

Dissemination Modernization at the U.S. Census Bureau

Liang-Chi Wang¹, U.S. Census Bureau*

Abstract

The Center for Enterprise Dissemination Services and Consumer Innovation (CEDSCI) at the U.S. Census Bureau is an innovative effort to transform and expand the dissemination of Census content and better serve our internal and external customers:

- Enable data to be easily discovered, accessed, and consumed
- Make data more useful for a diverse set of customer needs
- Take advantage of digital opportunities for securely interfacing with public and other government sources
- Centralize and standardize the metadata

Current Business Challenge

We currently have multiple disparate systems and processes not designed for interoperability.

- Over 100 applications to prepare and release data
- Lack of standardization and consistency across dissemination processes and sources
- Complicated and challenging to find data and create insights, especially across disparate programs, products, and geographies.

What we are trying to do:

- Combine and analyze data across statistical surveys, censuses, other programs, and external sources
- Maximize consistency in geographies among data sets, and programs to provide geographic tools to users
- Innovate and quickly provide products, customizable reports, visualizations and tools that respond to new and emerging data analytic needs of users
- Make available data, metadata, and analytic tools that are easy for data users to understand, locate, and use
- Release data quickly and efficiently following data collection while maintaining statistical data and confidentiality standards
- Test fully the integration, performance, and end-user experience of tools and products
- Collect, analyze, and integrate multi-channel customer feedback passively and actively in order to derive insights that are reported enterprise-wide

Intended benefits include:

- Cost savings through elimination of duplicate systems and processes
- Spurring greater consumer innovation

- Systematic quality assurance
- Improved customer satisfaction through consistent user experience
- Better utilization of existing tools to meet customer needs
- Greater insights into customer needs

¹ Liang-Chi Wang is Assistant Division Chief, Dissemination Development and Technical Integration at the U.S. Census Bureau in Suitland, MD. liang-chi.wang@census.gov

** Any views expressed are those of the author(s) and not necessarily those of the U.S. Census Bureau.*

CEDSCI

In 2013, U.S. Census Bureau leadership appointed a special task force, Data Dissemination Task Force (DDTF), which consisted of Bureau professionals from across the agency to look into the future of data dissemination and emerging technologies. Their charge was to help define how data dissemination services could evolve to better serve customers, including defining the core capabilities required in the coming years. In 2014, the DDTF produced a set of working papers and a plan to provide an approach to restructure and enhance dissemination activities across the entire agency to improve customer satisfaction, and grow our audience and customer base.

The plan calls for a multiyear, innovative effort to transform and expand the dissemination of Census content. It empowers data users to find, to access, to connect, and to use data in ways that better their lives and communities, and it stimulates economic invention and growth. Such a comprehensive transformation requires leadership and insight to achieve an efficient yet impactful outcome. The changes touch nearly every office and branch across Census headquarters and in the various census regions covering the U.S. To manage this effort, we have formed a new, collaborative Center for Dissemination Services and Consumer Innovation (CEDSCI) with representation from across multiple organizational directorates.

The CEDSCI team is an experienced set of Census experts with extensive knowledge of internal dissemination systems and external customer needs. Located within the Decennial Directorate, its diverse membership is also composed of technologists, scientists, and analysts from other Directorates including, Economic, Demographic, Communications, Research and Methodology, and other offices. Their challenge is to modernize our dissemination efforts through enabling technologies, governance, metadata standardization, and digital transformation integration.

The team needed to create a phased approach evaluating current technologies and existing innovations to provide a core set of shared dissemination services that enables both Census staff and our data consumers to do more with the massive amount of content we publish year round. This approach will allow a graceful transition that moves us from the suite of current systems and tools to a unified and more sustainable approach. The phased approach will also minimize risks across programs and surveys while standardizing and modernizing dissemination processes and technologies.

The Census Bureau's core mission is to produce high quality, trusted, independent measures of America's people, places, and economy. CEDSCI is helping this mission by transforming and expanding the dissemination of Census data to better serve our customers by enabling data to be easily discovered, accessed, and consumed.

Current Business Challenges

The Census Bureau had already begun making changes to information dissemination before the Digital Government Strategy was released, and many of the changes were directly aligned with the CEDSCI strategy. One of the challenges is to respond to rapidly changing technology in the consumer marketplace. Over the years, agency projects and programs had unique approaches to resolving consumer needs, which resulted in a variety of technologies, tools, standards, and user experiences. The lack of consistency across these dissemination processes and sources made it challenging for the Bureau to adapt to these marketplace shifts, has made finding and accessing information challenging because each tool has different features and core sets of capabilities. This approach makes it challenging for customers to find, access, and use census content.

Data are a significant commodity in the “information economy” that spurs innovation and the creation of new business enterprises. Keeping pace with these demands is important to the Census Bureau’s core business - data collection and dissemination. Strong Census Bureau brand awareness and acceptance along with high data quality drive higher response rates. The more households and businesses value and use Census data, the higher the likelihood they will cooperate with requests to participate in one of our surveys. In many ways, CEDSCI is perfectly paired with survey data collection processes. Both collection and dissemination efforts are connected in the sense they will help each other deliver efficiency and quality improvements, and ultimately will enhance the customer’s experience when interacting with the Census Bureau – and grow our customer base.

Technology Approach

Enabling Environment

CEDSCI is currently developing an enabling environment that is flexible and extensible. The environment is being developed in a componentized way so that as technology changes the environment can easily adapt. To further validate the flexibility of the environment, the phased approach exercises the componentized architecture by replacing older technologies with newer ones or enhances current environment with new features.

Some of the business challenges the technologies are addressing includes: enterprise metadata management, information discovery, and information access. Each business challenge is solved by different collections of technologies and evolving at a different pace. Even though each business challenge has different collections of technologies, CEDSCI has standardized the software stack, metadata, geography, and system interface parameters. Lastly, the environment will serve multiple presentations with a consistent experience through standard pattern libraries.

The dissemination environment will allow Census developers to innovate and quickly provide products and tools that respond to the emergent data needs of users in a consistent environment and experience. It will also allow the rapid and efficient release of data following Census data collection efforts while maintaining statistical and confidentiality standards. Lastly, it will provide a seamless user experience for the consumers of Census information.

Enterprise Metadata Management

One of the first goals is to establish a managed enterprise metadata repository that will be used for all information that is disseminated. The repository will consist of both geographic reference metadata and statistical metadata standards. This information will be the basis for all structured (statistical data) and unstructured (reports, visualizations, tools, etc.) content and will allow users to find the information they need more easily. Metadata standards also allow for improved data interoperability through content harmonization efforts and relationship mapping. The metadata will be used in several ways to help customers find information, including standardized searchable topics, faceted filtering, a unified search experience, and geographic equivalence.

The metadata model creates the foundation to allow for standardization across data products and geography. Geographic reference metadata repository allows the bridge between statistical and spatial information. The combination of the two repositories will also help prevent common issues like duplicate and inconsistent labeling. In addition, commonly defined code sets and geographies will allow for joining information from different datasets. By using consistent topics, code sets, and labels users will be able to find information more easily, and better understand similarities and differences across data sets and their contents.

Information Discovery

A primary goal of the platform is to enable users to search for information across different content types and sources without requiring the user to know or understand the Census Bureau's data holdings or structures. The enterprise metadata model is designed with this in mind, taking into account the requirements that comes from allowing users to use natural language to search against the data, but also the known business rules that will be placed on the catalog of metadata information by the Bureau's dissemination standards. This structure also allows for these dissemination standards to be maintained in a single location, and reused by all consumers of the data platform, and by using standard definitions and a systematic approach, a consumer is able to find the most accurate information based on their search query.

Parsing user queries and deriving intent is very important to the success of the search-first approach the platform is assuming. By using the user input as the primary search signal, the search can infer the user's preferred content type such as a map, a table, or webpage results, along with the relevant content within each type. By looking at the metrics of the search with website statistics, the platform can tune the dissemination environment to better serve our consumers.

Technology Environment

As part and in support of the above goals, CEDSCI is developing shared enabling technologies as a dissemination environment for the enterprise. The shared services are developed from a combination of data provider requirements and end user requirements, and the platform will serve as the intersection point for both parties. This will also create an ecosystem where data are not just consumed, but where usage statistics and analytics are captured and used to inform program areas as to how their data products are being used and what modifications may be helpful to our end users.

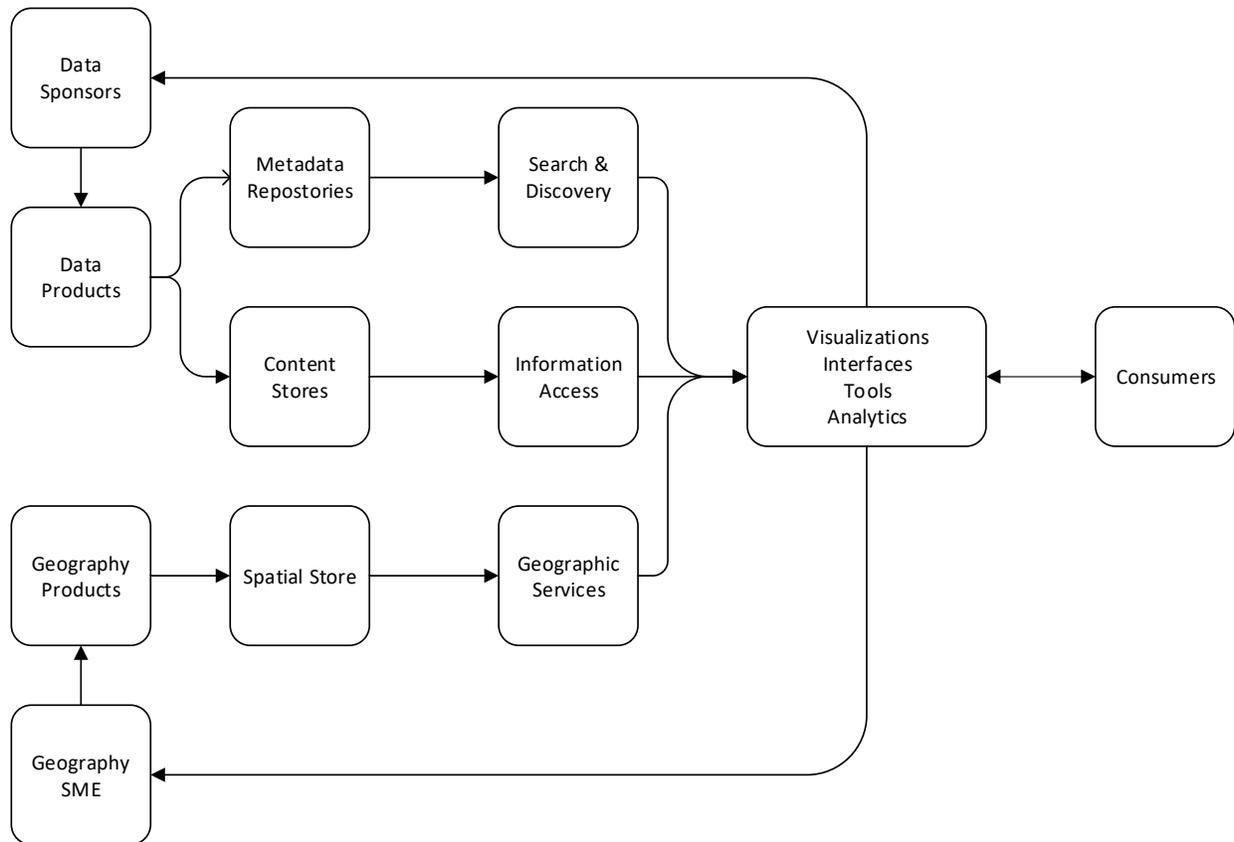


Figure 1: Dissemination life cycle

To construct the new dissemination platform, the Census Bureau is employing an agile development methodology that emphasizes iterative development and integrated customer feedback. Our two-week sprints aggressively incorporate datasets from Census Bureau programs. Our goal is to have a shippable codebase every forty business days allowing customers to see the progress.

As we modernize and centralize our dissemination efforts, we must similarly keep pace with technological innovations and changing expectations from our data sponsors and consumers. Deploying a componentized shared services platform across our diverse programs and datasets not only provides us with internal efficiencies, but it also allows us to more quickly respond to the changing digital landscape. We will demonstrate that the Census Bureau’s mission and activities delivers tangible value to the public, the private sector, data sponsors, and educators through the innovative and adaptive design patterns we are instantiating, while also meeting our budgetary and schedule constraints. This dissemination platform will not only provide for faster, more complete and accurate data services, but will also serve as a mechanism to capture user feedback across our surveys and programs to achieve more complete user experience, and will enhance Census data products and tools that empower others to put our data to work for them.