FINAL REPORT
November 2024

Data Concierge Model Report:

Two Proposed Data Concierge Models

Presented by:
NORC at the University of Chicago
Mike Berning
Seth Brohinsky
Sara Lafia
Brandon Sepulvado
Martha Stapleton
Nola du Toit

Presented to:
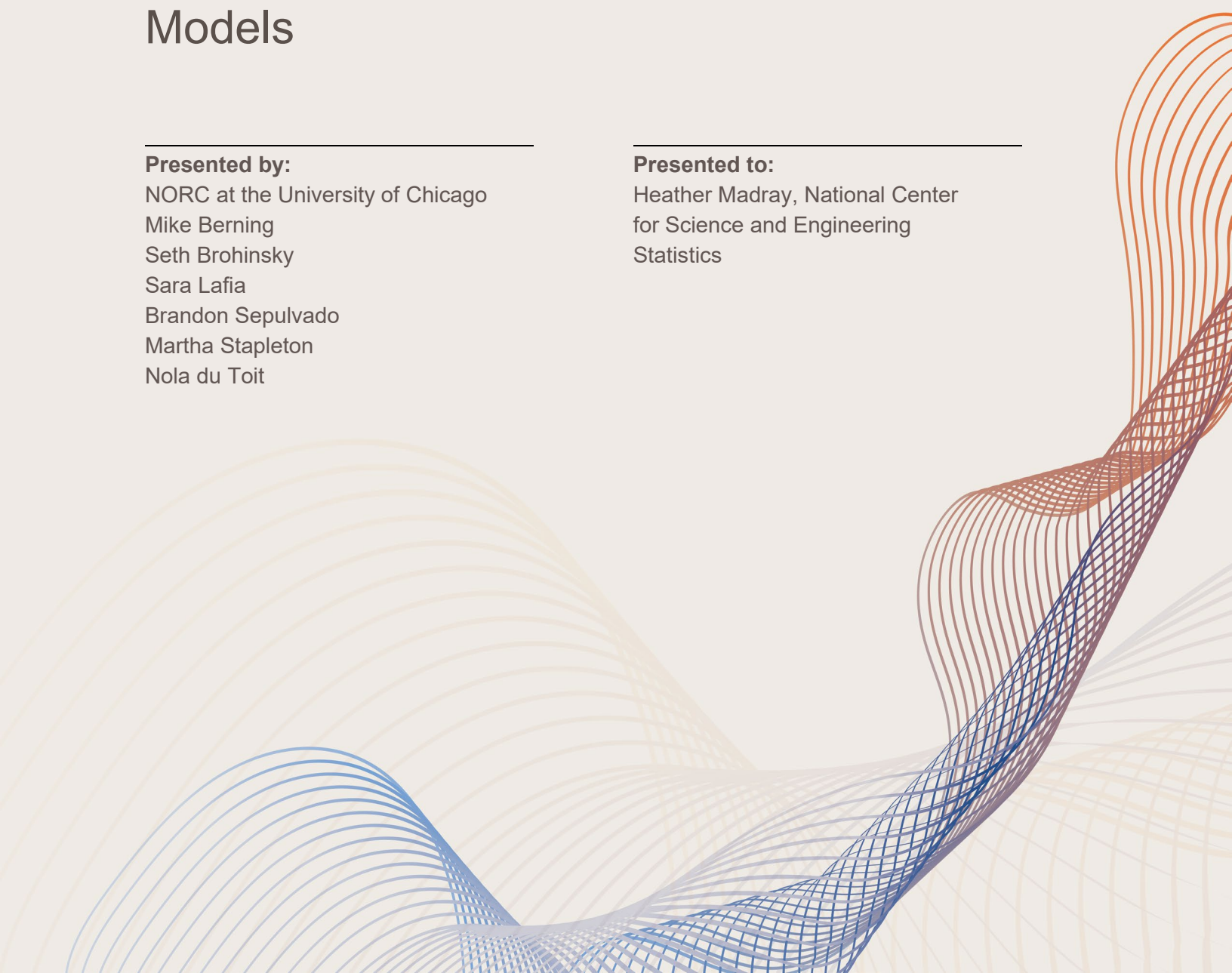Heather Madray, National Center for Science and Engineering Statistics

Table of Contents

Introduction

As part of the National Secure Data Service Demonstration (NSDS-D) project, the National Center for Science and Engineering Statistics (NCSES) has contracted with NORC at the University of Chicago (NORC) to explore models for a data concierge service for a potential, future National Secure Data Service (NSDS). The purpose of this project is to recommend approaches for providing a range of high-quality services to support data users in their pursuit of data for evidence-building research. This report drew upon three main sources – the Advisory Committee on Data for Evidence Building (ACDEB) Year 2 report recommendations for a data concierge service, synthesized feedback from interviews with statistical agencies and data users, and information gathered in an environmental scan about existing data infrastructure services. Based upon the information from those three sources, we developed a core set of data concierge service functions. This information is summarized in Part 1 and 2 of this report. Part 3 of the report synthesizes those insights to propose two Data Concierge Service (DCS) Models that build on one another - a Connector Model and a Full-Service Model. We describe how each model would implement the data concierge service core functionalities and provide details about their components and related staffing needs. Finally, user scenarios in Part 4 demonstrate the robustness of each model. The scenarios use generalized personas to illustrate how a user might interact with our proposed DCS models at different stages in the data discovery, data access, and data use continuum. Included in each scenario is a user journey diagram to facilitate a more comprehensive understanding of the model component integration.

Part 1 – ACDEB Year 2 Report

In this section, we review the guidance from the Advisory Committee on Data for Evidence Building (ACDEB) Year 2 Report related to the creation of a Data Concierge Service (DCS) for a future National Secure Data Service (NSDS).

ACDEB Year 2 Report Recommendations

The Advisory Committee on Data for Evidence Building Year 2 report lays out its recommendations for a National Secure Data Service, where a Data Concierge Service would maximize the utility of data holdings that span statistical and administrative data sources at both the federal and state levels, as trends of data use continually evolve. To match the breadth of user needs, the DCS should support a breadth of data uses, including data discovery, access, analysis, and reporting, across many data holdings and required processes. A DCS should represent a centralized, standardized, and seamless user experience that is approachable, complete, and efficient in meeting a broad range of needs for access to and use of resources.

Table 1 summarizes recommendations from the ACDEB Year 2 report for a DCS. These three recommendations reflect essential themes of user-centered design, centralized assistance with all data lifecycle processes, and resource efficiency.

Table 1. ACDEB Year 2 Recommendations

ACDEB Year 2 Section	Recommendation	Elements
3.2	NSDS Website as a "Front Door"	• Support user questions • Provide a webform to contact SMEs • Integrate lists of agency data assets • Select and link to data (catalogs, SAP)
3.3	Automation in the NSDS Intake Process	• Enable a self-service user experience (UX) • Consider the use of chatbots
3.4	NSDS Data Concierges and Service Providers and Partners	• Refer users to existing and accessible data • Prepare access agreements • Refer users to agency SMEs (for computing, tiered access, data cleaning, analysis, disclosure avoidance) • Publish service use metrics

Recommendation 3.2 – *the NSDS Website as a "Front Door"* – recognizes the breadth of user needs, as captured in different "personas", that span:

1. Interest in authoritative answers to straightforward, concrete questions
2. Discovery of all materials—including data, metadata, data access protocols, and related tools— relevant for a specific goal
3. Advanced analysis of complex and integrated data holdings

The ability to respond to this breadth of needs requires a DCS that maintains a unified account of existing knowledge, agency data asset rosters, materials, and processes related to data access. It must also sustain contact with subject-matter experts (SMEs) and connectivity to high-touch data support that serves as a single point of contact for users with all backgrounds and needs.

Recommendation 3.3 – *Automation in the NSDS Intake Process* – recognizes the potential for many solutions to be either self-service or provided via automated tools. This process of automation not only captures value from new and continually improving resources to reduce overall costs; it also ensures that available budgets maximize time coverage for high-touch support of more complex user needs.

Recommendation 3.4 – *NSDS Data Concierges and Service Providers and Partners* – recognizes the diversity of user needs at each stage of the research lifecycle. This could include a wide range of support: for example, including assistance clarifying research questions, discovering data, developing analytic plans, connecting with service providers and partners, arranging research approvals in the preparation phase; monitoring project compliance during the execution phase, helping finalize close-out activities (e.g., documentation, reporting, and assurance of reproducibility in the completion phase).

These recommendations span themes of designing resources so they completely address user needs, maintaining an expanding knowledge-base, coordinating across federal agencies, providing ongoing access to subject matter expertise, offering a range of tools that include user interfaces and organize back-end operations in serving multiple stages of complex user requests, and coordinating with staff of existing agency offerings so users experience a complete and integrated service. Over time, the DCS must be responsive to trends in user needs; the dynamics of data assets and protocols of access, usage, and reporting; and reflect continuous quality improvement by capturing and reflecting on user feedback and patterns of usage.

Part 2 – Review of Environmental Scan and Outreach

To help inform recommendations for DCS functions, NORC conducted a series of interviews with federal statistical agencies and data users. We simultaneously performed a review of the landscape of existing services. These data gathering processes identified existing agency services from which to leverage a foundation for the DCS models, uncovered gaps in services that the proposed DCS models can help fill, and revealed relevant user experiences and priorities.

Statistical Agencies and Data User Feedback

Key Stakeholder Interviews

Individual interviews with statistical agencies and data users offered insights into information needs for data discovery, data access, and data use activities, which are summarized in **Table 2**. We also learned more about the audience for each type of information needed. A general audience includes a wide range of users, while researcher-specific needs are relevant to users who are already familiar with using federal statistical data assets.

Most data users expressed a need for high-quality documentation about data and related services such as consultations across data discovery, access, and use. Users also want a way to provide feedback on their experiences using data and services. In our interviews, it was experienced researchers had interacted with confidential data the most. Although general users did not indicate the need for confidential data, most were not familiar with the standard application process (SAP). Connecting this audience with existing methods for discovering and requesting access to confidential data could lead to new types of data use.

Table 2. User information needs to address in the design of the proposed models

Persona	Information need	General Audience	Research Audience
Data discovery	Full-time equivalent (FTE) generalist and subject matter experts (SME) to answer data questions	✓	✓
	Integrated search across agency datasets, metadata, and documentation	✓	✓
	Searchable collections of data-related literature		✓

Persona	Information need	General Audience	Research Audience
Data access	Tiered access registration and login	✓	✓
	Single verified profile qualifying users for data access		✓
	Guidance navigating and revising confidential data applications		✓
Data use	Tools to explore, combine, and visualize data at a desired scale and frequency	✓	✓
	FTE generalist and subject matter experts (SME) to answer questions about data analysis and quality	✓	✓
	Tools to provide feedback, such as issues with data or services	✓	✓

Users' information needs are addressed to varying degrees by existing agency services. However, we learned from our environmental scan and interviews with federal statistical agencies that services provided by statistical agencies are not standardized and vary substantially in scale and scope. Our proposed DCS Models build on existing services to fill these gaps and provide a more consistent standard of service across the federal data ecosystem. The main areas in which existing services are provided by one or more agencies are covered in **Table 3**.

Most of the agencies we interviewed provide varying levels of support through staffed positions to assist users with data discovery and help answer user questions. Some agencies rely on documentation, such as Frequently Asked Questions, to help guide users, while other agencies staff librarians and subject matter experts who can respond to users' questions and provide consultations. A common issue arises when users ask about data outside of the agency's purview. Few agencies are equipped to address these types of questions but do try to refer users by either providing contact information or examples of other agencies' data. There are few services available, outside of the SAP, to assist users in navigating requests for confidential data access. Typically, only experienced researchers who had direct agency contacts were successful. This type of individualized service is mainly reserved for agencies that interface directly with experienced academic researchers and was described as involving an intensive level of interaction between the agency and the user. Additional services, like data linking and statistical consultation were only offered to researchers who qualified for restricted data access or were provided on a case-by-case basis for agency partnerships.

Table 3. Federal agency services to leverage in addressing users' information needs

Persona	Service	General Audience	Research Audience
Data discovery	Staff to interpret and answer users' questions	✓	✓
	Consistent, high quality, up-to-date, and complete metadata and data documentation	✓	✓
Data access	Public and restricted versions of data	✓	✓
	Consultations to help users meet requirements to gain secure access to confidential data		✓
Data use	Custom data products such as tabulations, linking, and quality assessment		✓
	Disclosure risk review and remediation prior to publishing results derived from data analysis		✓

Environmental Scan

We also reviewed examples of data concierge services offered by data infrastructure providers. We compared the software-driven and expert-driven services to determine how the solutions helped address users' information needs and the ways in which they resembled existing services offered by federal statistical agencies. Below is a brief summary; more detail about specific services is available in the Environmental Scan and Outreach Final Report.

Software Services

We found that software services play an important role in providing referrals and connecting service components. Examples include federated data catalogs for data discovery, user authentication services mediating data access, and data deidentification and online analysis tools to support data use. These examples of software connect distributed components in a service network to provide a seamless user experience. The development and adoption of software services may have a higher upfront cost but has greater potential for scalability and efficiency than staffed services alone. We will discuss relevant software solutions for the DCS in the upcoming Recommended Tools Report.

Expert Services

Whereas software solutions allow for self-service, expert services tend to be staffed and provide support in customer service, data management, and data analysis. Examples include webinars on how to use services, metadata preparation and cataloging, and consultations on confidential data access or analysis. We were not able to find detailed information about the level of staffing required to maintain

expert services, although many data infrastructure providers like ICPSR's NACJD archive and the FSRDC (Federal Statistical Research Data Center) system maintain full-time staff who provide a full range of expert services. Relevant to this report, we used these existing expert services to help define key components of our DCS models. In addition, we collected job descriptions from data infrastructure providers and have outlined the key duties and skills for DCS staff in Appendix B.

Core Functionality of a DCS

Based on the information collected during our key stakeholder interviews and environmental scan, we identified ten core components essential to a Data Concierge Service for a National Secure Data Service. Embedding these components in a DCS will best mitigate the challenges data users face when seeking to discover, access, and use federal statistical data. These core components are summarized in **Figure 1**. A complete matrix illustrating each DCS core functionality and model component is included in the **Appendix A**.

Standardize User Support. Data user support should be universal to users, standardized across agencies, and repeatable in terms of its process.

Improve Data Navigation. The current process of negotiating terms of access should be streamlined. Tiered access, through techniques such as Privacy-Preserving Record Linkage (PPRL), provide ways to help improve access to data.

Connect Users to Experts. Data users should be able to pose a research question and get connected with someone who has research expertise in that area. The expert should:
- Understand the "universe of what's available" and help a user determine whether an answer could be found in published data
- Help users "learn the ropes" and build institutional knowledge
- Guide users in the right direction and direct requests to relevant experts to save time
- Respond in a timely and tailored manner for best customer experience

Direct Users to Existing Services. Services should be well-integrated into the communities they serve by interfacing with familiar platforms and meeting users where they are by:
- Guiding general audiences who tend to use self-directed Google search or Data.gov to existing data-related products and services
- Assisting research audiences, who use word of mouth to discover and access restricted data, to the SAP (though there is a need to increase the visibility of the SAP outside of power user communities)

Enable Data Evaluation. Users should be able to evaluate different data products by reviewing additional context information, such as how data were produced and for what purpose. This core component would:

- Help refine research projects and streamline access to data by helping users identify appropriate datasets and processes for accessing them
- Assist accessing metadata documentation to evaluate the value of a data asset
- Draw from grey literature, such as documents written for sponsors, that include lessons learned, failures, and findings that aren't otherwise published

Manage User Relationships. An NSDS DCS should use customer relationship management (CRM) tools to ensure inquires receive coordinated and efficient responses. A feature of the CRM should be the ability to log events in a customer interaction database that can provide a source of training data and analytics for identifying gaps in services and assets.

Support Cross-Agency Requests. This core component would support working with data from multiple agencies by:

- Referring users to resources across agencies and helping them way-find as they navigate from the SAP Portal into agency sites
- Submitting a single request across multiple data sources
- Supporting researchers who draw from a wide variety of datasets (e.g., to study social and structural drivers of health outcomes)
- Explaining each agency's rules and disclosure authorities
- Walking a user through steps of applying for data across agencies

Communicate Coverage and Completeness. A DCS should help users understand how complete and current inventories of data are and what plans exist to add additional sources.

Create a Corporate Voice. Effectively communicating and promoting service offerings is a core component that serves to continuously remind users of its existence and utility for their evidence-building efforts by.

- Leveraging existing newsletters such as the SAP Newsletter to, for example, highlight available services, spread information on new improvements, and list contact information
- Producing shareable documentation that removes "word of mouth" communication
- Having a presence at relevant conferences, organization meetings, and events

Uphold Data Governance. A service as visible as the DCS for an NSDS should be guided by an expert steering committee that continuously evaluates and proposes services to meet user needs.

Figure 1. Core functionalities of a Data Concierge Service for an NSDS

Group Interviews

Following the development of DCS model components, NORC conducted two follow-up group interviews with select federal statistical agency representatives and data users. The goal of this second round of interviews was to confirm users' information needs and agencies' existing services and to garner feedback on the proposed DCS models and components. The group interviews reinforced our understanding that a DCS should include a staffed help desk, unify existing documentation, and provide multiple ways for users to access data products and agency-specific tools.

We heard about the importance of including mechanisms for users to give feedback on data and services. A DCS should also provide extensive documentation on data and tools that anticipate users' questions. As much as possible, this information should be derived from materials that agencies already maintain to ensure timeliness, relevance, and consistency. It would likely be harder for staff outside of an agency to maintain this level of documentation for agency-specific resources. Agencies stressed that the creation of new materials should be avoided if there are existing resources that address users' information needs. Duplication of effort and inconsistencies should be avoided by centralizing access points in a DCS.

Static documentation is unlikely to address many users' needs however, and it will be important to offer support staff who can help fill gaps. Some agencies already have dedicated staff in place who can take on this role. For these agencies, the DCS would connect general users to their existing staff, services, and tools. A DCS should help users navigate across siloed agency resources. This is important because many data users are unaware of the boundaries between agencies. A DCS should help map users' topical questions to relevant resources, which include statistical, administrative, and state-level data; computing environments; and subject matter experts housed in multiple federal agencies.

Data users mentioned that it can be difficult to access the same materials at multiple points in time because agencies change their policies or services. A DCS can help users identify "paths of least resistance" to access relevant materials that meet their needs when these kinds of changes occur. It is imperative that the DCS support all users, not just experienced researchers who already have privileged positions within the federal data ecosystem in terms of access to tools and services that cater to their needs. Individuals who work in state data centers, for example, expressed similar needs for tools and services to access restricted data through FSRDCs. However, the timeline and requirements for accessing data through an FSRDC model makes it nearly impossible for analysts at state data centers to acquire needed data in a timely manner. A DCS could provide alternative approaches to help users identify and access data needed for activities that require quick response times, like meeting state reporting requirements.

Some of the capabilities of the DCS, especially understanding and articulating user requirements, could be augmented by AI tools. The proposed DCS models should explore these capabilities to scale up and serve more users. The DCS could also function as a knowledge base of user questions and experiences with data, which could be leveraged to better meet users' needs over time. Centralizing and automating support services could provide higher quality information by helping users and support

staff obtain comprehensive answers from a single, shared source of truth. Currently, users experience varying levels of service, quality, and accuracy within single agencies based on the staff that they interact with and their knowledge of existing resources.

Part 3 – Proposed Data Concierge Service Models

Based on the feedback from statistical agencies and data users, as well as the environmental scan and ACDEB Year 2 report recommendations, we propose two data concierge service (DCS) models. The proposed Connector Model leverages existing services including data owner contact information, terms of data access and FAQs, and metadata, while the Full-Service Model proposes enhanced staffed services beyond those currently offered by agencies. These models provide groups of services with complementary sets of functions for different levels of user needs. The models have enough overlap that users can be advanced to higher levels of support when warranted. Importantly, both models fulfill the core requirements identified during the feedback sessions and environmental scan. Throughout this section, we use the icons introduced above in **Figure 1** to illustrate which core requirements are met by each model component.

As illustrated in **Figure 2**, the base components in the Connector Model provide a platform of existing resources that can be expanded with customized services into a Full-Service Model. Importantly, we envision these services as cutting across agencies and providing more comprehensive user support by building on the wide variety of services currently offered by individual agencies. We also note that the staffed services proposed in the model components of this report can be enhanced by automation. More fully developed proposals of specific software and tooling solutions will be presented in the subsequent Recommended Tools for a DCS Report.
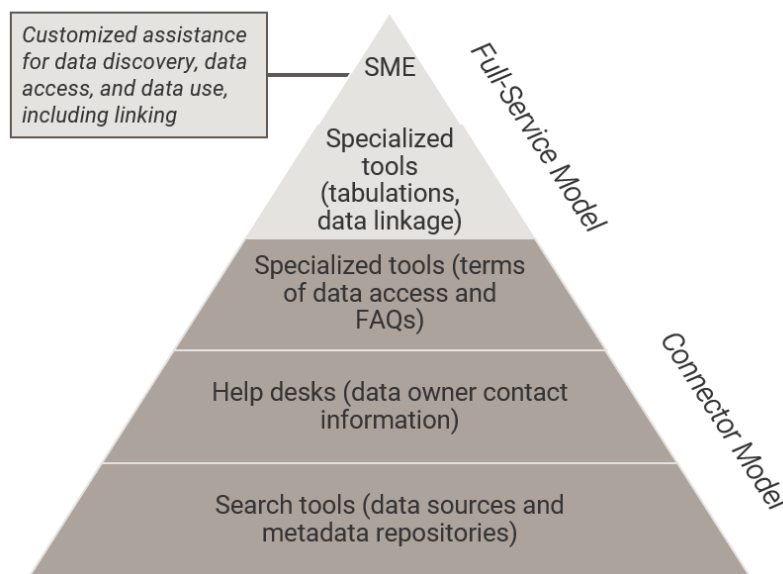
Figure 2. Components of the proposed Connector Model (*dark gray*) at the base of the pyramid and Full-Service Model (*light gray*) at the top

Figure 3 illustrates the list of actions a data user of each of the DCS Models could take across the continuum of data discovery, access, and use. The actions in the Connector Model would be primarily self-service, while the actions in the Full-Service Model would involve more interactions with staff. For example, while help desks are envisioned in both models, the Connector Model help desk would rely more on existing agency contact information to help users identify relevant experts to address detailed data-related questions. Alternatively, a help desk offered in the Full-Service Model would be more fully staffed and better able to provide thorough user consultations. We expand on these models, their components, and implementation details in the following sections.

Figure 3. Actions and services available to a user within the proposed Connector Model (*dark gray*) and Full-Service Model (*light gray*)

Connector Model

Our proposed Connector Model (CM) would establish a public facing website to centralize access to existing agency resources. The CM serves as a first-tier response to core functionalities identified in our Outreach and Environment Scan Report. There are two categories of components to the Connector Model. Category 1 includes the components that would be enabled by a public-facing Customer Relationship Management System (CRM), which DCS staff would create to triage and respond to user inquiries and maintain an FAQ knowledge base. In addition, the automated system would collect and categorize feedback to inform the quality of the CRM content and identify gaps that should be addressed by DCS staff. Category 2 includes components that establish and centralize linkages to existing data and human resources, which would assist data users in discovering, accessing, and using federal data for evidence-based research. The DCS staff could consolidate the individual links in a table that shows data sources, connected to data owner contact information, connected to the terms of use and process to access data.

Each of the components of the Connector Model presents an opportunity for tools and automation to enhance data users' experience. The next phase of this project will identify where and what software and automated tools will add efficiencies to the model components.

Customer Relationship Management System

The underlying support for the Connecter Model is a Customer Relationship Management System (CRM). The automated system would include, but not be limited to the following functionality:

- Enable data users to search a "living" knowledge base of frequently asked questions (FAQ) and other user experiences that DCS staff add to and curate over time
- Open direct interaction with DCS staff by telephone, email, or electronically through the CRM system
- Receive questions or feedback from DCS staff with the expectation of DCS response with solutions or referrals within a specified time (e.g. 48 hours from receipt)
- Categorize requests for assistance and update/edit existing resources or create new content for the knowledge base repository
- Act as a bridge from the Connector Model to the Full-Service Model user consultations
- Direct data users to relevant resources as contained within a full-service knowledge base
- Gather and report data usage metrics

Linkage Components of the Connector Model

During the research phase of the project, Federal Statistical Agencies and data users identified information needs and core competencies that should be addressed in the proposed DCS Models. Our descriptions of those recommendations and the responding proposed solutions are detailed below.

Link to the Standard Application Process (SAP) Portal

Rationale: The SAP provides access to the Data Catalog, a searchable inventory of confidential data assets from federal statistical agencies, and the SAP's application form, where users can request access to confidential data. Feedback from statistical agencies indicated that most agencies are aware of the SAP and participate to different degrees. Conversely, most data users are not aware of the SAP and have never used it to discover and/or apply for access to restricted data for evidence-based research. Importantly, the Connector Model components will support and supplement the SAP, not replace it.

| STANDARDIZE USER SUPPORT | IMPROVE DATA NAVIGATION | CONNECT USERS TO EXPERTS | DIRECT USERS TO EXISTING SERVICES | ENABLE DATA EVALUATION |
| MANAGE USER RELATIONSHIPS | SUPPORT CROSS AGENCY REQUESTS | COMMUNICATE COVERAGE AND COMPLETENESS | CREATE A CORPORATE VOICE | UPHOLD DATA GOVERNANCE |

NORC Recommendation: The DCS should contain a direct link to the SAP and DCS staff should have knowledge of the information and resources contained therein as a core component of the Connector Model. DCS staff should be well versed in the SAP Data Catalog as well as the restricted data application process. Staff should be able to advise and promote these resources and capabilities to data users seeking information about data discovery and data access.

Links to Federal and State Data Inventories

Rationale: Many data users are unaware of the full range of federal, state, and third-party data (public and confidential) available to use. Data users rely on legacy data assets and typically conduct internet searches when new data are needed. Statistical agencies consider the SAP their inventory for most confidential data. However, data users are interested in links to other federal, state, and third-party data beyond confidential data accessed through the SAP. Organizing and centralizing links to data inventories is likely to improve the user experience and ultimately advance the use of data for evidence-based research.

STANDARDIZE USER SUPPORT   IMPROVE DATA NAVIGATION   CONNECT USERS TO EXPERTS   DIRECT USERS TO EXISTING SERVICES   ENABLE DATA EVALUATION

MANAGE USER RELATIONSHIPS   SUPPORT CROSS AGENCY REQUESTS   COMMUNICATE COVERAGE AND COMPLETENESS   CREATE A CORPORATE VOICE   UPHOLD DATA GOVERNANCE
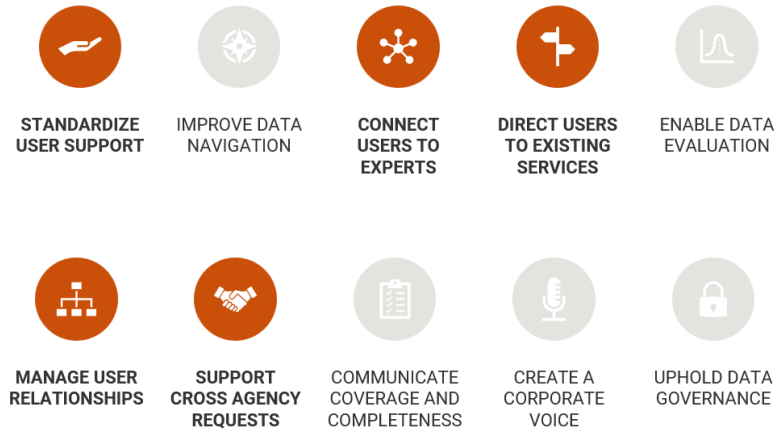
NORC Recommendation: The future DCS staff should compile, organize, and maintain links to federal (if not already in the SAP), state, and third-party data assets. The DCS staff should be familiar with the types of data available and sort the links by the categories including, and not limited to, demographic, income, health, education and business data. Importantly, the data inventory will require maintenance and expansion over time. DCS staff should consistently update the current list and search and catalog new inventories.

Link to Agency Contact Lists

Rationale: Many less experienced data users expressed frustration with how difficult and time consuming it is for them to find detailed information about the fitness for use and process to access data assets. "Power users", such as university faculty, with a long history of working with data to support evidence-based research typically have established relationships with agency contacts to aid the data acquisition process.

STANDARDIZE USER SUPPORT · IMPROVE DATA NAVIGATION · **CONNECT USERS TO EXPERTS** · **DIRECT USERS TO EXISTING SERVICES** · ENABLE DATA EVALUATION

MANAGE USER RELATIONSHIPS · **SUPPORT CROSS AGENCY REQUESTS** · COMMUNICATE COVERAGE AND COMPLETENESS · CREATE A CORPORATE VOICE · UPHOLD DATA GOVERNANCE

Most inexperienced data users do not have these relationships and described challenges with getting answers to questions about a specific data asset.

NORC Recommendation: The future DCS staff should maintain a list of contact information for specific data assets. The DCS should ensure that contact information is current and correct and provide data users with recommendations based on project requirements. Like the data inventories, the contact list should also be organized into categories including but not limited to demographic, income, health, education and business data.

Help Users Understand Data

Rationale: Data users expressed frustration at the lack, inaccuracy, or limited extent of metadata available. A common complaint was that the only way to determine fitness for use of restricted statistical metadata was to first gain access to the data. That prolonged and burdensome process often concluded with a determination that the quality and content of those data were insufficient to answer the research question.

STANDARDIZE USER SUPPORT   **IMPROVE DATA NAVIGATION**   **CONNECT USERS TO EXPERTS**   **DIRECT USERS TO EXISTING SERVICES**   **ENABLE DATA EVALUATION**

MANAGE USER RELATIONSHIPS   **SUPPORT CROSS AGENCY REQUESTS**   **COMMUNICATE COVERAGE AND COMPLETENESS**   **CREATE A CORPORATE VOICE**   **UPHOLD DATA GOVERNANCE**

NORC Recommendation: DSC staff should post links to existing metadata repositories that inform data users of the quality and fitness for use of their data for specific research projects. Types of metadata that could be considered are record layouts, data dictionaries, data profiles, and citations where data have been used to support research.

Help Users Understand Terms of Access

Rationale: Data users noted that when they identify a data asset of interest, often the path to access those data is not clear. Data users need assistance to determine the type of data access agreement that is required and the terms of access to those data.

STANDARDIZE USER SUPPORT   **IMPROVE DATA NAVIGATION**   **CONNECT USERS TO EXPERTS**   **DIRECT USERS TO EXISTING SERVICES**   **ENABLE DATA EVALUATION**

MANAGE USER RELATIONSHIPS   **SUPPORT CROSS AGENCY REQUESTS**   **COMMUNICATE COVERAGE AND COMPLETENESS**   **CREATE A CORPORATE VOICE**   **UPHOLD DATA GOVERNANCE**

NORC Recommendation: DCS staff should populate a knowledge base with background information and processes to gain access to data contained in the data inventory. The DCS should maintain and be able to recommend memorandum of agreement templates tailored to data of interest.

Provide Other Relevant Links

Rationale: During the initial phase of the project, data users suggested other links that would be helpful as Connector Model components.

STANDARDIZE USER SUPPORT

IMPROVE DATA NAVIGATION

CONNECT USERS TO EXPERTS

DIRECT USERS TO EXISTING SERVICES

ENABLE DATA EVALUATION

MANAGE USER RELATIONSHIPS

SUPPORT CROSS AGENCY REQUESTS

COMMUNICATE COVERAGE AND COMPLETENESS

CREATE A CORPORATE VOICE

UPHOLD DATA GOVERNANCE

NORC Recommendations: DCS staff should develop and post links to materials that include but are not limited to:

- Understanding the NSDS and role and responsibilities of the data concierge service.
- Data Gems: Short videos offering guidance on the use of data to support research (e.g., drawing from the SAP newsletter and highlighting data sources from all agencies).
- Tiered Access: Recommend range of options for accessing restricted data, such as personal protection privacy linkage methods.
- Impact of disclosure avoidance methods to final product (content and schedule), such as what can and cannot be released.
- Data visualization tools, together with high-level information about their functionality and their use.

Staffing the Connector Model

The staffing complement for the Connector Model should consider the following skillsets:

- Web development to create a web presence that will support the connector model content and to be scalable to add full-service model content.
- Data program management to identify content of interest and establish links to that content.
- Metadata management to identify and maintain links to metadata repositories.
- Data science to advise on analysis and to create custom products as required.
- Customer service to manage requests for assistance, route those requests to the proper specialists, and track the status of response.

Full-Service Model

The Full-Service Model builds on the existing services leveraged in the Connector Model to provide additional tailored expertise. These additional capabilities were identified during our outreach and environmental scan activities as important to data users and federal statistical agencies. The main category of functionality introduced in the Full-Service Model is enhanced user support, which is achieved through expanded staffing via a network of subject matter experts. Such a network is not envisioned to supplant existing services but rather can help fill gaps in the user experience and respond to user requests. The main areas that the Full-Service Model addresses include high touch support for data discovery, data access, and data use needs. Staff services will use the scaffolding provided by the Connector Model to enable a secondary tier response to user needs and establish shared services, such as a network of subject matter experts to help provide more in-depth responses to user questions. The automated software and tools that could enable or expand each of the services here will be more fully developed and elaborated on in the upcoming Recommended Tools Report.

Network of Subject Matter Experts

The underlying support for the Full-Service Model is a network of subject matter experts (SMEs), which could include the following services and functionality built atop the Connector Model:

- A bridge from the Connector Model to the Full-Service Model in the form of curated resources and a customer relationship management (CRM) system.
- The ability for users to submit general research questions about topics of interest and receive recommendations for relevant data assets along with guidance on how to get started with accessing the data within a specified time (e.g. 48 hours from receipt).
- The ability for users to submit and get answers to questions about known data assets related to their acquisition, requirements, or analysis within a specified time (e.g. 48 hours from receipt).
- The ability for users to request custom tabulations of restricted data through a Custom Tabulation Service (CTS).
- Capability for data users to directly interact with the subject matter expert network by either telephone, email, or electronically through the CRM system.
- Requests for assistance using this capability will be addressed on a case-by-case basis and responses will be inventoried in a CRM system.
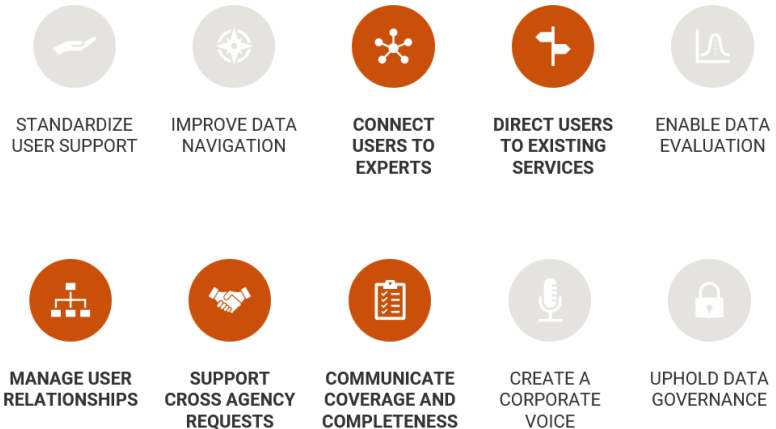
One model for a subject matter network and a potential source of existing expertise that the Full-Service Model may leverage is the network of FSRDC administrators who serve as points of contact on behalf of partner institutions (https://www.census.gov/about/adrm/fsrdc/contact.html). Another potential source of such expertise might be organizations who already provide data concierge services for particular data or domains. Two tiers of staff could be established as part of this network, comprised of dedicated staff from participating agencies and another tier of future DCS staff who have specialized subject matter expertise.

Concierge Components of the Full-Service Model

During the outreach and environmental scan, statistical agencies and data users recommended information needs and core competencies that should be addressed in the DCS models. Provided below are our additional recommendations that constitute the Full-Service Model.

Assist Users with Data Discovery

Rationale: Data users often struggle to articulate and then match their needs to existing data during the data discovery process. Users would benefit from advanced help desk capabilities where they could ask a wide variety of questions, both about general research topics and about specific data assets once they have been identified. Supporting users in understanding when publicly available versions of data could meet their needs rather than applying for access to confidential data could expedite and improve the data discovery process. A DCS could also support users in evaluating data quality during the discovery process, specifically data fitness for use. Tools such as the FCSM Data Quality Assessment Tool[1] and visual and interactive tools for assessing data quality[2] can support this process. This is important when users are considering data collected for statistical purposes as well as public or private program administration data. A DCS could leverage these tools to guide users through considerations for appropriate data reuse. A more complete discussion of tools will be provided in the upcoming Recommended Tools Report.

STANDARDIZE USER SUPPORT

IMPROVE DATA NAVIGATION

CONNECT USERS TO EXPERTS

DIRECT USERS TO EXISTING SERVICES

ENABLE DATA EVALUATION

MANAGE USER RELATIONSHIPS

SUPPORT CROSS AGENCY REQUESTS

COMMUNICATE COVERAGE AND COMPLETENESS

CREATE A CORPORATE VOICE

UPHOLD DATA GOVERNANCE

NORC's analysis of the services that agencies currently offer to support data discovery assistance, along with users' expressed needs for additional support, particularly for cross agency requests, indicate a specialized offering in a Full-Service Model that provides additional assistance to help users search for and identify relevant data assets.

---

[1] Iwig, W., Berning, M., Marck, P., & Prell, M. (2023). Data quality assessment tool for administrative data.

[2] Seeskin, Z. H., & Rice, K. Visual and Interactive Tools for Assessing Data Quality.

NORC Recommendation: DCS should provide a dedicated help desk where users can ask general, cross-agency questions framed around topics as well as specific questions about data assets. This service could be integrated with existing search catalogs so that users who are initiating a self-service data discovery process could be connected to more specialized data discovery support in the form of an inquiry. Many agencies provide some level of user support but are not typically able to support cross-agency requests or answer questions about data from other agencies. Furthermore, staff who are familiar with confidential data may be siloed from staff who work with public data. This requirement would mean having dedicated staff in place who have the requisite knowledge to address these types of questions, as well as having a system in place to help users make requests for assistance to refine their requirements.

Assist Users with Refining Requirements

Rationale: Part of the challenge users face includes having an incomplete understanding of the full scope of available data assets. This includes both confidential and public use data, as well as data products and reports that are relevant to their needs. Some users may not necessarily require data access to answer their questions and could have their needs met through existing tabulated, summary level data or a

STANDARDIZE USER SUPPORT · IMPROVE DATA NAVIGATION · **CONNECT USERS TO EXPERTS** · DIRECT USERS TO EXISTING SERVICES · ENABLE DATA EVALUATION

MANAGE USER RELATIONSHIPS · SUPPORT CROSS AGENCY REQUESTS · COMMUNICATE COVERAGE AND COMPLETENESS · CREATE A CORPORATE VOICE · UPHOLD DATA GOVERNANCE

relevant report. Additionally, users are often unfamiliar with all requirements associated with requesting access to confidential data, for example, and would benefit from understanding this prior to initiating a request. A consultation service would help match users' requests to a full set of relevant data assets and data products so they can make an informed decision about fitness for purpose.

NORC Recommendation: The DCS should provide a consultation service where users can describe their requirements and receive expert advice about the tradeoffs associated with specific data assets. This service would allow users to be more proactive and anticipate some of the challenges that may arise prior to initiating requests for data. This service would also help make recommendations about which data may be best fit for purpose given a user's research question. To provide this service, the DCS would draw on a subject matter network of staff familiar with a wide range of data assets who can provide users with relevant context and guide them through the various restrictions associated with confidential data use.

Negotiate Data Use Agreements

Rationale: Many agencies currently rely on the SAP to guide users through requests for confidential data; however, some agencies also described cases where they work with qualified researchers or other agencies to provide access to data on a case-by-case basis. Consequently, many potential data users who are unfamiliar with the mechanisms for partnering with agencies to request access to restricted data, or who run into roadblocks when applying for access to data through the SAP, may give up on their requests or use other, less suitable data for their needs. Users would benefit from a Custom Tabulation Service (CTS), which would allow them to access the appropriate level of data for their needs rather than negotiate direct access to microdata, for example.

STANDARDIZE USER SUPPORT IMPROVE DATA NAVIGATION CONNECT USERS TO EXPERTS DIRECT USERS TO EXISTING SERVICES ENABLE DATA EVALUATION

MANAGE USER RELATIONSHIPS SUPPORT CROSS AGENCY REQUESTS COMMUNICATE COVERAGE AND COMPLETENESS CREATE A CORPORATE VOICE UPHOLD DATA GOVERNANCE

NORC Recommendation: The DCS could provide a Custom Tabulation Service (CTS) within a consultation service where prospective data users could submit requests and receive feedback on the appropriate tier of data access based on their qualifications and needs. If a user qualifies for access to microdata, the CTS could help expedite user access to the data in an appropriate computing environment through centralized credentialing for unique, approved users. Otherwise, the service could make a recommendation about the appropriate level of access and provide users with a custom tabulation using the confidential data.

Consult on Data Linkage

Rationale: While relatively few users interviewed described linking data, those that did mentioned how difficult it can be to assess the quality of linked data assets, especially when linkage is performed on their behalf. Additionally, few agencies provided explicit mechanisms for linking data assets such as data dictionaries or tools. Users would benefit from a consultation service and access to a suite of tools to help guide record linkage activities and quality assessment.

| STANDARDIZE USER SUPPORT | IMPROVE DATA NAVIGATION | CONNECT USERS TO EXPERTS | DIRECT USERS TO EXISTING SERVICES | ENABLE DATA EVALUATION |
| MANAGE USER RELATIONSHIPS | SUPPORT CROSS AGENCY REQUESTS | COMMUNICATE COVERAGE AND COMPLETENESS | CREATE A CORPORATE VOICE | UPHOLD DATA GOVERNANCE |

NORC Recommendation: The DCS could develop and catalog measures of data linkage quality. Those measures may include, but not be limited to:

- Degree of coverage of various characteristics (e.g. demographic, geography) of source files available for linkage.
- Available link keys on each source file.
- Where linked files already exist, the match rate of that linkage.

Navigating Tiered Access Via Privacy Preserving Tools

Rationale: Data users expressed concern with the lengthy and complicated process to access confidential data for linkage projects. The personally identifiable information (PII) contained in confidential data optimizes the quality of the linkage and it is those data elements that pose the greatest disclosure risk. Additionally, it is the reason confidential data have restricted terms and conditions for access. A solution or mitigation to these restrictive terms is privacy preserving techniques, such as privacy preserving record linkage (PPRL) applications. ACDEB Recommendation 1.7 states: "OMB, in coordination with the ICSP should promote the use of privacy-preserving technologies in

| STANDARDIZE USER SUPPORT | IMPROVE DATA NAVIGATION | CONNECT USERS TO EXPERTS | DIRECT USERS TO EXISTING SERVICES | ENABLE DATA EVALUATION |
| MANAGE USER RELATIONSHIPS | SUPPORT CROSS AGENCY REQUESTS | COMMUNICATE COVERAGE AND COMPLETENESS | CREATE A CORPORATE VOICE | UPHOLD DATA GOVERNANCE |

the tiered access framework required under Title III of the Evidence Act by identifying an initial set of promising tools over the next 1 to 3 years."

NORC's analysis found that a specialized service to advise data users of existing PPRL capability and/or create and manage an NSDS-owned PPRL would help users acquire a high quality, anonymized linkage key usable across multiple data files.

> **NORC Recommendation:** The DCS could offer data users advice and recommendations about existing PPRL solutions. The DCS would also establish its own organic tiered access tools, including a PPRL capability, to be available within the NSDS secure computing environment (SCE). The organic NSDS tool responds to limitations of using a third party PPRL platform noted during NORC's Foreign Born Scientists and Engineer project. In that project, the terms to access sensitive data complicated third party PPRL solutions for accessing those data.
>
> Tiered access capability will enable the DCS to accept sensitive data from a data owner and process the data through a privacy preservation process. For the PPRL, the DCS will append an anonymous link key and drop the PII from the records. The processed version of the file can either be returned to the data owner or placed in the SCE to support an existing project as approved by the data owner.
>
> The DCS role may be to facilitate access to the NSDS PPRL for approved data users. There will likely be cases where confidential data cannot leave their protected environment for a PPRL process. In cases where the sensitive data are too restricted to transfer to the SCE solution (bringing the data to the solution), the DCS staff will explore the potential and manage the process to download the PPRL application to the user's secure environment (bringing the solution to the data) for linkage projects.

Staffing the Full-Service Model

The staffing complement for the Full-Service Model should consider the enhanced skillsets are listed below. For a more comprehensive discussion of job descriptions and requirements, see **Appendix B**.

- Policy experience to guide compliance requests and help users and agencies navigate data use and sharing agreements.
- Customer service experience to respond to requests pertaining to more than one agency, or which would require knowledge of state, local, or tribal data.
- Technical experience to help users access restricted data in secure computing environments and address analytical questions that arise.
- Subject matter expertise in the areas of health, income demographics, education, and business, which could be matrixed from relevant agencies; these staff would augment the core DCS staff as expert advisors in these subject areas.

Part 4 – Personas

To illustrate how users can discover, access, and use data for evidence-building activities within each proposed model, we have developed scenarios consisting of user personas and associated user journey diagrams. The personas cover users who may be interested in obtaining authoritative information in response to a question (*data discovery*), users who require support in accessing known data assets (*data access*), and users interested in analyzing one or more data assets (*data use*). The user journey diagrams provide examples of services from each model that would support users' needs. Each scenario is not exhaustive and does not touch on all model components; rather, we select relevant components to help illustrate how they function together to serve users. Importantly, the diagrams also provide an overview of the sequence in which each persona might functionally engage different components of the proposed models given their specific information needs. Each diagram (**Figures 4, 5,** and **6**) shows the user services and how they are linked.

Data Discovery

Scenario: A user with a general question wants to explore topically relevant resources to learn more about how to begin approaching their question. The user is not familiar with the federal data system and is not sure how to search for relevant resources that could help address their question.

> "What's the median wage for a graduate of Great State University?"

Connector Model

The user starts by browsing data-related literature using keyword searches. Through this process, they search for data and variables about wages and university graduates in a federated data inventory and identify agencies that produce relevant data. They then reach out to ask general questions about the data such as its coverage, completeness, and availability. This path is illustrated with the orange arrows in **Figure 4**.

Full-Service Model

After proceeding through self-service search and identifying relevant agencies, the user can connect with a subject matter expert (SME) who is familiar with the data sources and can provide a consultation on the research topic. The SME also points the user to auxiliary data sources at the state level that have high fitness for their analysis. This path is illustrated with a dashed line in **Figure 4**.

Figure 4. Data discovery scenarios in the Connector Model and Full-Service Model
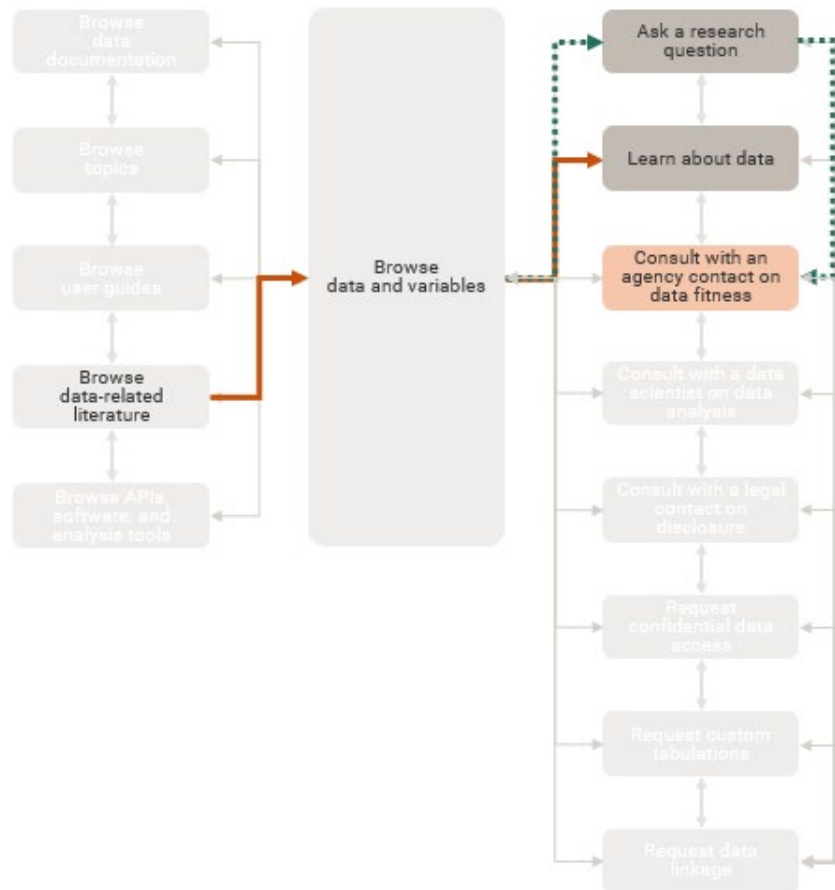
Data Access

Scenario: A user has a specific need to access data that they have already identified as relevant. The user is familiar with federal statistical data, is looking for guidance on how to negotiate access to the data if it is confidential and wants to learn more about how to qualify for access to the data. If accessing this specific data asset proves to be too difficult, the user would also like to know whether other existing data might be useful.

> "How can I access data from the NCES School Survey on Crime and Safety?"

Connector Model

The user starts from a self-service search of documentation to learn more about terms of access for the data asset in question. They also conduct general searches to learn more about how they reach out to a relevant agency contact for assistance and then request access to the data asset through the SAP. This path is illustrated with the orange arrows in **Figure 5**.

Full-Service Model

After reviewing self-service documentation, the user schedules a consultation with a Subject Matter Expert who clarifies the user's needs. After this review of the user's request, the expert makes recommendations as to whether the user should request access to the confidential data or whether they would be well-served by a custom tabulation. Based on the determination, the SME refers the user to the Custom Tabulation Service where they have the opportunity for additional consultation. This path is illustrated with a dashed line in **Figure 5**.

Figure 5. Data access scenarios in the Connector Model and Full-Service Model
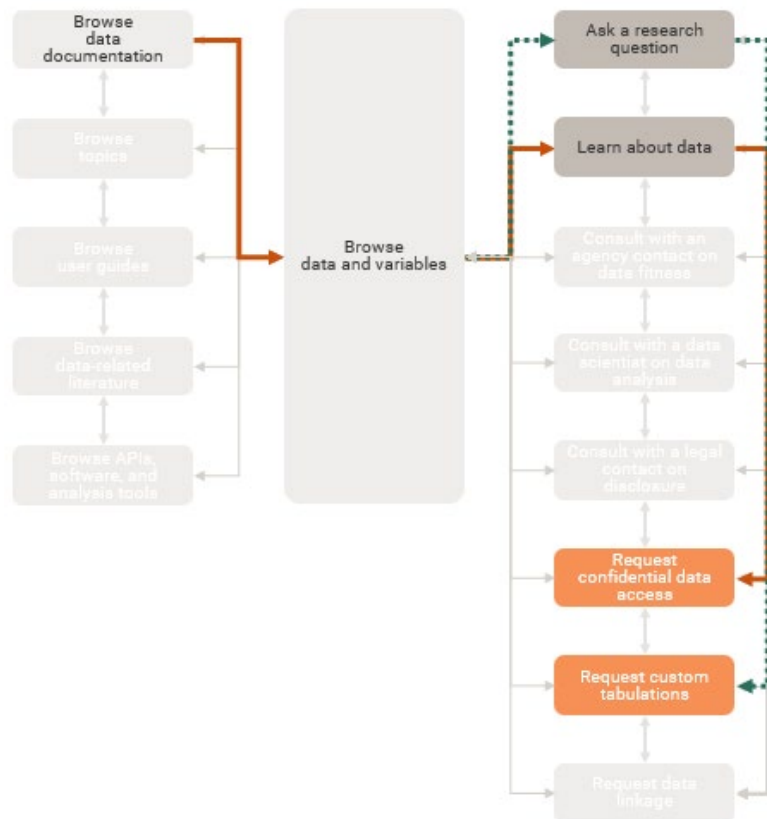
Data Use

Scenario: A user has a specific research question spanning the usage of data assets from multiple agencies. They are familiar with the types of data available from various agencies but are not familiar with data linkage. Their question can only be addressed by jointly linking and analyzing these data assets. The user is not aware of the accessibility of the data assets in question and is unsure how to proceed with their analysis if they are unable to access the required data assets.

> "How many farmers in USDA's Census of Agriculture receive Veteran's Affairs benefits?"

Connector Model

The user starts by searching a self-service catalog for relevant data linking tools. From the self-service catalog, they can narrow down the number of data assets to consider and review relevant documentation explaining data availability and variables available for linking. Based on available variables, the user can proceed with following agency-specific guidance for linking and analyzing the data. This path is illustrated with the orange arrows in **Figure 6**.

Figure 6. Data use scenarios in the Connector Model and Full-Service Model

Full-Service Model

After identifying relevant data assets using the self-service catalog, the user can request a consultation with a SME to obtain guidance on the feasibility of linking data given the variables and documentation provided in each data asset. The user performs the linkage, possibly using a secure computing environment suggested by the SME and can follow up for additional guidance in reviewing the quality of the linkage. This path is illustrated with a dashed line in **Figure 6**.

Part 5 – Future Phases

For the next phase of this work, we recommend identifying a partner within the federal statistical system that would work with NCSES and NORC to implement a pilot of the Connector Model. The pilot project could begin identifying and compiling the linkage components of the Connector Model, while at the same time developing a future public facing website as the entry point for the Data Concierge Service. To prepare for an NSDS 1.0 launch, this site should be developed to provide intuitive access to the suite of services and/or tools that constitute the DCS and seamless integration into the larger network of NSDS offerings, including a potential broader "front door" to the NSDS as a whole. The website could include:

- Information on the DCS and NSDS generally;
- DCS helpdesk contact information;
- Links to the SAP;
- Links to a complete list of the Agency's data inventory;
- Links and contact information for Agency representatives;
- Links to any additional data-related services the Agency offers;

Once the website, data inventory, and contact information are compiled, the pilot could include outreach to potential users to elicit user feedback. We could also consider developing a communication plan around the pilot, to raise awareness among the broader research community, as just one element of preparing potential and future users for an eventual NSDS. Outcomes from the pilot would include not only recommendations for improvements to the model itself, but also the process for engaging the research community with such a service. Additionally, the pilot could serve as a testbed to support other NSDS-related shared services. For instance, user information shared with DCS staff could inform account provisioning in the secure computing environment. Finally, the pilot project could develop detailed job descriptions for initial DCS staff. Key elements have been identified in **Appendix B**, and the pilot project would more fully refine the descriptions into detailed job listings.

Appendix A

Core Functionality Overview

The following table summarizes the model components described graphically in the Proposed Data Concierge Service Models, Connector Model and Full-Service Model, sections of the report.

	Link to SAP	Link to federal and state data inventories	Link to agency contact lists	Help users understand data	Help users understand terms of access	Provide other relevant links	Assist users with data discovery	Assist users with refining requirements	Negotiate data use agreements	Consult on data linkage	Navigating Tiered Access
STANDARDIZE USER SUPPORT	●	●	●	●	●	●					●
IMPROVE DATA NAVIGATION	●				●	●			●		●
CONNECT USERS TO EXPERTS	●	●	●		●	●	●	●	●	●	●
DIRECT USERS TO EXISTING TOOLS	●	●	●	●	●	●	●			●	●
ENABLE DATA EVALUATION				●	●	●				●	●
MANAGE USER RELATIONSHIPS	●	●	●	●	●	●	●	●	●		●
SUPPORT CROSS AGENCY REQUESTS		●	●	●	●	●	●	●		●	●
COMMUNICATE COVERAGE AND COMPLETENESS				●	●		●	●		●	
CREATE A CORPORATE VOICE					●	●					●
UPHOLD DATA GOVERNANCE					●	●			●		●

Appendix B

Job Descriptions

The following duties and skills for staffing a DCS have been gleaned from existing descriptions of relevant positions with technical, policy, and customer service requirements. These lists have been provided in anticipation of the types of tasks and requests that staff hired to support a potential future DCS may support and could be used to help develop future materials for recruitment.

Duties	Skills
Technical	• Prepare, archive, and share data • Review customer experience and engagement analytics • Provide access to restricted data through secure computing environments • Provide usability assessment for materials and end-user research
Policy	• Ensure data access agreements comply with policies and regulations • Review and negotiate data use agreements with data providers • Ensure adherence to policies for handling confidential data
Customer service	• Create instructional guides and materials for data end-users • Serve as initial point of contact for users by telephone or help desk • Gather information from users and route inquiries • Respond to data requests with recommendations • Serve as a data point contact for external requests from other agencies • Monitor and resolve user requests • Review data requests and project requirements • Acquire data from other organizations to meet end-user requirements • Conduct data searches and distribute data files • Process applications for restricted-use data • Engage with submitters to resolve application issues

Requirements	Elements
Experience	• Subject matter expertise (Criminal Justice, Public Health, Education) • Library and information sciences • Restricted data user support and data access agreements • Confidentiality, disclosure, and related data distribution issues • Training and workshop coordination
Web analytics	• Google Analytics
Data analysis and statistical applications	• SPSS • Stata • SAS • R
Data management	• Cataloging • DBMS • SQL
Service management	• Jira • Asana • Smartsheet
Training and workshop coordination	• Curriculum development and instructional design • Audience engagement • Logistics management