The Future Challenge for Federal Chief Data Officers

By Paul Eder, PhD and Eiman Behzadi

Dr. Eder is a Director at The Center for Organizational Excellence. Mr. Behzadi is a Director at Zenledger.

As Government leadership and processes mature, the ways in which agencies use data for decision-making and modeling also mature. Over the last decade, the level of advancement and attention given to using quality data in ethical and valid ways has grown exponentially. This maturity is partially the result of laws that have promoted the availability of data and required its use in strategic planning and decision-making processes.

Title II of [The Foundations for Evidence-Based Policymaking Act of 2018](https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf) (Evidence Act) requires the establishment of the Chief Data Officer (CDO) role in every Federal agency. Among other duties, the CDO is responsible for the management of data assets in an agency. Accordingly, the CDO must maximize the use of agency data by encouraging innovative uses and maintaining a comprehensive data catalog with metadata suitable for user consumption.

An immediate by-product of more usable data is the advent of advanced data models and data assets enhanced by emerging technology. Although not specifically detailed in the Evidence Act, we propose that CDO responsibilities are subsequently expanded based on the maturity of data use in an organization. A CDO must plan for the cataloging and maintenance of data, not only in its current form but in the projected ways that it can be used by Chief Evaluation Officers, Performance Improvement Officers, and government statisticians.

We view the following emerging technological trends as posing challenges for the traditional ways that data are catalogued:

1. Increasing use of blockchain data and cryptocurrency trends
2. Data lake assets
3. Artificial intelligence
4. Machine learning

The government is used to very cleanly defined data sets; they have rows and columns and easy-to-specify meta-data for each element. It is a relatively straight-forward (but nonetheless complicated) task to develop governance processes around these types of data. Not surprisingly, data governance is viewed as an important role by [100% of CDOs](https://www.datafoundation.org/cdo-insights-report-2021).

As CDOs establish their footing, it is natural to focus first on the data assets that are readily available and easy to define. The maturity of governance is a welcome response to the Evidence Act. But we encourage government agencies to think beyond the data-in-hand or traditional datasets when determining governance procedures.

Emerging technological trends are not going away and will soon be a key component of the data practices in most agencies. Accordingly, CDOs must plan now to incorporate these trends into their governance practices. A [study by the Data Foundation](https://www.datafoundation.org/effective-data-governance-a-survey-of-federal-chief-data-officers-2020) suggested that increasing data literacy in government was a major priority for CDOs. As agencies mature, data literacy becomes bigger than understanding what data assets exist. Leaders must also know how data is evolving, how new trends impact the storage of data, and how new data can be generated from the models structured based on existing assets. Below, we outline three ways that CDOs and agency leaders can ensure they are strategically considering emerging trends when establishing their governance procedures.

1. Document governance for blockchain based assets. Many government agencies do not maintain their own blockchains. Others only analyze blockchains created by other entities. Nonetheless, as this data emerges and becomes a go-to method for decentralized storage, agencies must be prepared for how to handle these assets. By getting in front of the curve, agencies can ensure when these assets are created, implemented, and analyzed that they are backed by sound governance principles. For example, the Internal Revenue Service (IRS) must determine the best way for analyzing blockchain data associated with cryptocurrency and the best way for cataloging that data as an available asset.
2. Develop meta-data practices for data lake assets. [As multiple different data types are comingled and interconnected via various linkages](https://governmenttechnologyinsider.com/the-dod-and-u-s-census-bureau-turn-to-data-lakes-for-modernization-and-greater-insights/), agencies should understand how these data were chosen, why they were combined, and how they are maintained. Agencies must plan for the evolving use of data lakes and ensure they are developed in a way consistent with evaluation needs and validation capabilities.
3. Develop a repository of machine learning and artificial intelligence model meta-data that details how data is used for description, prediction, and prescription. [Studies from McKinsey](https://www.mckinsey.com/industries/public-and-social-sector/our-insights/when-governments-turn-to-ai-algorithms-trade-offs-and-trust) have documented the importance of ensuring models are accurate, explainable, stable, and adopted as trustworthy sources. Trust in these types of models has its foundation in the effective governance of the data supports them and the effective management of the models themselves, for ensuring data feeds are updated, learning models are continuously retrained as data is updated, and results of business intelligence procedures based on models are evaluated for accuracy and opportunities for improvement.

Of course, the future cannot be fully predicted, but the trends are clear. The landscape of data is evolving. CDOs must be at the forefront of that evolution, guiding their agencies to establish effective governance procedures before they are backlogged with numerous complex datasets and models that they are not prepared to monitor and maintain. Data governance is not a one-time process; it must focus on present needs but also look forward and serve as a pathway for agencies to ensure the future accuracy and stability of their data assets.