

A Data Governance Roadmap Even a Robot Would Love

George T. Tziahanas VP of Compliance

© ARCHIVE360 2023 ALL RIGHTS RESERVED

Agenda

- Past is Prologue
- The Future of Data
- A Common Approach
- AI/ML and Robots
- Integrated Data Governance



Chips and Rice

Since 2003, the world economy is producing more transistors (building block of chips) at lower cost than grains of rice



Wavelength of Visible Light Now Too Big for Chips



1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1991 1998 1999 2009 2001 2002 2004 2005 2004 2005 2006 2011 2012 2013 2014 2015 2016 2011 2018 2019 2020 2021 2022 2021 2022

Chips are now fabricated with 3 nanometer transistors, well beyond visible wavelengths of light (into extreme ultraviolet)

The Future of Data



CAGR for Structured Data is 8.5%, Unstructured is 15.2%



The Future of Data and AI

McKinsey estimated impact of Al across core functions

- Customer (or govt. constituent) operations greatest impact
- Software engineering will accelerate creation and consumption of data
- Legal and Compliance high impact as % of spend

Using generative AI in just a few functions could drive most of the technology's impact across potential corporate use cases.



M-19-21/ M-23-07

Transition to Electronic Records

Purpose:

Improve public access Reduce costs Extract value from data

How?

Transfer permanent records & metadata in electronic format







Federal Government Direction

Federal Data Strategy calls for the federal government to replicate, accelerate, and scale leading practices related to government data (M-19-21/M-23-07)

- Provide consistent, reliable, and privacy-preserving access to federal government data to the public, businesses, and researchers for commercial and other public purposes
- Permanent federal records and appropriate metadata must be created, managed, and maintained in electronic format

Balanced objectives for deriving insight and value from data, while managing risk and adhering to obligations



It is All Just Data: Obligations and Decision-Making

- Risks and affirmative obligations traditional drivers for data governance:
 - Volume, velocity, mobility, and sensitivity of information
 - Volume and complexity of legal, security and privacy requirements
 - Hybrid technology and operating environments
 - Headline and business interruption risks
- Al will allow <u>improved analysis and decision-</u> <u>making</u> in modern governance strategies



Balancing Value and Risk









Modern Example of Real-Time Analysis and Decision-Making (Risk/Reward)

For example, many insurance companies now offer potential savings in automotive insurance if allowed to monitor driving habits in real-time.

- These applications capture tremendous amounts of information, from duration, distance, acceleration, speed and other attributes for a given individual.
- This allows the companies to create models of risk and alter coverage rates based on this analysis but create vast amounts of new data.

Insurance companies also develop insurability scores and models, based on extraordinary aggregation of publicly and privately available data.

• The aggregation of this data comprises some of the most expansive views of an individual's habits, practices and personal information.



National Interest in Data Sovereignty-US CUI Requirements

Country	Relevant Law
China	Data Security Law of People's Republic of China, severely limits data that can be shared outside of country
France	Proposed data security law effectively precluding the use of cloud- infrastructure not owned, controlled, or managed by France-based resources for ~600 "vital" or "essential" companies
KSA and Emirate of Dubai	Both have classification regimes as a predicate for data that can be moved to cloud-environments, and initiatives to fund major cloud-vendors in country
Germany	Sovereign Cloud Initiative, run in country by T-systems (subsidiary of Deutsche Telekom) for mandated classes of information
U.S. Controlled Unclassified Information (CUI)	Establishes classes of government and contractor information subject to specific protections and sovereignty requirements



Key Elements for Roadmap on Controlled Unclassified Information (CUI)

Establish

Establish clear understanding of critical data and workloads

Understand Understand where relevant data is created, disseminated, and stored geographically (including third-parties)

Leverage In-country cloud-infrastructure if possible, to avoid setting up internally managed data centers

Classify Look at solutions that assist in classifying data, either with embedded AI/ML, or ability integrate AI/ML solutions



ML and Al

The algorithms behind many ML and AI tools are not new

- Linear regression, support vector machines (SVM), Bayesian inference, and Euclidian distance models have been around for decades or more
- Using these algorithms at scale required advances in compute resources recently available

Legal and Compliance solutions started using various ML models ~10-years ago

- Predictive Coding, Technology Assisted Review (1.0/2.0/etc.), Classification and Surveillance; available in most platforms today
- Use has been inconsistent but value and efficacy increased significantly in recent years

Advanced analytics (AI/ML) allows organizations to process and derive insight from enormous sets of data

• This creates both opportunities and risk



AI/ML Example Use Cases for Your Roadmap

Use Case	AI/ML Used
Content Classification	Combination of supervised and unsupervised learning. Will become critical throughout data governance models
FOIA Content Identification	Primarily supervised learning, with some unsupervised learning for clustering/categorization
Contract or Document Construction	Generative AI, creating contracts or documents based on an LLM. May be industry or sector/agency specific models
Service Interaction	Generative AI, <u>providing self-service answers</u> to questions/queries, without having to search through results



Additional Considerations AI/ML

Models	Considerations
Your data, your models	Provides highest levels of control, and also higher levels of precision based on your data, operations, business, etc. Downside is lack of leverage what others have created
Other's model, your data	Using third-party models, but applied against your data. May require additional levels of training, but leverages learning from others while limiting access to your data
Sharing your data to create models	Everybody benefits from improved models, but key is what data is actually shared and stored elsewhere. Non-generative Al lower risk than generative Al use cases
Sharing your queries to create content/models	Questions/queries are shared, which could include things like source code, copyrighted material, etc. High risk, unclear reward



Integrated Governance

Many organizations mistakenly treat their obligations and risks separately

However, a given piece of information may have compliance, security, and privacy related requirements

Integrating common policies and controls across sets of information minimizes redundancy across technology and security







Managing the Environment Customer Service EO 14028 **NARA/Capstone** M-19-21/M-23-07 **Cloud Computing On-Premises** Automating Information **Better & More Efficiently** (Microsoft 365 / Google Workspace) **Record Storage Governance & Compliance Informed General Public** Reaching M365 Regulatory / internal Migration to M365 Reporting & or Cloud Archive compliance storage limits Analytics Litigation / internal Paper record Retiring legacy digitization investigations systems



Conclusions

- Understand what you have
 - Data classification as a predicate
 - Move critical workloads and systems to modern architectures
- Have a common framework that applies compliance, legal, security, and privacy related obligations
 - Leverage technologies across domains where you can
 - Lean into cloud solutions that will accelerate access to the latest capabilities
- Value of data decays over time...so
 - If you do not need it, get rid of it
 - If you need it, get it out of operational systems as soon as you can and "retire" it to secure archival or similar environments
- Leverage AI/ML capabilities, but make sure you apply the right model