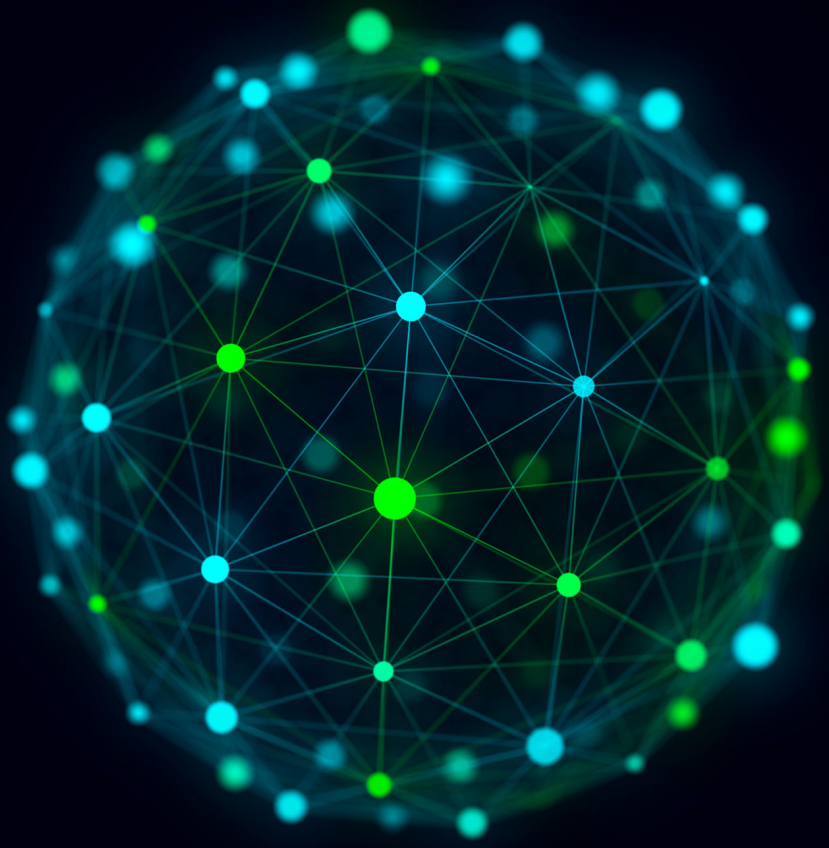
Deloitte.



Is your data AI-ready?

As a CDO preparing for the latest with AI, what's next for your leadership and organization? Begin with assessing where your data is and where it needs to go for AI

Authors: Adita Karkera, Mike Greene, Landon Henderson, Jonathan Newman

MAKING AN IMPACT THAT MATTERS since 1845

The Path to AI ready: CDO Considerations and a Rubric for Data Readiness

In the rapidly advancing era of AI, data leaders face a central challenge as they consider the next stage of AI adoption: the immensity of data being collected. How does this leader harness data and AI for citizen benefit? It starts with assessing data readiness. To meet this future with success, there are several data readiness and governance initiatives that are necessary to prepare agencies and businesses alike to adopt AI.

AI is Changing Mission Operations

Artificial Intelligence (AI) has only just begun, quickly changing the way we work and perceive the world. As of now, business adoption of AI is only at operational lift-off: 7% of potential but not current AI-users are "looking into it", while 14% have tested a few concepts with limited success and 25% have processes that are fully AI-enabled.[1] Artificial Intelligence corporate benefits include better and quicker insights (machine learning can sift through data faster than humans to find patterns and generate helpful predictions), efficiency (AI automates repetitive tasks that take humans a long time), improved accuracy (programs don't make mistakes like humans do - assuming they are programmed correctly and monitored) and cost reduction (humans need rest, but AI doesn't sleep!).

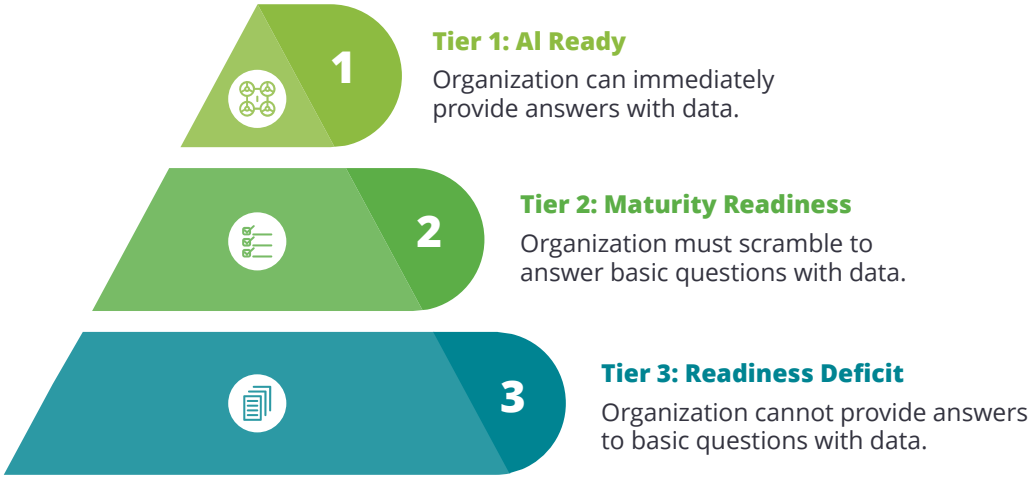
Data Preparation & Assessment

AI flourishes with clean data. Thus, CDOs must find a way to acquire this clean data and store it (in vast quantities) so that it is highly accessible and reusable for their trajectory to AI adoption. The Federal Data Strategy states that the two most challenging preparatory requirements for government agencies' ability to acquire and store large amounts of data are "updating data inventories" and "identifying data for open data plans".[2] Only 35% of CDOs feel they have "a lot" or "completely" enough resources to fulfill these requirements. This suggests there are many executive officers, within federal agencies, who are under-resourced for success as acquiring relevant clean data becomes critical. To facilitate data readiness, CDOs must ask: where does their organization stand now – in **data findability, accessibility, interoperability, and reusability** ("FAIR")? [3] Then, gaps can be filled to get the data AI ready.

Leveling Up Step by Step

CDOs would be wise to follow a data preparation rubric to "level up" towards data readiness and AI adoption. (For a play by play detail on how to overhaul your data strategy, reference the Deloitte CDO Playbook, which delves into defensive and offensive data strategies beyond the scope of this article.) In terms of leveling up step by step, an example of a rubric and the progression from Tier 3 (Readiness Deficit) to Tier 1 (AI Ready) is illustrated in the below diagram.

Data Preparation Rubric to Level Up towards Data Readiness & AI Adoption



Organizations can use the steps below to help them "level-up" toward being AI-ready, from whatever initial condition they reside.

CDO Engagement in the Progression to AI ready

Progressing towards Data Readiness and AI Adoption	How a CDO can help the organization move forward
Step 1. Build a key stakeholder network	
Successful AI starts with connecting key stakeholders to the data products your organization is deploying. For instance, take a form that is part of applying for a benefit, license, or other government service. Policy makers will decide on the components of this form. Data engineers will decide on how best to collect and transform the data collected from this form. Data scientists analyze the data and apply AI. Constituents are the users of the form. All of these stakeholders have influence in the development of the data final product and must work together. Increased communication and tracking across all those who place value in data will make the data more **findable** by all stakeholders - scientists, researchers, data scientists, constituents, etc. Data location, provenance, and citation must be well documented.	The CDO is the bridge between the stakeholders who use the data to make informed decisions. By fostering connections amongst stakeholders and balancing the needs of each, the CDO can facilitate a community of engaged individuals, helping the organization achieve their mission. The CDO can represent the needs and data interests of each group while shepherding technological advancement toward usable, safe and sustainable solutions. Not only findability but also the accessibility of the data must be honored for the broader public to gain benefits from information. It is important to keep in mind if the data will be **accessible** - how will the broader public and researcher stakeholders come to use and access this data?

Step 2. Develop a robust data transformation strategy

Data is fuel for AI, and that fuel needs to be refined for optimal performance. To use data in AI solutions, many steps may be required, from documenting meta-data in consultation with subject matter experts to more technical data cleaning and formatting such as processing text with regex or geocoding location information. Data transformation must also be contextualized with awareness of data security and privacy, implementing security protocols as necessary and as required by law (HIPPA, e.g.).

An organization that wants to undertake AI will develop a strategy for these data transformations, enabling the value of AI solutions. The organization will bear in mind: is the data **interoperable**, or able to be used across multiple platforms?

A CDO can craft the data transformation strategy by understanding the transformations needed, bringing together the community to document the data, and providing tools to empower a community of data consumers to transform the data for eventual AI solutions. Will the data be **interoperable** across many communities and software? How does the CDO include encryption, user access controls, machine learning algorithms to identify security vulnerabilities, and anonymization techniques into their organization's data transformation strategy, for example?

Step 3. Enable data security, privacy, & trust

Privacy and confidentially need to be intrinsic to data architecture for robust security. Lack of privacy and confidentiality can sap trust from stakeholders. Strengthening trust requires: (1) maintaining accurate data and (2) safeguarding the privacy of the users.

For example, when managing Personal Health Information (PHI), Health Insurance Portability and Accountability Act (HIPAA) privacy rules safeguard against misuse of sensitive medical records and other personally identifiable information. Carefully following these requirements is necessary to building and maintaining trust. Also, it is good to bear in mind security, privacy and trust while determining approaches to making the organization's data **reusable**. I.e., how will the data be reused longitudinally or destroyed right away after use?

CDOs can promote security, privacy and trust by implementing data protections such as regulating user access, which restricts entry and egress from other systems. Another method could be to implement cloud-based machine learning algorithms that proactively search for algorithmic weaknesses in security protections.

Data transformation can safeguard data privacy. To do so, CDOs can promote privacy practices such as the generation of synthetic data for use in AI solutions. Such data is "generated to emulate certain key information found in the actual data and provide the ability to draw valid statistical inferences" [4] while improving privacy and trust by serving as a substitute for the original data.

As we know, AI solutions are only as useful as the good, copious data by which they are trained and built. The CDO is in a unique position to be able to provide this data - **helping organizations take honest assessments of their current data readiness and AI adoption and assessing future AI and data possibilities**. The CDO's role will evolve even more in the coming years alongside advancements in technologies, new regulation, and changes to data use patterns.

By developing important, reciprocal relationships between key stakeholders, implementing powerful data transformation strategies, and protecting data privacy through data security, the CDO will help build a stronger foundation for AI readiness and achieve the vast promise of AI in government – from improving professional and humanitarian services to reaching mission goals more effectively and augmenting the public service workforce.

Authors

Adita Karkera, GPS Chief Data Officer
Chief Data Officer, Deloitte Government & Public Services

Adita Karkera serves as the Chief Data Officer for Deloitte Consulting LLP's government and public services practice and is a fellow at the Deloitte AI Institute for Government. With more than 23 years of industry experience, Adita Karkera is an award winning executive, dedicated to improving public service. She serves on numerous industry boards and data management industry forums, including the CDO Magazine Editorial Board. She is a pioneer in articulating the importance of data literacy especially in accelerating advanced analytics and trustworthy AI adoption in government.

LinkedIn: www.linkedin.com/in/aditakarkera/ **Twitter:** www.twitter.com/AKark_DataGirl

Mike Greene
Technology Fellow, Artificial Intelligence & Data Engineering

Mike is a data scientist passionate about helping our clients use data and advanced analytical solutions to crack the biggest problems. For over 15 years he has designed, built, and implemented statistical, behavioral, and machine learning solutions for public and private sector organizations to improve outcomes. Mike holds an AB in Mathematics from the University of Chicago, and an AM in Statistics from Harvard University. He has been a member of the American Statistical Association since 2003 and proudly knows more computer languages than spoken languages.

LinkedIn: www.linkedin.com/in/michael-greene-a026221/

Landon Henderson,
Senior Consultant
Deloitte Consulting,
Government and Public Services

Jonathan Newman,
Specialist Senior
Deloitte Consulting,
Government and Public Services

Endnotes

1. Eliaçık, E. (2023, January 5). Early bird benefits in AI adoption are about to end. DataConomy.
2. Hart, N., Jones, T., Lawton, J., Sheldon, L. & Joe Willey. (2021, September) CDO Insights: 2021 Survey Results On the Maturation of Data Governance in U.S. Federal Agencies. https://www.datafoundation.org/cdo-insights-report-2021
3. Wilkinson M., Dumontier, M., Aalbersberg, J. et al. (15 March 2016). "The FAIR Guiding Principles for scientific data management and stewardship". Scientific Data. 3 (1): 160018. doi:10.1038/SDATA.2016.18
4. Raghunathan, Trivellore E. "Synthetic Data". Annual Review of Statistics and Its Application 2021 8:1, 129-140.

Acknowledgements

The authors thank Ashley Hall, Harlan Simpson, and Jolito Rivera for their assistance drafting and editing. The authors also thank the numerous reviewers who provided invaluable feedback on earlier drafts of this paper.

Deloitte.

About Deloitte

As used in this document, "Deloitte" means Deloitte Consulting LLP, a subsidiary of Deloitte LLP. Please see www.deloitte.com/us/about for a detailed description of our legal structure. Certain services may not be available to attest clients under the rules and regulations of public accounting.

Designed by CoRe Cretive Services. RITM1554299