Sector - State
- Public Sector

The following are not included in the prototype:

- Civilian data
- National totals
- Private sector data by Metro Areas
- Premium / contribution / cost distributions

Below is a list of the data visualizations developed for MEPS IC along with indications (x) of the visualization types and key features that are available.

Data Tool: MEPS IC					
Data Visualization	**Graph & Tabular Data**		**Map**	**Standard Errors Display Option**	**Download/Share**
	Line/Trend	**Bar/ Cross-Section**			
Private Sector - National	x	x	N/A	x	x
Private Sector - State	x	x	x	x	x
Public Sector	x	x	N/A	x	x

In the prototype, the tabular data tables are shown at the bottom of the graph within the same visualization page.

The project team received an estimated **70%** of the available actual MEPS IC data to upload and use in the prototype.

2.6 NHQDR

Data visualizations for NHQDR were created to access and present the following levels of data:

- National
- State

The types of the initial data visualizations and key features that are currently available in the Prototype for NHQDR are indicated (x) in the table below.

Data Tool: NHQDR					
Data Visualization	**Graph & Tabular Data**		**Map**	**Standard Errors Display Option**	**Download/Share**
	Line/Trend	**Bar/ Cross-Section**			
National	x	x	N/A	x	x
State	x	x	x	x	x

In the prototype, the tabular data tables are shown at the bottom of the graph within the same visualization page.

The project team received an estimated **65%** of the total available data in the NHQDR database to upload and use in the prototype. The data that was uploaded and being shown by the prototype is limited to only one subject area, which is the “Priority Areas” chapter/category.

APPENDIX A: ACRONYMS AND DEFINITIONS

Table 1. List of Acronyms and Definitions

Acronym	Description
AHRQ	Agency for Healthcare Research and Quality
CAHPS	Consumer Assessment of Healthcare Providers and Systems
COTS	Commercial off-the-shelf
HC	Household Component
HCUP	Health Cost and Utilization Project
HHS	Department of Health & Human Services
IC	Insurance Component
MEPS	Medical Expenditure Panel Survey
NHQDRnet	National Healthcare Quality and Disparities Reports Data Query Tool
POC	Point-of-Contact

APPENDIX B: COPY OF THE ALPHA PROTYPES REPORT

The following is a copy of the report and guide that accompanied the Alpha Prototypes deliverable. Please note that the Tableau data visualizations have been updated since the Alpha Prototypes were submitted.

B.1 INTRODUCTION

The Agency for Healthcare Research and Quality (AHRQ) provides several data tools and summary data tables that have dedicated user groups, but currently these tools are siloed and lack a common user experience (UX) and branding among them.

Using the visual prototype examples developed in Task 5 as guide, two functional Alpha prototype portals were created mainly to allow AHRQ to assess the features of the two candidate data visualization products (i.e., for Tableau and for Qlik).

- **These Alpha Prototypes are conceptual and exist to support AHRQ staff in evaluating the feasibility of a harmonized data tool portal and data visualization products.**

- **The intention of the Alpha Prototypes is to implement high-priority features of the two data visualization tools, Tableau and Qlik, in order for AHRQ to be able to evaluate each product and determine which of the two will best address AHRQ's data tool harmonization needs.**

- **The prototypes will not be for public view or use.**

- **The portals have key links and navigation that allow users to test out the data visualization products. These prototypes are works-in-progress, and some of the links, navigation, and features that are not critical to the assessment of the data visualization products are not active in the Alpha prototypes (e.g., the links in the main header where the US Department of Health and Human Services and AHRQ logos are located, the tiles in the Featured Section).**

For Alpha, Landing Pages and queries to the data visualization products are available for the following data tools:

- The Health Cost and Utilization Project's (HCUP) HCUPnet
- Medical Expenditure Panel Survey (MEPS) Summary Tables: Household Component (HC)/MEPSnet (HC)
- National Healthcare Quality and Disparities Reports Data Query Tool (NHQDRnet)

The other data tools will be incorporated to the Beta version of the prototype:

- Consumer Assessment of Healthcare Providers and Systems (CAHPS)
- HCUP Fast Stats
- MEPS Summary Tables: Insurance Component (IC)/MEPSnet (IC)

B.1.1 Guide to Reviewing the Alpha Prototypes

Use this link to go to the Alpha Prototypes: (link unavailable at this time). This link will take you to the initial Alpha Prototype Page where you will be able to see the links to both the Tableau version and the Qlik version.

Please note that you will be asked to sign-in to Tableau and Qlik to be able to see the visualizations.

When asked to **sign-in to Tableau**, you can use the following login information:

Username: (removed)

Password: (removed)

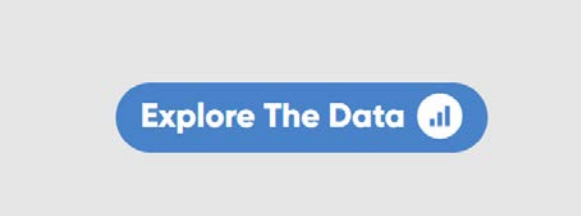
When asked to **sign-in to Qlik**, you can use the following login information:

User name: (removed)

Password: (removed)

How to Get to the Data Visualizations

As this is the Alpha Prototypes, the focus is on the **Query Pages** of the MEPS HC, HCUPnet, and NHQDR. When reviewing the prototypes, please use the **"Explore the Data"** button located in each of the three data tools' Landing Pages to get to the Query Pages.



Note that the Query Forms in the Landing Pages may not look, or flow as described in the previous Style Examples due to either limitation of the visualization product (Qlik or Tableau) or may need further coding/configuration in the succeeding versions of the prototypes. For example, the Query Forms in the Qlik version of the prototypes appear with unwanted white space around them. These Query Forms are works-in-progress (for both the Qlik and Tableau versions) that we plan to address further for succeeding versions of the prototypes. The Atlas Team emphasizes **that reviewers use the "Explore the Data" button to get to the Query Pages and test the visualizations through that page.**

B.2 ASSESSMENT OF SEVERAL PRODUCT FEATURES

The following features are available and can be viewed in the Alpha prototypes:

- Tabular data - single variable and crosstabs
- Standard errors toggle
- Plots - including line plots over time
- Maps
- Download options

For some features that are not part of the Alpha Prototypes, the Atlas Team did an assessment based on research and experience working on the two data visualization products. These assessments are in the following subsections.

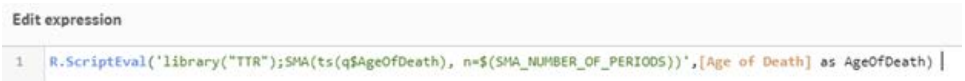
B.2.1 Significance Testing (Z-test)

Z-Scores can be coded as calculated fields within Qlik and Tableau for AHRQ data that is loaded up into the data visualization products. However, because the development of the Beta and Final Prototype will not utilize APIs, which could easily push distinct Z-scores into an application to determine the p-value, more conventional means will be considered by the Atlas Team instead. Alternatively, the Atlas team could pursue a static Z-test application, such as the one available in HCUP, and included that in the HTML code of the product, however this was not included in the wireframes and would have to be considered within a redesign of the pages.

B.2.2 Integrating with R

Integrating R with Qlik Sense is a simple setup, requiring the user to install R on the same server as Qlik sense, installing packages and then connecting R to the Qlik Sense engine. The Qlik support team is able to provide test scripts to ensure there's a connection between R and Qlik and the Atlas Team was able to successfully connect up Qlik to R.

Running R basic scripts in Qlik is simple and easily achievable. In the Qlik UI, the user is able to create functions using R code through Qlik script by using the Qlik syntax. Qlik formulas compatible with R include R.ScriptAggr, R.ScriptAggrEx, R.ScriptAggrExStr, R.ScriptEval, R.ScriptEvalStr. These functions allow the Qlik to not only execute R functions on data loaded into Qlik, but also allow R functions to run on the selections of the Qlik user, allowing the user to build on their work.



However, the team ran into issue when attempting to run more complex scripts involving package installs. Moreover, attempting to troubleshoot the query R Code requires external support as errors in the script cannot be troubleshot in the Qlik back end.

Establishing the integration between R and Tableau Desktop is relatively simple, requiring only a few steps to start the connection. Once the integration is established, the relationship between the two programs becomes very powerful, allowing the statistical analysis of R to work with the visualization of Tableau.

Any R script, function, or package (any calculation) can be called from Tableau, so long as the script, function, or package is installed on the Rserve server used for the integration. New scripts, functions, and packages can easily be installed on the server.

The only limitation to the integration is that visualizations in R cannot be imported directly into Tableau (i.e., graphs/charts/etc. that are created directly in R cannot be converted to a Tableau integration). Although, the image file or URL of the R visualization can be used in the Tableau dashboard. However, since Tableau (if selected as AHRQ's data visualization product) will be used for creating the visualizations themselves (not R), this limitation will not affect the purpose of this integration.

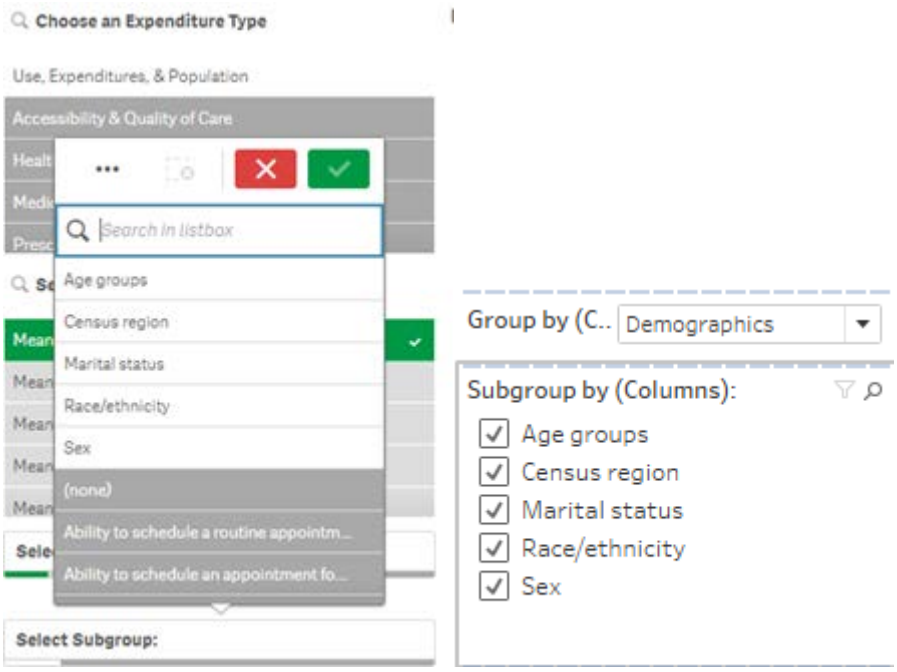
Data Analysis

The Atlas Team ran the R Sample Code provided by the AHRQ team in R and attempted to include the code into Qlik, which can run R code in Qlik Sense, although the script has to be substantially reformatted. Moreover, due to the limited ability to troubleshoot R script in the Qlik UI, the Atlas Team is currently unable to conclude whether the R Sample Code can be run in either COTS product. Based on the selection by the AHRQ team for the Beta Prototype, the Atlas Team will work closer with a product representative and with AHRQ to confirm how to execute the exact R sample code through the COTS product.

Additionally, the Atlas Team would like to schedule time with a MEPS representative during the Beta development phase to review out of the box functionality in the selected COTS product to determine if existing functionality can provide the desired outcomes for MEPS data analysis

B.2.3 Variable Selection/Recoding/Subsetting

Variable selection and subsetting features are included in the Alpha Prototypes as shown in this screenshot below.



The Subset of Variables for "Demographics" in MEPSnet HC in Qlik and Tableau

Recoding, however, is not a functionality that's immediately available to front end users. The Atlas Team will need more time to investigate whether recoding can be implemented, either in the succeeding prototypes or in the actual and full implementation of the AHRQ Data Portal application.

B.2.4 Query Sharing & Downloading

The style examples include a universal feature that allows users to share query results, both within the query interface itself and in the premade "Featured" queries sections. The functionality varies according to the tool, but generally includes: Copy Query Link, Email Query Link, Download in PDF, Download in PowerPoint, and Download in Excel. Both Qlik and Tableau offer this functionality in the App.

However, the Atlas Team noticed a decrease in functionality with Qlik once the visualizations were embedded into the iframes. In fact, the only functionality available to front end users for sharing data is using a Qlik function that copies the URL to the user's clipboard.

B.2.5 Natural Language Data Queries

An up and coming feature in data dashboards is the ability to form a natural language data query in a search bar and receive relevant data insights. In Tableau, this feature is called "Ask Data", and it was released on February 14th, 2019 in Tableau version 2019.1. It is available for

any pricing level of Tableau. Qlik Sense's Smart Search feature is more dependent on user configuration to make data more easily searchable using natural language.

Based on initial demonstrations of the product by the Atlas Team, the service relies on phrasing provided by the user, which lacks accuracy. Moreover, the service relies on how the data loaded is structured, rather than analyzing the data inside of the tables, which takes a considerable amount of power to analyze and understand. At this point, if you know your data enough to use this feature, then you likely know how to structure the data through Tableau's easy to use drag and drops. Nevertheless, this is a field where Tableau can make some considerable strides in the future where it could prove to be much more valuable than its current state.

B.2.6 Mobile

After delivering a completed proof of concept of the tool in desktop format, the design of the portal will need to be optimized for mobile use. While many elements of the existing user interface might easily translatable to mobile viewing, the smaller form factor and idiosyncrasies of mobile use, will require a complete redesign of the site for a mobile version of data.ahrq.gov (e.g. m.data.ahrq.gov). In this redesign, the concept, design, HTML coding, and COTS dashboard visualizations need to be built from the ground up with this mobile view.

The level of effort for this redesign would take approximately 75% of the effort put forth to the Alpha prototype, and involve a similar flow, such as creating low-fidelity wireframes and soliciting user feedback. Moreover, there would need to be time devoted to developing the concepts and wireframes along with ensuring that dashboard configurations are mobile-ready.

B.2.7 508 Compliance

While the prototypes do not have to undergo 508 review if staff don't require accessibility considerations, per AHRQ's 508 Coordinator (Web). The Atlas Team did conduct a preliminary 508 compliance review using Wave (a web accessibility evaluation tool). However, the results weren't sufficiently comprehensive as the Alpha Prototype uses images as placeholders for areas where HTML coding will exist for the Beta Prototype and for Final delivery. The Atlas Team will consider running a future and more built-up version of the prototype on Wave or a similar tool in order to provide a report about the accessibility of the portal for documentation and future reference for AHRQ.

B.2.8 Ease of use (back-end) and Data Manipulation (back-end)

On the back-end, Qlik utilizes a somewhat convoluted method for loading data. First, data needs to be loaded onto a shared folder on the Qlik server. From there, data is loaded into a temporary layer. In this layer, Qlik automatically makes data transformations, and translates these transformations into a SQL-like language. From there, the Administrator can modify the language to complete the upload. For users less comfortable with SQL-like languages, Qlik provides a GUI to create table joins or to make modifications to the data. While a user can accomplish a lot using the basics of Qlik, most valuable functionality, and a large portion of Qlik's value add, is available only by using the SQL-like language. It is essential that AHRQ has an Administrator who can become very comfortable with Qlik syntax to get the most out of the product. This is particularly relevant too for creating visualizations once the data is loaded. Using this SQL-like language, users can create advanced formulas to drive visualizations. The learning curve for the language based on Atlas's experience sits in between Excel formulas and SQL.

Navigating Qlik's backend for creating data visualizations is complicated. The hierarchy begins with the Qlik Hub, which is where all Qlik applications are developed. In each Qlik application, back end users create a sheet, which houses the actual visualizations. When data is loaded, it is loaded on the application level, so an application could use multiple data sources to power many visualizations. Navigating between applications and views is counterintuitive because once inside the Qlik hub the user is in one long session of Qlik, which means basic functionality like a back-button can't be used. Instead, users have to use a set of panes, arrows and new tabs opened by Qlik to get around.

Once a visualization in a view is complete, an Administrator can either publish an individual view or an entire application to make it visible to other Qlik users. Once a view is published however, nothing on that view can change. The administrator must choose to unpublish the view or duplicate the view and publish a revised view, which creates a new URL. This makes any necessary live edits an involved process that will affect front-end users.

Tableau's back-end begins with Tableau Prep, which provides a GUI for adding and transforming data files. The drag and drop GUI uses simple and intuitive gestures to join or union data, and buttons are available for more complex tasks, like pivoting the data or transforming the data. For more involved calculations, Tableau prep offers a simpler-than-Excel formula syntax allowing users to create conditional statements for their data. The transformation of data in Tableau prep is saved as a flow and can be transferred between parties, so Atlas will be able to provide the Tableau prep flows used for this project.

From there, users create worksheets and dashboard views in Tableau Desktop. Tableau Desktop provides significantly more functionality and complexity than Tableau Prep. And since it's a standalone application and not a web application, it is very easy to use once the basics are mastered. Backend users click-and-drag fields from data sources that are loaded into Tableau, and new calculated fields can be created using the same simpler-than-Excel syntax. Only one visualization can be built per worksheet, so a workflow can get crowded when building out a dashboard, but once a dashboard is created with those visualizations the user can hide them. The dashboard is made by simply clicking and dragging the desired workflows, and from there, the user can publish the dashboards to be available for public consumption.

At any part of the flow, any change can be made to the data and it will be reflected in Tableau with a quick refresh. Users can modify the raw data, the Tableau Prep flow or the Tableau Desktop, but once the revision is made, it's available to view the final dashboard. Moreover, a user can push these updates into the final Tableau dashboard for public consumption, allowing for changes to be easily made while ensuring maximum uptime for front-end users.

B.2.9 Side-by-Side Compare

A side-by-side using two visualizations can be implemented in the Beta version of the Prototype. The Atlas Team does foresee its implementation in Qlik to be of more considerable effort compared to implementing this feature in Tableau. Implementation of buttons in Qlik to direct users to the preferred visualization updates that will be needed for a side-by-side compare currently has provided challenges to the Atlas Team when developing the Alpha Prototypes.

B.2.10 Toggle Between Map and Chart

For the Tableau Alpha Prototypes, buttons to switch from map to chart have been implemented.

A toggle in Qlik between map and chart visualizations can be implemented in the Beta version (if Qlik is selected as the data visualization product to move forward with). For Qlik, the buttons to toggle between map and chart can be implemented similarly to what has been implemented in the Alpha Prototypes for *show/hide* container used to show and hide the standard errors in the MEPS HC visualization. This implementation involved more considerable effort to implement and maintain compared to implementing the buttons used in the current Tableau visualizations that you can see in the Alpha Prototypes.

B.2.11 Treatment of Suppressed Values

Suppressed values in AHRQ's data take on many forms, such as an NA, absent value, or asterisk marking low sampling rates for significance. Firstly, the Atlas Team made sure to treat fields with these values as strings, so that they could be correctly loaded into the applications. For example, loading a number with an asterisk might cause the data loader to skip this record, or treat it as a regular number, which is the opposite of what the asterisk is trying to achieve. The Atlas Team recommends, as part of a long-term solution, to change the data at its source by creating a separate field to capture exactly what the asterisk or the other such values are trying to convey.

APPENDIX C: COPY OF THE BETA PROTYPE REPORT

The following is a copy of the report and guide that accompanied the Beta Prototype deliverable. Please note that the data portal and the Tableau data visualizations have been updated since the Beta Prototype was submitted.

C.1 INTRODUCTION

The Agency for Healthcare Research and Quality (AHRQ) provides several data tools and summary data tables that have dedicated user groups, but currently these tools are siloed and lack a common user experience (UX) and branding among them. The AHRQ data tools included in this project are:

- Consumer Assessment of Healthcare Providers and Systems (CAHPS)
- The Health Cost and Utilization Project's (HCUP) HCUPnet
- HCUP Fast Stats
- Medical Expenditure Panel Survey (MEPS) Summary Tables: Household Component (HC)/MEPSnet (HC)
- MEPS Summary Tables: Insurance Component (IC)/MEPSnet (IC)
- National Healthcare Quality and Disparities Reports Data Query Tool (NHQDR)

After evaluating the potential for harmonizing these AHRQ statistical data tools through previous tasks (General Specifications, Environmental Scan, Product Assessment, and Style Examples), the project team is developing prototypes to demonstrate how the selected commercial off-the-shelf (COTS) data visualization product could deliver the important features of these data tools to the public.

During the previous prototyping phase (i.e., Alpha), the project team developed initial data visualizations (for MEPS HC, HCUP, and NHQDR) that were developed using two different data visualization products, Tableau and Qlik. The Alpha Prototypes enabled AHRQ to assess the features of Tableau and Qlik, and to consequently select which product to proceed developing the Beta and Final Prototypes with.

For this Beta Prototype, the project team has developed this Prototype using AHRQ's selected COTS data visualization product, **Tableau**.

- **The Beta Prototype is conceptual and exists to support AHRQ staff in evaluating the feasibility of a harmonized data tool portal using Tableau.**

- **The goal of the Beta Prototype is to implement and feature more fully developed data visualizations for MEPS HC, MEPS IC, and HCUPnet.**

- **In addition, the Beta Prototype includes initial data visualizations for NHQDR, HCUP Fast Stats, and CAHPS.**

    - **These initial data visualizations will be used for scheduled work sessions with the data tools' Points-of-Contact (POCs).**

    - **These work sessions are being held with the POCs to iteratively develop data visualizations for the Final Prototype.**

- **These prototypes are for demonstration purposes only and will not be for public view or use.**

- **The website portal landing page is now simplified to encourage focus on the actual data visualizations for the Beta and Final prototypes.**

C.1.1 Guide to Reviewing the Beta Prototype

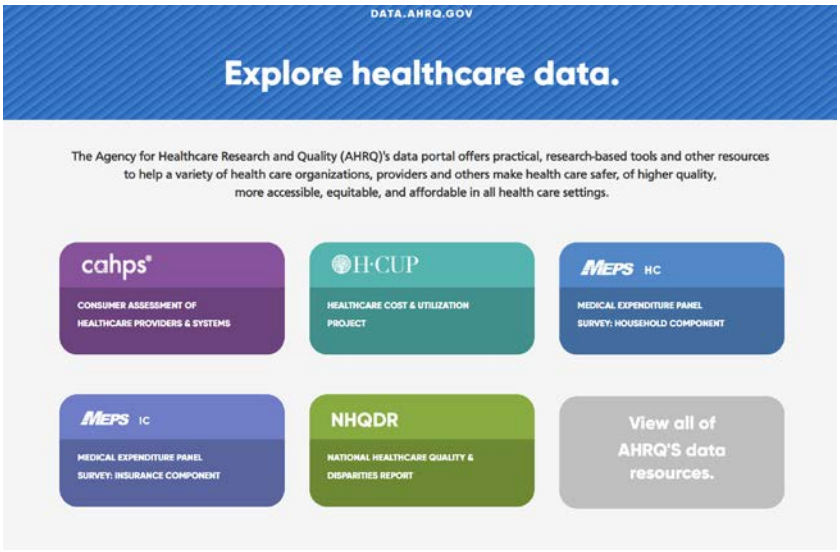
If you're accessing this within the AHRQ network, please use **Mozilla Firefox** as your browser to access the Beta Prototype.

Use this link to go to the Beta Prototype: (link unavailable at this time) This link will take you to the Beta Prototype Introduction Page. Click on the "Portal" link on the upper right-hand side of the screen.

How to Get to the Data Visualizations

On the portal's main page, click on any of the data tool tiles. Each tile will direct you to that specific data tool's UI Query Page where you can see the Tableau data visualizations. Once on that data tool's UI Query Page, you will be asked to sign-in to Tableau to see the data visualizations. When asked to **sign-in to Tableau**, use the following login information:

Username: (removed)
Password: (removed)



C.2 DATA VISUALIZATIONS

The Beta Prototype includes more fully developed data visualizations for MEPS HC, MEPS IC, and HCUPnet. The project team had work sessions with POCs of these three data tools in order to clarify and expand the data visualizations that were first presented during the Alpha Prototypes. For each of these data tools, the data visualizations that have been developed, so far, are noted.

C.2.1 MEPS HC

For MEPS HC in the Beta Prototype, you will be able to generate visualizations of data from the following HC summary tables:

- Use, Expenditures, and Population
- Health Insurance
- Accessibility and Quality of Care
- Medical Conditions
- Prescribed Drugs

Below is a list of the data visualizations developed for MEPS HC so far, along with indications (x) of the visualization types and key features that are available. For visualization types and key features that are left blank, these will be finalized and filled out in the Final Prototype Report.

Data Tool	Data Visualization	Tabular Data		Graph		Map	Standard Errors Display Option	Download/ Share
		Trend Data	Cross Sectional Data	Line	Bar			
MEPS HC								
	Use, Expenditures, & Population	x	x	x	x		x	x
	Health Insurance	x	x	x	x		x	x
	Accessibility & Quality of Care	x	x	x	x		x	x
	Medical Conditions	x	x	x	x		x	x
	Prescribed Drugs	x		x			x	x

Please note that in working and discussions with the POC, it was decided that the MEPSnet/HC Trend Query tool in the current MEPS website will not be included in any of the prototypes.

C.2.2 MEPS IC

For MEPS IC in the Beta prototype, you will be able to currently generate visualizations of the following:

- Private sector data
- Public sector data
- State-level data

Below is a list of the data visualizations developed for MEPS IC so far, along with indications (x) of the visualization types and key features that are available. For visualization types and key features that are left blank, these will be finalized and filled out in the Final Prototype Report.

Data Tool	Data Visualization	Tabular Data		Graph		Map	Standard Errors Display Option	Download/ Share
		Trend Data	Cross Sectional Data	Line	Bar			
MEPS IC								
	Private	x	x	x	x		x	x
	Public	x	x	x	x		x	x
	State	x	x	x	x	x	x	x

C.2.3 HCUP

For HCUP, the Beta Prototype includes data visualizations for HCUPnet. You will be able to currently generate visualizations of the following:

- Inpatient Stays
- Emergency Department

The following are not part of the Beta Prototype but are being considered to be developed as part of the Final Prototype, if the appropriate data will be available to support their development:

- Ambulatory Surgery

- Community

The table below contains a list of the data visualizations developed for HCUPnet so far, along with indications (x) of the visualization types and key features that are available. For visualization types and key features that are left blank, these will be finalized and filled out in the Final Prototype Report.

Data Tool	Data Visualization	Tabular Data		Graph		Map	Standard Errors Display Option	Download/ Share
		Trend Data	Cross Sectional Data	Line	Bar			
HCUP - HCUPnet								
	Inpatient	x		x		x	x	x
	Emergency Department	x		x		x	x	x
	Ambulatory Surgery							
	Community							

C.2.4 Initial Data Visualizations for CAHPS, HCUP Fast Stats, and NHQDR

For the following data tools, the project team created initial data visualizations as starting points of discussion with the tools' respective POCs. Work sessions have been scheduled with the POCs where we will go over the initial visualizations, discuss which features to develop, gather feedback, and build out the required data visualizations that can be included for the Final Prototype.

CAHPS

For CAHPS, the project team developed initial data visualizations for:

- Clinician & Group Survey
- Health Plan Survey

There is a scheduled work session with the CAHPS POC on May 13, 2019, where the project team will start with a demo of these initial data visualizations. The work session will aim to flesh out the specifics of the data visualizations for CAHPS that will be developed for the Final Prototypes.

For the two CAHPS surveys, the types of the initial data visualizations and key features that are currently available in the Beta Prototype for CAHPS are indicated (x) in the table below. For visualization types and key features that are left blank, these will be finalized and filled out in the Final Prototype Report.

Data Tool	Data Visualization	Tabular Data		Graph		Map	Standard Errors Display Option	Download/ Share
		Trend Data	Cross Sectional Data	Line	Bar			
CAHPS								
	Clinician & Group Survey	x	x	x	x			x
	Health Plan Survey	x	x	x	x			x

The initial visualizations include views to top box scores and percentiles.

HCUP Fast Stats

Initial data visualizations for four statistics available in Fast Stats have been developed as part the Beta Prototype. These four are for:

Under "State Trends in Hospital Use by Payer"

- Inpatient Stay
- Emergency Department Visit

Under "National Hospital Use and Costs"

- Trends in Inpatient Stays
- Most Common Diagnoses for Inpatient Stays

Not included in the Beta Prototype are initial data visualizations for:

Under "National Hospital Use and Costs"

- Most Common Operations During Inpatient Stays

Under "Opioid-Related Hospital Use"

- Trends in Inpatient Stays and Emergency Department Visits

These other data visualizations will be included in the Final Prototype pending the work session with the Fast Stats POCs.

There is a scheduled work session with the Fast Stats POCs on May 6, 2019, where the project team will start with a demo of these initial data visualizations. The work session will aim to flesh out the specifics of the data visualizations for Fast Stats that will be developed for the Final Prototypes.

The types of the initial data visualizations and key features that are currently available in the Beta Prototype for HCUP Fast Stats are indicated (x) in the table below. For visualization types and key features that are left blank, these will be finalized and filled out in the Final Prototype Report.

Data Tool	Data Visualization	Tabular Data		Graph		Map	Standard Errors Display Option	Download/ Share
		Trend Data	Cross Sectional Data	Line	Bar			
HCUP - Fast Stats								
State Trends in Hospital Use by Payer								
	Inpatient Stay	x	x	x		x		x
	Emergency Department Visit	x	x	x		x		x
National Hospital Use and Costs								
	Trends in Inpatient Stays	x	x	x	x			x
	Most Common Diagnoses for Inpatient Stays	x	x	x	x			x
	Most Common Operations During Inpatient Stays							
Opioid-Related Hospital Use								
	Trends in Inpatient Stays and Emergency Department Visits							

NHQDR

Initial data visualizations for NHQDR were created to access and present the following levels of data:

- National
- State

There is a scheduled work session with the NHQDR POC on May 6, 2019, where the project team will start with a demo of these initial data visualizations. The work session will aim to flesh out the specifics of the data visualizations for NHQDR that will be developed for the Final Prototypes.

The types of the initial data visualizations and key features that are currently available in the Beta Prototype for NHQDR are indicated (x) in the table below. For visualization types and key features that are left blank, these will be finalized and filled out in the Final Prototype Report.

Data Tool	Data Visualization	Tabular Data		Graph		Map	Standard Errors Display Option	Download/ Share
		Trend Data	Cross Sectional Data	Line	Bar			
NHQDR								
	National	x	x	x	x		x	x
	State	x			x	x	x	x

RFTO #10: Developing Strategies for the Harmonization of AHRQ Data Tools (Data Tools)

ROM Estimates for Actual Development of AHRQ's Data Tools Portal

July 9, 2019

CONTENTS

1. INTRODUCTION

The Agency for Healthcare Research and Quality (AHRQ) provides several data tools and summary data tables that have dedicated user groups, but currently these tools are siloed and lack a common user experience (UX) and branding among them. The project's goal is to evaluate the potential for harmonizing AHRQ statistical data tools so that they use a common or limited set of user interfaces, are clearly branded as AHRQ, are less costly to expand and maintain, and more easily transferred between vendors.

The AHRQ data tools included in this project are:

- Consumer Assessment of Healthcare Providers and Systems (CAHPS)
- The Health Cost and Utilization Project's (HCUP) HCUPnet
- HCUP Fast Stats
- Medical Expenditure Panel Survey (MEPS) Summary Tables: Household Component (HC)/MEPSnet (HC)
- MEPS Summary Tables: Insurance Component (IC)/MEPSnet (IC)
- National Healthcare Quality and Disparities Reports Data Query Tool (NHQDR)

After developing the final prototype using Tableau, the project team has developed **a Rough Order of Magnitude (ROM) estimate for fully implementing the AHRQ data tools portal and building data visualizations using Tableau**.

2. ESTIMATES FOR DEVELOPING AHRQ'S DATA TOOLS PORTAL

The project team developed the following estimates by building on the outcomes and findings of the various task areas that were executed as part of the "Developing Strategies for the Harmonization of AHRQ Data Tools" project, as well as utilizing the knowledge and experience that were gained from working on the actual development of the final Tableau-based prototype. Throughout the development of the various versions of the prototype, the project team worked with the various AHRQ Points-of-Contact (POCs) to obtain as much of the specifications, requirements, and priorities that may be included in the development of an actual AHRQ Data Tools portal.

In this report, the project team is providing a Rough Order of Magnitude (ROM) that estimates the effort for building a harmonized AHRQ Data Portal using Tableau. To come up with the estimates, several assumptions were made. Please refer to the assumptions that are listed after the ROM Estimates table below.

In analyzing the estimates, please note of the following:

- The ROM estimates are provided in labor hours
- Separate ROM estimates are provided for
    - Development Project (from requirements gathering up to implementation)
    - 6-months of Operation & Maintenance (O&M) after implementation
- The "Developing Strategies for the Harmonization of AHRQ Data Tools (Data Tools)" project's resulting designs and final prototype are used as basis in estimating the ROM
- The ROM takes into consideration general estimates presented in a low and high range. This range accounts for variables and dependencies including implementation timeline, required resources and which teams will be responsible for certain tasks (government or vendor), and any additional requirements that could be identified closer to or during the actual implementation project. Specific requirements that will eventually factor in the determination of the actual estimated costs will include the features that will be required in the data portal and what actual Tableau data visualizations and functionalities may be required for each of the data tool.
- For the ROM on the O&M, the low range represents the minimum labor hours to address defects that may be identified post-launch plus a viable budget to implement simple new features and additional data visualizations for the data tools in scope. The high range for the O&M is provided to represent a more extensive support period that includes addressing defects, as well as development of more complex features identified post-launch or to potentially add initial data visualizations for any new AHRQ data tool that is not included in the scope of the development project.

- Based on Atlas' experience with the AHRQ Data Tools project and collective past experiences in developing prototypes and applications, and using the assumptions listed below, the team came up with an estimated Level of Effort (LOE) for each task area for an actual development project. Using a specific 12-month projection for the development project, the team's ROM estimates have been adjusted and shows a relatively tight variance between its low and high range endpoints.

ROM Estimates		
Development Project Task Areas	**Hours**	
	Low	**High**
1. Requirements and Design	6,000	8,000
2. Data Structure Optimization	1,000	1,500
3. Development	8,000	10,000
4. Testing and Implementation	5,000	7,000
Total Development Project Hours	**20,000**	**26,500**

O&M Hours (for 6 months)	**4,500**	**10,000**

- The team further broke down the LOE to further specify the range for each data tool per task area. The estimates for each data tool are based on the following factors which the team assigned based on information and experience gained through the development of the prototype:
  - Percent of the assumed total number of visualizations that were prototyped per data tool
  - Percent of data that was available and used in the prototypes
  - Perceived additional implementation of customized solutions that will be required specifically for the development of some data tools outside of Tableau (e.g., additional HTML-based solutions)

AHRQ Data Tools	Factor based on the Visualizations that were Developed during the Prototype	Factor based on available data in the Prototype	Factor based on additional implementation of non-Tableau solutions	Total Factor	General Percent of the Total LOE
MEPS-HC	1	0	0	1	**10%**
MEPS-IC	2	0.25	0.25	2.5	**20%**
HCUPnet	2.5	1	1	4.5	**35%**
HCUP Fast Stats	1	0	0	1	**10%**
NHQDR	1	0.25	1	2.25	**15%**
CAHPS	1	0.5	0	1.5	**10%**

- For the purpose of creating the ROM estimates, the team then used the resulting total factor per data tool to have a general idea of the percent of the LOE that it will take to develop each of them.
- Applying the General Percent of the LOE, the ROM estimates for each data tool per task area come to about:

Development Project Task Areas	Hours	
	Low	High
1. Requirements and Design	**6,000**	**8,000**
1.a. MEPS-HC	600	800
1.b. MEPS-IC	1,200	1,600
1.c. HCUPnet	2,100	2,800
1.d. HCUP Fast Stats	600	800
1.e. NHQDR	900	1,200
1.d. CAHPS	600	800
2. Data Structure Optimization	**1,000**	**1,500**
2.a. MEPS-HC	100	150
2.b. MEPS-IC	200	300
2.c. HCUPnet	350	525
2.d. HCUP Fast Stats	100	150
2.e. NHQDR	150	225
2.d. CAHPS	100	150
3. Development	**8,000**	**10,000**
3.a. MEPS-HC	800	1,000
3.b. MEPS-IC	1,600	2,000
3.c. HCUPnet	2,800	3,500
3.d. HCUP Fast Stats	800	1,000
3.e. NHQDR	1,200	1,500
3.d. CAHPS	800	1,000
4. Testing and Implementation	**5,000**	**7,000**
4.a. MEPS-HC	500	700
4.b. MEPS-IC	1,000	1,400
4.c. HCUPnet	1,750	2,450
4.d. HCUP Fast Stats	500	700
4.e. NHQDR	750	1,050
4.d. CAHPS	500	700
Total Development Project Hours	**20,000**	**26,500**

2.1 Assumptions

Here are the following assumptions in coming up with the ROM estimates:

- The development project will be completed in a 12-month period.

- The main task areas to be included in the development project are:
    - **Requirements and Design**
        - Using proposed elicitation techniques similar to ones used during the prototype development, generate and finalize the product requirements (functional, technical, system, etc.) with AHRQ Team's leadership to implement a solution for the AHRQ Data Tools Portal project, with Tableau as the data visualization product
        - Analyze the requirements and make recommendations based on prioritizations and potential value across most, if not all, of the AHRQ data tools
        - Design potential solutions for prototyping (designs are to be presented in wireframes for approval)
    - **Data Structure Optimization**
        - Analyze and understand the existing data architecture of each data
        - Design the data architecture to optimize efficient data retrieval and maintain a defined baseline system performance
    - **Development**
        - Develop the necessary code for the AHRQ data portal and all its features using similar approaches as in the prototype development
        - Create optimal data structure, build tables, and upload data for each data tool
        - Develop all Tableau-based visualizations defined within the project's scope
        - Embed the Tableau-based data visualizations and make any customizations to the Tableau embed code necessary to satisfy users' requirements
    - **Testing and Implementation**
        - System and Acceptance testing
        - Bug fixes, modifications based on user feedback, and adjustments to data/data structure
        - Migration of code base and data visualizations to production
    - **Note:** Project Management hours are included in the estimates for each task area.
- A separate ROM is included for a 6-month O&M period after implementation
- The estimates for the **Operations & Maintenance (O&M)** will include the following type of work:
    - Updates and fixes of identified bugs and defects on the data portal and Tableau visualizations
    - Implementation of new features and visualizations that will be agreed upon as can be accommodated within the O&M period

  - Refresh of data based on current data tools' update schedule
  - Provide support to AHRQ IT in their management and maintenance of Tableau and related server and network components
  - Provide support and responses to AHRQ IT in addressing technical issues and questions related to security (e.g., security controls, security certificates)

- The same data tools that were part of the prototype phase will be included:
  - MEPS HC
  - MEPS IC
  - HCUPnet
  - HCUP Fast Stats
  - NHQDR
  - CAHPS
- Incorporating SOPS and other data tools that were not part of the prototype will not be included in the ROM estimates
- Data visualizations will be built using Tableau
- 90% of the requirements and features that will be built are expected to be addressed using and configuring Tableau's out-of-the-box features (with 10% requiring customized solutions)
- Assumptions related to some key data tool-specific, non-Tableau-based features that were discussed during the course of the prototype project:
  - MEPS-HC will not have an integration with R
  - HCUP and MEPS-IC will have a Z-Test Calculator
  - The static notes and comments feature will apply similarly to all data tools (i.e., it's not a feature that will be specially implemented to only one data tool)
  - NHQDR will have state dashboard web pages
- The total amount of data that was uploaded and used in the final prototype is estimated to represent the following percentage relative to the actual data currently available for each data tool:
  - MEPS-HC: **99%**
  - MEPS-IC: **70%**
  - HCUPnet: **Less than 1%** of the metadata in HCUPnet*
  - HCUP Fast Stats: **95%**
  - NHQDR: **65%**
  - CAHPS: **20%**

  * This is due to the inclusion of detailed ICD9 and ICD10 condition codes in the current HCUPnet tool, which are not included in the Tableau prototype. However, the

POC also noted that the current loaded data in the prototype accounts for handling an estimated **70%** of the inquiries and data manipulations by users.

- The ROM estimates will factor in designing, developing, and testing the portal and Tableau views to support accessibility (508-compliance)
- The process for loading the data will be identical to how data has been loaded for the prototypes, using CSV files that will be provided by those who own/maintain the data tools' databases (it is noted that the creation of the CSV files will be more effort for certain tools compared to others)
- Data that will be uploaded to Tableau will be public-use data files and will not contain PII or sensitive data
- The harmonized data portal will be fully public facing
- The project will be utilizing iterative Agile methodologies and project management will be included throughout the task areas
- The ROM estimates will not include the following:
    - Designing and developing a native mobile application (app) version of the data portal/Tableau visualizations
    - Training
    - Procurement of Tableau License Fees

APPENDIX A: ACRONYMS AND DEFINITIONS

Table 1. List of Acronyms and Definitions

Acronym	Description
AHRQ	Agency for Healthcare Research and Quality
CAHPS	Consumer Assessment of Healthcare Providers and Systems
HC	Household Component
HCUP	Health Cost and Utilization Project
HHS	Department of Health & Human Services
IC	Insurance Component
MEPS	Medical Expenditure Panel Survey
NHQDRnet	National Healthcare Quality and Disparities Reports Data Query Tool
O&M	Operations & Maintenance
POC	Point-of-Contact
ROM	Rough Order of Magnitude
UX	User Experience