Semantic Data Portal and Data Catalog

⛰ The Challenge

As part of their efforts to improve overall data quality, data usage, and coordination, the Chief Data Officer of a federal agency was seeking to analyze current data management practices and identify ways to improve the Office's existing processes. One of the most pressing challenges identified was that data scientists and economists were finding it difficult to make efficient use of siloed data sources in order to easily access, interpret, and track data and its history. Each researcher had anecdotal knowledge of what data resources were available, and they were often recreating similar data manipulations and research that other analysts had done previously in their own departments. The agency needed a more efficient and centralized way to capture key contextual information to drive the use, discovery, and reuse of data. The solution had to enhance and modernize their metadata management practices through improved access and visibility across agency data resources, while maintaining the appropriate security measures.

◉ EXPLORE KEY INFO FROM DIFFERENT PERSPECTIVES
- Ability to explore available data according to different information needs.
- Allow for a dynamic search experience by traversing connections in contextual information between data

🔒 RESPECT USER AND DATA SECURITY ROLES
- Track security requirements for the underlying data for both the data elements and metadata itself.
- Only reveal information based on user access permissions.

🕐 TRACK DATES, REVISION HISTORY, AND OTHER CHANGES
- Reveal how data and its contextual information changes over time
- Feature key deadlines and important changes so situate data in its appropriate temporal context

⚛ UNIFY ALL RELEVANT CONTEXT ABOUT DATA INTO A SINGLE RESOURCE
- Relate information on banks, regulatory forms and statutes, data sets, and individual data elements.
- Gather a 360 degree view of what the data means, and why it is valuable.

🗄 CONNECT INFORMATION WITH ACTION
- Provide key information on how to access the data so users can immediately start using what they have discovered.
- Provide sample queries, database connection information, and data access info to remove obstacles to utilize data.

💡 The Solution

Enterprise Knowledge (EK) started this effort with a strategy engagement to define an overarching strategy, identify business requirements with prioritized use cases, and design a roadmap to guide the scale of the overall effort. As an initial step towards implementing that strategy, EK led the development of an advanced, semantic data catalog prototype, leveraging a knowledge graph to provide key contextual and descriptive information that helped map relationships across the agency's regulatory data sources, including collected data, metadata repositories, and publicly available financial information. For the data catalog, EK also developed an intuitive front-end, user interface that enabled end-users and data researchers to explore and access the data within the model. To support the catalog application, the knowledge graph was custom modeled to integrate information from a variety of data sources, such as structured databases, regulatory manuals, existing metadata repositories, and public websites. The integrated semantic layer contained the relationships that a user could leverage to explore and traverse information on relevant datasets, regulatory language, financial institutions, and data elements, so that they could discover what they needed regardless of their starting point.
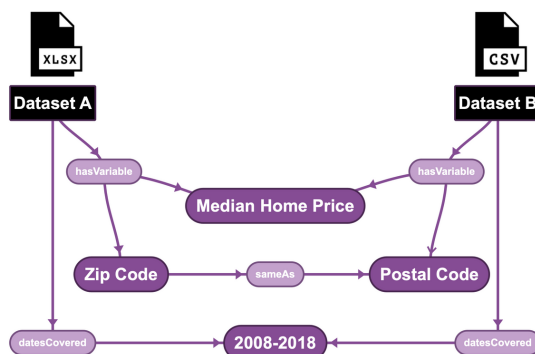
SEMANTIC DATA PORTAL AND DATA CATALOG

THE EK DIFFERENCE

EK has designed taxonomies, ontologies, and data/metadata models to enable data integrations and modernizations for dozens of large and complex organizations. Relying on this experience, EK was able to rapidly develop the knowledge graph in this solution while providing the agency with a sustainable and scalable model that could readily integrate their data and information across departments, processes, and systems. We were also able to simultaneously ensure the graph's effective governance, standardization, and, most importantly, cross-department usability.
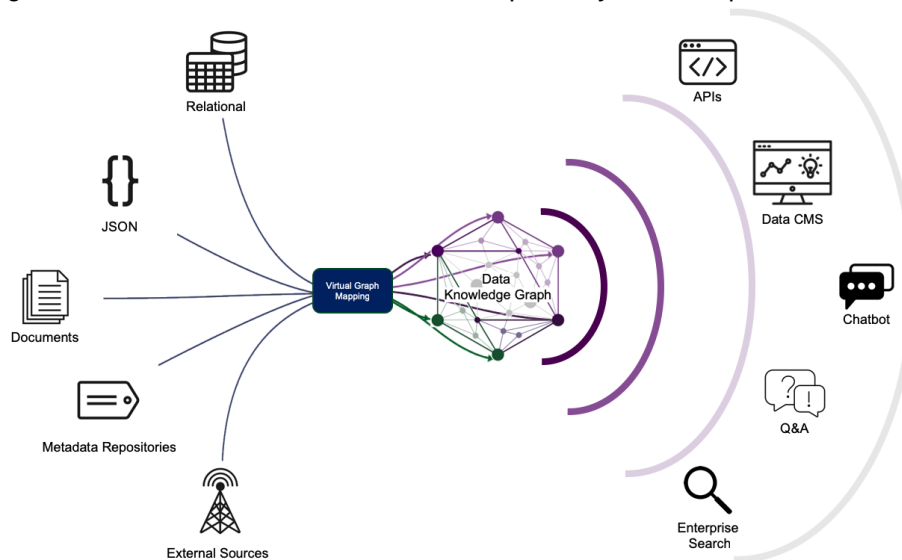
Beyond the model, we also drew upon our end-to-end technology solutions capabilities to develop the catalog application using full-stack development methodologies. This development competency allowed our proposed solution to involve not only the remodeling and provision of a new data layer for the agency, but also to develop a demoable prototype to highlight the deep value of the knowledge graph in a very tangible and interactive way.

SEMANTIC DATA PORTAL AND DATA CATALOG

⌁ THE RESULTS

The catalog serves as a visual demonstration of the value of having a semantic data layer to organize, relate, and standardize metadata use at the agency. The catalog also makes it easy to find and connect relevant data with business users to view key information at a glance. Overall, the agency continues to realize the capabilities and associated business outcomes from the phased implementation of the data solution that provides an array of the agency's stakeholders and business users with the ability to:

- Capture information about data and relationships between data to power the findability and usability of data sets across the agency. The intuitive, front-end user interface reduced the amount of time data scientists and other SMEs spent tracking or processing data for non-technical users, as they can now directly access and explore the data for their own decision-making purposes.
- Ensure that only the appropriate people have access to data assets and the information about those data assets.
- Understand the history, context, and processes behind each data set.
- Relate data elements more easily and more consistently. The tool allows data analysts and researchers to access the agency's data resources in a single tool that makes data stored in multiple locations available without having to move or copy the data.
- Enhance the overall quality and efficiency of the agency's data through improved awareness, collaboration, and consistency.

The data portal continues to serve as a scalable model of how the agency can modernize its metadata management practices while making its data easier and more readily available for use.