Using AI to Eliminate Information Silos and Accelerate Government Efficiency

Federal agencies face a significant challenge in overcoming information silos, which can hinder their efficiency, accountability, and effective use of taxpayer dollars. These silos, driven by fragmented and outdated systems, misaligned priorities, and conflicting privacy regulations, limit programs' ability to deliver timely and non-duplicative services. The executive order on eliminating information silos underscores the need to address these challenges and encourages greater momentum toward enhanced intra- and inter-agency coordination and data sharing to reduce waste, fraud, and abuse.

Approaches tailored to federal agencies' priorities that combine **cutting-edge data linkage and AI-driven analysis** can help break information silos. In the private sector, electronic health records are consolidated and analyzed to improve patient care, reduce redundancies, and detect fraud. Similar methods in the public sector can improve interoperability within and between agencies while prioritizing data security and privacy. By employing this approach and working collaboratively, federal agencies can address waste, fraud, and abuse while enhancing the efficiency and effectiveness of service delivery.

Introduction

Artificial intelligence (AI) has rapidly advanced in recent years, offering opportunities to enhance how agencies manage and integrate data. With AI, agencies can more easily map similar concepts in their data sets to a common data model, a task that has traditionally been manual and subjective.

Adopting a common data model would provide the technical foundation for agencies to achieve the executive order's goals efficiently and systematically, helping to enhance collaboration between agencies with varying data capacities and structuring data in a unified format that supports seamless links between agencies and data sets. AI can also support version control, data governance, and access management to help ensure that data assets are responsibly managed across the data life cycle. Yet cross-agency collaboration remains a significant challenge, and many agencies lack the staff capacity or specialized expertise needed to support data standardization and AI integration at scale.

Technological advancements must also be paired with strategies that address redundancies across agencies that collect the same data, such as the U.S. Department of Veterans Affairs, the Centers for Medicare & Medicaid Services, and the Social Security Administration. Sharing eligibility, enrollment, service use, and claims data would enable policymakers and program leaders across agencies to estimate future costs, identify cost-saving strategies, and address gaps or barriers in service delivery. These strategies should also address increased concerns about data privacy and security because, with broader access to sensitive data, there is a larger opportunity for a breach to occur. Each additional access point increases the potential for unauthorized access.

Although the private sector offers solutions to some of these problems, expanding on the right commercial products will help agencies transition from outdated and cumbersome systems to more

sustainable cost-effective solutions. The executive order on Ensuring Commercial, Cost-Effective Solutions in Federal Contracts supports this shift, urging agencies to leverage proven commercial tech. Some agencies already do this, with cloud-based platforms such as Databricks used by the Centers for Medicare & Medicaid Services and Palantir used by the U.S. Department of Defense. But, to eliminate information silos, choosing and installing a product is just the start. Real impact will come from working with experts who understand federal programs and data and can connect the right information in ways that will support agencies' missions and workflows.

The federal government is in a unique position to make these investments in infrastructure for public use in a way that private entities are not incentivized to do. Our proposed solution to address these challenges and connect and integrate key data from different federal and state agencies is to **create a common federated platform using privacy-preserving record linkage and multi-party computing technologies.**

Solution

Solving the problem of siloed data requires involved parties to adhere to accepted methods across several dimensions: data creation and measure or construct development, cross-data linkage practices, data exchange infrastructure, appropriate use covenants, digital security measures, and legal safeguards. Many of these pieces have become accepted standards or are being tested and validated.

The use of AI to create a secure multi-party computing network that allows questions to be answered across different data systems without revealing sensitive data has been tested in healthcare with a clinical research network. Northwestern University has worked with multiple health system sites to ensure that data are mapped using multi-dimensional and multi-modal data (imaging, notes, labs, diagnoses, and other structured health data), feeding the data into different models to create a concept map that can then serve to identify and diagnose conditions. Except for some very complex conditions, the models were just as accurate as humans. Mathematica is uniquely equipped to build on these advancements, combining deep expertise in working with administrative datasets and data governance—which includes privacy-preserving record linkage and secure multi-party computing—to enable protected, collaborative data sharing and actionable insights across networks.

Establishing a federated and connected data network across federal agencies allows that network to be connected to state and commercial networks. Existing health information networks and health information exchanges provide a model for how to structure such partnerships and highlight the various requirements that they must satisfy. The Chicago Area Patient Centered Outcomes Research Network (CAPriCORN) is one of the first such examples. It was established in 2014 as a collaborative research network involving a wide range of actors in Chicagoland healthcare delivery with the objective of promoting patient-centered clinical research. CAPriCORN's member providers serve more than one million patients in the region and connect patients' electronic health record data from health services delivery institutions with researchers, using data use agreements that limit data access to only those uses specified in their agreement. Other networks, such as the eHealth Exchange, Carequality, and the CommonWell Health Alliance, do not yet connect to federal agencies but have demonstrated secure patient linkage and cross-party data exchange technologies that can inform federal solutions.

Ultimately, these networks can connect to certified identification vendors that allow consumers to access the complete set of information regarding their benefits. Data ecosystem vendors, such as Mathematica's partners at Datavant and HealthVerity, are already doing this with data covering a large part of America. Datavant, for example, leverages data connections with more than 70,000 hospitals and clinics, and HealthVerity's data ecosystem includes claims from more than 190 million people under commercial, Medicare Advantage, and Managed Medicaid plans. These data ecosystems play an instrumental role in supporting clinical research using observational data and providing market sizing information to investors and innovators. But, so far, they have only seen limited use in improving federal healthcare spending efficiency.

In addition to enhancing security and data structure consistency, AI applied to unsiloed data is expected to significantly improve data quality. For example, AI models can engage in outlier detection and flag anomalous charges for a specific procedure or prescription against the universe of available data. AI can also identify incongruities between a patient's diagnoses and the procedures performed on them, ensuring that clinicians only provide care justified by evidence-based medical guidelines. Such data review systems will have the ability to raise flags for follow-up manual review by using a human-in-the-loop model.

Partner with trusted data solutions experts.

Our experts are ready to help you identify and adopt the right AI-driven solutions to surface data quality issues and inconsistencies and improve the efficiency of data sharing. Together, we can optimize federal programs and policies for efficiency, cost savings, accountability, and measurable impact.

Contact Ngan MacDonald, Director of Data Innovations, at nmacdonald@mathematica.org.