May 28, 2024

Submitted via https://inquiry.healthit.gov/

Micky Tripathi
National Coordinator
Office of the National Coordinator for Health IT
U.S. Department of Health and Human Services
330 C Street SW
Mary E. Switzer Building
Washington, DC 20201

Dear National Coordinator Tripathi,

Thank you for the opportunity to comment on the *Federal Health IT Strategy, 2024-2030*.

Datavant's mission is to make the world's health data secure, accessible, and usable. Datavant is a data logistics company for healthcare whose products and solutions enable organizations to move and connect data securely through proprietary technology, the world's most robust healthcare network, and value-added services. Datavant enables more than 60 million healthcare records to move between thousands of organizations, more than 70,000 hospitals and clinics, 70% of the 100 largest health systems, and an ecosystem of 500+ real-world data partners.

Our specific comments on your goals and selected objectives and strategies are outlined below. At a high-level, we agree with the goals outlined in the draft strategic plan. We are working to realize these goals every day through our partnerships with providers, payers, research institutions and other stakeholders across the health care system.

Goal 1: Promote health and wellness
Objective: Individuals are empowered to manage their health
Strategies:
- Support individuals in accessing and using their EHI securely, privately and without special effort.
- Protect individuals' right to share their EHI with third parties, including third party applications, of their choice.
- Develop educational resources for choosing and using secure technologies that incorporate privacy protections.

We support the strategic goal of promoting health and wellness, and agree that patient privacy is an essential component of any effort to empower patients to be shared decision makers in their health and healthcare.

Datavant works to reduce the friction of data sharing across the healthcare industry by building neutral, trusted, and ubiquitous technology that protects the privacy of patients while supporting the exchange of identified and de-identified health data across tens of thousands of healthcare institutions. Datavant created the nation's largest health data ecosystem, powering secure data connectivity on behalf of thousands of providers, payers, health data analytics companies, patient-facing applications, government agencies, research institutions and life science companies.

Datavant works closely with tens of thousands of hospital systems and medical practices across the country, supporting their efforts to provide patients with seamless access to their health information, digitizing electronic health records and ensuring interoperability of data, and unlocking the potential of data in medical records. This commitment is exemplified by Datavant's recent initiative to eliminate industry-standard charges for patients requesting their health records, a first-of-its-kind decision aimed at improving patient outcomes and breaking down barriers to accessing personal health information. We fulfill more than 100 million requests for health records annually. In addition, Datavant has just launched a record request automation solution for healthcare organizations, designed to streamline the handling of high volumes of medical record requests from health plans, further enhancing efficiency and compliance in data sharing.

As exemplified by these efforts and others, we encourage efforts to support and prevent the erosion of patient privacy. We believe that individuals should be able to easily and seamlessly access their own health information in the form they choose, and that individuals should also have the right to permit, authorize, or direct access to their health information, particularly for health care purposes. We also understand and appreciate that patients today do not have visibility when their data crosses from being HIPAA-protected to not protected, and that patients may not fully appreciate how their information may be shared, sold or used by third parties.

Today, we have experienced third parties that are not HIPAA-covered entities or business associates use the HIPAA right of access pathway to access patient data, and then subsequently use, monetize, and resell that information however they choose, often without patients having any awareness of the third party's intent or actions. This is now happening at increasing volumes as an ever-expanding group of third party entities misuse data. These medical records often contain sensitive information, including information about reproductive health.

We agree that any strategy to empower patients should have education at its core. We also believe that there are steps that HHS can and should take to ensure that third parties do not abuse the pathways they are given to support patients in making key health care decisions.

Goal 2: Enhance the delivery and experience of care
Objectives:
- *Providers deliver safe, equitable, high quality and improved care.*
- *Providers experience reduced regulatory and administrative burden*

Strategy: Advance standardization and interoperability of social determinants of health data.

Datavant strongly supports efforts to encourage the use of social determinants of health data ("SDOH") to enhance quality improvement activities, to track factors that influence people's health, and to eliminate health inequities. Our experience is that social determinants of health data can be extremely valuable and powerful tools to improve patient outcomes while reducing costs. and to improve healthcare.

For example, Datavant's Social Determinant Insights product, like other similar products in the industry, helps payers to supplement their members' health data with over 400 socioeconomic attributes, ranging from financial security to transportation availability and demographic information, including phone number, email address, and a deceased indicator. Datavant's Social Determinants Insights tool can also layer Z codes and other social determinant keywords and references from patients' health data to create a more comprehensive, accurate, and timely data set than Z codes would provide.

Although we agree that there is value in providers engaging with patients to better understand their holistic needs, we caution that relying on data collected solely through screenings is insufficient to fully inform policymakers and health care practitioners as part of efforts to enhance quality improvement, track factors, or identify and monitor disparities, given the very low current rates of collection and reporting of social needs data at the point of care.

There are a number of reasons why there may be less routine documentation and reporting of SDOH in the inpatient setting. We offer a few examples that we have witnessed, which largely align with the reasons CMS has listed, below.

- Screening for and collection of SDOH data is currently voluntary, and is generally not incented through financial or performance mechanisms.

- Social needs screening has not traditionally been a part of providers' workflows, and may require additional time to perform the screening and interpret and address any needs identified, increasing the burden on already overburdened providers.
- Providers are often hesitant to screen for social needs, due to a lack of wide-scale education and training on social determinants, lack of information on the availability of screening tools, and insufficient training for screening for social needs in a socially and culturally appropriate way.
- Patients may feel uncomfortable sharing their social needs information with their provider, resulting in inaccurate data.
- Patients may not see the relevance of providing information to their providers about their social needs and may simply not provide responses to those questions.
- Studies have shown that many providers are wary of screening for social needs if they feel they do not also have the ability to make referrals or to connect patients to resources to address their needs.

Strategy: Promote the safe and responsible use of AI tools.

AI tools have many applications in the health data sector. First, AI can help enhance research powered by health data by automating and improving processes to make health data more accessible and useful. Second, AI can support privacy preservation in health research (e.g., by powering synthetic healthcare datasets), or by enabling federated approaches. Third, AI can help better interpret health data to generate insights at both the personalized medicine and the large-scale predictive analytics levels. Finally, to advance AI models across the healthcare value chain, health data, including novel sources such as imaging and genomic data, can be utilized to help train, fine-tune, and validate these models.

The healthcare data industry should ensure that the intersection of healthcare data and AI is powered by robust data to avoid bias and discrimination, and that AI efforts are built on a responsible, ethical, and privacy-preserving foundation.

Our current privacy research and development initiatives are investigating methods to leverage AI to enhance privacy assessments and data handling within the health sector. First, we are leveraging AI to evaluate unstructured health data types, such as imaging and genomic data. Second, we are developing privacy metrics with quantifiable thresholds and fostering industry-wide dialogue to achieve consensus on privacy standards and risks associated with AI and related technologies. This consensus is critical in guiding public and private sector contributions to privacy practices.

With respect to managing the performance of AI in the health data sector, we offer the following thoughts:

- **Clear Objectives:** "AI" is a broad term, so agencies should clearly specify what they want to procure and why. It is easy to "run an algorithm/model," but this does not mean that the output will be what is needed, will be reliable, etc. At the base level, agencies should distinguish between "predictive AI" (forecasting based on historical data, more established approaches) and "generative AI" (largely new content creation that resembles existing data).
- **Data Completeness:** Missing data leads to flawed science, resulting in biased inferences and predictive models. We urge establishing benchmarks to help understand and compare the data underlying AI models. However, this is a nuanced topic as the same data may be more or less appropriate depending on the context.. For example, if one was creating a model based on sickle cell patients, it is entirely appropriate for it to represent the population of patients most impacted by sickle cell. It may not be appropriate to do so in other contexts.
- **Data Integrity:** Data access and quality are immensely important. Any government led initiatives where data is supplied must provide data of high quality and whose profile and provenance is understood.
- **Privacy Safeguards:** To protect the privacy of highly sensitive personal data, federal initiatives should employ strong protective measures such as the use of enterprise versions of AI tools (preventing personal data from being used for machine learning or distributive purposes outside the contracting entity), strong encryption at rest and in transit, limits on retention and use of personal data, de-identification, and other privacy preserving technologies.
- **Performance Quality Measurement:** Federal agencies should establish performance benchmarks to ensure appropriate decision making. For example, data should be accurately labeled. There is also a need for continual maintenance and monitoring of models, outputs and data to avoid drift, overtraining, AI hallucination incidents, etc.
- **Validation Requirements to Avoid Automating Bias:** There must be careful consideration to remove bias in AI models where output is used for decision making. To promote equity, additional consideration should be given to overtraining, drift, data quality, and ensuring representative data is used.
- **Sandbox Testing:** Federal agencies should consider whether and when it would be useful to require a sandbox environment for prototyping prior to deployment of AI within and across data sets.
- **Deriving Lessons from User Testing:** To ensure appropriate privacy controls, federal agencies should include some commentary or direction on privacy thresholds (including test types, number of tests, and thresholds).

Federal agencies may also want to direct their partners to use small cohorts and privacy preserving technologies to improve data quality in a privacy preserving fashion, and to take steps to relieve administrative burden in data creation. Federal agencies should work with leaders in the security industry to develop clear guidance to manage the security risks and

limitations of individual AI tools where health data will be processed. At a minimum, this should include guidance to understand and manage hallucination risk, which could impact patient safety, as well as mitigate prompt injection or attempts to bypass controls designed to protect patient privacy.

Goal 3: Accelerate research and innovation
Objective: Individual and population-level research and analysis are enhanced by health IT.

Strategy: Protect de-identified health information from re-identification

We strongly support widespread federal adoption of privacy-preserving record linkage ("PPRL") approaches as a mechanism to enable research uses of de-identified health:

- **Preservation of Patient Privacy:** PPRL enables data linkage without revealing sensitive attributes, mitigating unauthorized access and potential breaches. Employing advanced de-identification techniques, PPRL irreversibly transforms data, preventing re-identification. This balance between linkage and privacy empowers responsible data analysis while upholding ethical standards.

- **Accuracy & Data Quality:** PPRL facilitates accurate and reliable data linkage by enabling deduplication without exposing patient identifiers. The ability to use fine-grain features and attributes, combined with modern solutions that use machine learning models, result in accurate disease prevalence, particularly in care settings and conditions that have a high degree of care fragmentation.

- **Cross-Domain Insights:** PPRL empowers researchers and policymakers to discover insights from diverse datasets without compromising privacy, thus fostering collaborations and accelerating advancements in healthcare research.

- **HIPAA-Compliant Framework:** PPRL implementations within the health sector must operate within a HIPAA-compliant framework, specifically meeting the Expert Determination Standard of the HIPAA Privacy Rule §164.514(b)(1). This standard requires that an expert performs a statistical assessment of the PPRL tokens to confirm that it poses a very small risk that it can be used alone or in combination to identify the individual.

- **Data Reuse in Accordance with FAIR Principles:** Enabling PPRL within databases and repositories, enables data linkage across data repositories and enclaves to be re-used without needing to take a wholly centralized data approach. This data reuse

in accordance with FAIR principles, ensures that disparate datasets can still be **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable (FAIR). This data minimization approach also ensures only minimum necessary linked data would need to be pooled and aggregated to formulate a relevant dataset.

- **Robust Privacy Risk Disclosure Assessments for Responsible De-identification:** In bringing together de-identified datasets, more sophisticated techniques and methodologies to ensure datasets remain de-identified are required. The HIPAA Privacy Rule at §164.514(b)(1) uses the Expert Determination standard, which provides a more fit-for-purpose assessment of de-identified datasets regardless of patient consent. This approach provides a data governance methodology that addresses a higher-level of assurance when de-identified datasets are constructed.

PPRL empowers use cases that extend across various domains within healthcare. It transforms how researchers and policymakers derive insights while respecting individual privacy. Use cases that may be of interest to the federal government include:

- **Public Health Analysis:** PPRL can be utilized to link various health databases to monitor disease prevalence, assess treatment outcomes, and evaluate the efficacy of public health interventions. For instance, researchers can study the impact of global pandemics to shape health policies while maintaining patient anonymity.

- **Clinical Trials:** By linking electronic health records to trial eligibility criteria, records match precisely without compromising privacy. This accelerates patient recruitment, preserves data confidentiality, and optimizes resource use, enhancing both research efficiency and ethical considerations.

- **Healthcare Policy Evaluation:** Government agencies can leverage privacy-preserving record linkage to assess the effectiveness of healthcare policies by linking datasets from hospitals, insurance providers, and public health agencies. This enables comprehensive analysis without violating patient confidentiality.

Strategy: Enable data exchange to support retrieval of patient records from a broad array of entities in a responsible manner

In support of data reuse and lowering patient and study participant burden, we encourage the following:

- **Reuse of Data from a Broad Array of Entities:** Participants should be able to direct data releases of their medical data from health care provider organizations,

registries, and administrative claims from payors and other claims data providers. This ability to direct data release can take the form of direct consent for *specific studies* or for patient consented research uses of de-identified data which is typically contained within health care provider patient consents.

- **Reuse of Data in Broad Array of Formats and Data Models:** Participants who wish to permit data to be released for specific studies (e.g. registry participation, observational data studies, clinical trials) should have the ability to request data releases without being subject solely to TEFCA standards and requirements. For example, when readily available data has been transformed to data models and formats that are optimized for research-readiness such as OMOP, entities that are able to supply data in the research-ready formats should have the option to do so. This is not intended to be counter to TEFCA as a way to accelerate data-sharing for research, but rather preserving the ability to enable data exchanges in formats optimized for research readiness.

Strategy: Promote increased transparency into the development and use of AI algorithms in health care settings.

Decision support interventions (DSIs) are often built using patient training data. Previous studies have identified biases, such as when a DSI is trained using one set of data, but the DSI is applied to a different population of patients. Providing transparency around the data that was used in developing or training these models can improve the quality of the DSI, and allow users to evaluate for themselves whether the algorithms can safely be applied to a target population.

Predictive DSI technology can change rapidly, and simple updates or the addition of new features can change the nature of the underlying models and decision support. This review should include updated testing on more recent data to see if there has been a shift or change in the data that the predictive DSI is being used on, compared to the initial data used to train the model. The results of that testing should be made available as part of the source attribution and the annual review.

Goal 4: Connect the health system with health data
Objective: Health IT users have clear and shared expectations for data sharing.
Strategies:
- Promote information sharing practices.
- Advance TEFCA that creates a universal governance, policy and technical floor for nationwide interoperability; enables individuals to access their EHI; and simplifies connectivity for organizations to securely exchange information.
- Improve interoperable exchange among different health systems, devices, and applications, and maintain the ability to exchange and use health information seamlessly.

Datavant applauds the HHS Office of the National Coordinator for Health IT (ONC) and the RCE for actions taken to date to implement the Trusted Exchange Framework and Common Agreement ("TEFCA") as a voluntary health information exchange network. Given that data fragmentation is the largest challenge facing the health data industry, Datavant supports robust health information exchange – through TEFCA as well as other information exchange channels – to reduce fragmentation, promote interoperability, and improve access to and exchange of health information.

First, Datavant strongly supports efforts to ensure that people have access to their own healthcare data in a private, secure manner. We applaud ONC and the RCE for prioritizing Individual Access Services (IAS), and agree that this use case is a top priority for health information exchange. That is why we are focused on building an open data ecosystem that allows healthcare stakeholders to readily exchange data while protecting patient privacy.

Like the Department of Health and Human Services (HHS), we believe that protecting patient privacy is paramount when using health data to improve health and healthcare. We are supportive of additional privacy protections built into version 2 of the TEFCA Common Agreement and QHIN Technical Framework (QTF), and urge ONC and the RCE to consider additional steps that can be taken to safeguard patient data flowing through TEFCA and to ensure that patients know and understand how their data may be used by TEFCA QHINs, Participants and Subparticipants.

Furthermore, there continues to be pathways for third parties to access, misuse, and profit from healthcare data while asserting compliance with the HIPAA framework. It is imperative to implement additional measures to prevent the inappropriate use of patients' data by any entities associated with TEFCA, thus safeguarding patients' intentions and privacy.

Finally, the 21st Century Cures Act specifically envisioned TEFCA as being a voluntary structure, and one that would avoid the disruption of existing exchanges between participants of health information networks. TEFCA participation must continue to be voluntary, so that it does not impede innovation or result in duplicative work that wastes taxpayer dollars.

As TEFCA implementation moves forward, we are exploring leadership roles with the RCE so that we can share our insights and experience more robustly. We look forward to serving as a resource to ONC and the RCE, and to continuing to collaborate on TEFCA implementation.

*****

Thank you for the opportunity to provide comments. Please do not hesitate to reach out to me at ssegall@datavant.com if we can be a resource to you.

Sincerely,

Samantha Segall

Samantha Segall
VP, Head of Government & Public Affairs