

WHITE PAPER

USING ARTIFICIAL INTELLIGENCE (AI) TO GET MORE VALUE FROM GOVERNMENT DATA

CONTENTS

- /3 DATA GOVERNMENT'S BEST ASSET & BIGGEST CHALLENGE
- /5 WHY AI IS THE SOLUTION TO A PUBLIC-SECTOR PROBLEM
- /6 SUCCESS STORIES AND THE LESSONS LEARNED
- /7 HOW TO START USING AI TO GET VALUE FROM DATA
- /7 CHOOSING THE RIGHT PARTNER

DATA – GOVERNMENT'S BEST ASSET & BIGGEST CHALLENGE

Data is a strategic asset, and it is everywhere, so generating it is not the challenge. The challenge is how to find, connect, share and derive value from the data to make more informed, insightful decisions. Trustworthy data is therefore critical. Data quality and integrity, security, and privacy must be managed within an overarching agency information governance and policy framework.

The amount of data in the world is increasing exponentially and will continue to do so, as more and more activities are conducted online and in digitally connected spaces. Artificial Intelligence (AI) and Machine Learning (ML) have become synonymous with solving the most complicated data issues facing the government, but is the technology being applied effectively? Are we solving a "data problem" or a citizen service problem? Focusing on the agency's mission and strategy recenters the funding and resources around outcomes and results.

Society has become experts at monitoring and capturing information. As predicted, by 2025, 463 exabytes of data will be generated daily¹. Individuals simply uploading their lives and stories to Facebook create four-petabytes of data every day. In addition, the current generation of connected cars produces up to 10-terabytes of information daily via the array of cameras, sensors, and scanners they carry.

The U.S. government is a major contributor to the information surge, "creating more data today than at any point in history," according to a White House fact sheet², which also acknowledges a correlation between volume and value. "Open data has fueled economic growth, innovation and opportunity."

Cities around the country are "smart cities" where public services and transportation systems managed by municipal governments are digitally enabled. Internet of Things (IoT) devices are collecting data and delivering information to public safety officials and frontline warfighters. Civilian workers are collecting electronic records from citizens and delivering services digitally. All of these activities generate data, adding to the information overload within the public sector.

¹Techjury - <u>How Much Data Is Created Every Day in 2021?</u> ²White House Fact Sheet - **Data by the People, for the People**

HOW BIG IS DATA?



MEGABYTE =





One trillion, 600 bilion books

By 2025, 463 EXABYTES of data will be generated every day.

STRUCTURED VERSUS UNSTRUCTURED DATA



By definition, structured data is labeled information. A straightforward example is to imagine a report on the number of cars that pass along a stretch of road. The time and color of each passing car is recorded manually by hand. The result is structured data that can be easily analyzed to tell you what time of day the road is busiest and what color the passing cars were.

On the other hand, if you were to put up cameras and film the road, you would still end up with data on how busy the road is and what color cars are passing by. However, the data would be unstructured – just moving images. This means that everyday computer tools, like spreadsheets and databases, would be useless for trying to make sense of it. Without specialized tools, or someone sifting through it and labelling it manually, it cannot tell us anything worthwhile.

Unstructured data is everywhere in government; as much as 90 percent of the data generated today is unstructured³. It is in legacy archives, employee/citizen records, chat logs, legal documents, spreadsheets, HR documentation, and a multitude of other places. Data is often siloed under the stewardship of different departments – usually by whomever collected it. Others who might find it useful, whatever their role, have no way to access it or are not even aware of its existence. This means the value of our data is locked effectively away – either because we do not know where, or what it is or we do not have the tools to digitally analyze and understand it.

	STRUCTURED DATA
> Large in size	> Small in size
> Expensive to store	> Less costly to store
> Time-consuming to manage	> Efficiently managed by Al
> Vulnerable to human error	> AI reduces potential errors

Data also brings issues of compliance and regulation. Laws around the collection, storage, and use of data are becoming stricter, as shown in areas such as the Health Insurance Portability and Accountability Act (HIPAA). Other examples include the National Archives and Records Administration (NARA) using a general retention schedule to manage records, including how to dispose of records no longer needed for business purposes. In addition, the General Services Administration (GSA) issues rules for handling personally identifiable information (PII), including consequences and corrective actions in response to breaches.

At the root of all these issues is the sheer size and variety of data that agencies have at their disposal and, in particular, its messy and unstructured nature. Manually formatting, cleaning, and structuring it is one possible solution for datasets up to a certain size – but this will usually be expensive, time-consuming, and open to human fallibility. For other solutions, we can look to new and emerging technology.

WHY AI IS THE SOLUTION TO A PUBLIC-SECTOR PROBLEM



Algorithms provide machines the ability to learn, which is why these technologies are considered Artificial Intelligence.

Rather than a vast initial outlay on skills and infrastructure, they can simply tap into the existing architecture of service providers to use AI tools if/ when they need them on a pay-as-you-go basis. In 2021, the Defense Department's Joint Artificial Intelligence Center (JAIC) rolled out three new contract vehicles intended to speed acquisition of AI products and services. A center of excellence, JAIC is harnessing "the game-changing power of AI [to] transform the DoD and achieve mission impact at scale," stated the Department's CIO ⁴.

Artificial intelligence refers to computer software that can become continually better at carrying out tasks by repeating a process and learning from its mistakes without human intervention. A more precise term for this functionality is machine learning, which refers to a subset of AI technologies that simulate, as closely as possible, the learning and understanding faculties of humans and animals.

These technologies include computer vision – training computers to analyze and understand visual information. They also include natural language Thanks to the rapid emergence and maturation of cloud-based, AI as-aservice tools, ML platforms and tools are within the reach of government agencies of just about any size. Government agencies can now focus their resources on understanding the mission critical challenges they face and partner with both the private sector and other agencies to apply cloud-based AI/ML capabilities.

processing, which allows computers to understand and process human language, either spoken or in text, and respond to it using the same language. This technology is already in everyday use in consumer products such as Amazon's Alexa or Apple's Siri. Both of these technologies, as well as others made possible by ML, operate by taking messy, unstructured data and giving it structure by analyzing it according to sets of rules, referred to as algorithms. These algorithms can improve themselves; by monitoring their own output, they can learn to adjust the weighting they apply to different factors that influence the decisions they make.

For the purpose of the simple example we have been using, an AI solution to the problem of recording information about cars travelling along the road would be to use computer vision algorithms to extract the time of journey and color of the car and convert it into structured data. Natural language processing would report the results of the analysis in language necessary to take action. These applied principles can convert virtually any data an agency is gathering or storing into valuable, actionable insights.

SUCCESS STORIES AND THE LESSONS LEARNED

For every type of unstructured data an agency can generate, there are Al tools that can extract insights. The following regulated industryspecific use cases leverage Al analysis of unstructured data enabled organizations to overcome real-world challenges.



STATE GOVERNMENT ENHANCED SERVICES TO CONSTITUENTS

In the first few months of the COVID-19 pandemic, businesses closed, workers lost jobs and unemployment claims surged, from 1,000 per week to 27,000 in a day in one state. The Commonwealth of Pennsylvania's Department of Labor and Industry, unable to process claims in a timely way, turned to Iron Mountain for help. Unemployment claims processors, who were working from home as a safety precaution, lacked access to the documents needed to process the two million claims received. Iron Mountain deployed a combination of technology and a digital-service solution enabling the department's workers to remotely process unemployment claims.



ENERGY IMPROVED DECISION-MAKING

The oil and gas industry generates vast amounts of data during exploration, extraction, and supply. Datasets, including geological samples, seismic images, well logs, and lithographic surveys, hold the key to accurately assessing and realizing the value of wells. An oil company seeking to identify potential targets for acquisition worked with Iron Mountain to scan and extract information from 15,000 documents relating to 500 wells leveraging AI technology. If done manually, this task would have taken several months, and even then, there would be no guarantee all data was evaluated.

Another multinational power infrastructure provider wanted to identify and classify sensitive records detailing construction of facilities. The data could have been anywhere within six kilometers of binders stored across five premises. Iron Mountain deployed technologies, including computer vision and natural language processing, to locate the relevant information so that it could be securely stored or disposed of, as appropriate.



FINANCIAL AUTOMATED VALIDATION TECHNIQUES

Financial analysts predict that AI will disrupt the mortgage industry to increase efficiency, lower costs and improve compliance with government regulations. On a day-to-day basis, approving mortgages is a timeconsuming process that can take months and requires the completion of a sizeable amount of paperwork by both the applicant and the lender.

Iron Mountain has shown this process (as well as other standardized processes) can be automated with the deployment of automated classification, metadata extraction, and validation techniques. Al can automatically extract all relevant metadata, whether they are on paper or in digital storage. Then, every document, PDF, or spreadsheet can be index-able and searchable through a visual database, meaning questions such as whether or not a document represents the most up-to-date agreement, or how many nonstandard clauses are included, can be answered instantly.



INSURANCE SWIFT PROCESSING AND FRAUD REDUCTION

Insurance companies must have constant access to the most up-to-date information when processing claims including submitted forms, photographs, and statements. Iron Mountain has successfully deployed AI to decrease the risk posed by inaccurately assessed coverage, as well as to speed up the processing of claims and to drive increased customer satisfaction. Key AI functionality that benefits this situation includes document classification through evaluation of information, analyzing legacy data alongside newly generated claims to assist in identifying fraudulent claims, and predicting the rate of cancellations.

In the public sector, AI tools will help government agencies to combat insurance fraud and similar schemes designed to cheat funds from public programs, including Medicare, Medicaid, Social Security, and other insurance programs.

HOW TO START USING AI TO GET VALUE FROM DATA

FIT TECHNOLOGY TO YOUR AGENCY - NOT THE OTHER WAY AROUND

This simply means identifying the problem that needs solving, then finding the data and tools needed to do it. Government agencies often approach this the other way around; first identifying a technology they believe they should be using and then looking for an opportunity to use it in their agency.

2

LOOK FOR QUICK WINS WITH DEMONSTRABLE ROI

Many AI data projects have faltered due to insufficient buy-in among all stakeholders to secure the necessary commitment and investment. Today, the availability of "as-a-service" AI and ML means it is possible to deploy quick-win initiatives that demonstrate how simple it can be to generate results with data and automation. This is done best by looking for repetitive tasks that take up the time of staff who could be more productive if they were doing other things.

Engaging in pilot programs is an excellent place to start for government agencies. By pursuing a pilot on one or two data / image repositories, it is possible to disprove common misconceptions that hamper the uptake of AI within an agency – such as the concern that it is expensive, requires the recruitment of highly trained specialists, or threatens to replace human workers in their jobs.

3

BUY VERSUS BUILD

In-house development of AI capabilities requires higher amounts of up-front investment, so budget will be a key consideration when addressing this issue, as will the skill-set required to produce such programs. Agencies are better suited to rely on industry partners that have technical resources, capacity to scale, and investments to put the research and development into these capabilities. Using vendors with a prebuilt solution that can be supplied "as-a-service" is likely to result in far quicker deployment and ROI.

CHOOSING THE RIGHT PARTNER

Government agencies that partner with AI solution vendors should look carefully at the options available to them. Vendors should be able to prove how effective they have been in delivering solutions to other clients with similar requirements. Quantifying the results should be straightforward, and an AI provider should be able to demonstrate the ROI they have achieved for existing clients or the impact they have had on other significant metrics.

Security should also be a primary consideration. Agencies cannot have an instance where sensitive data is lost or stolen, even if it happens when the data is in the custody of a third-party contractor. Particularly when a project involves the processing of personal or sensitive data, every effort needs to be made to ensure contractors are compliant with best practice security protocols, as well as all data stewardship regulations. A prospective partner should always be able to explain every step taken to ensure data is safe while it is on their premises or in their cloud.

OTHER CONSIDERATIONS SHOULD INCLUDE:

- Establishing one-on-one meetings with demos, or releasing RFI's with follow-on demos
- Ensuring solid past performance with AI/ML and commercial experiences, with the ability to showcase significant ROI
- Ensuring the presence of a long standing reputation for utilizing best practices in security protocol. Security should be built into every step

CONCLUSION

Data is often siloed under the stewardship of different departments – usually by whomever collected it. Others who might find it useful, whatever their role, have no way to access it or are not even aware of its existence. This means that a huge amount of the value their data contains is effectively locked. This information needs to be unlocked from multiple standalone, on-premises repositories.

Iron Mountain InSight® provides the industry's most advanced AI-driven content services platform (CSP) to help government agencies unlock the hidden value of their content through modern, AI-infused contentdriven applications, and federated search capabilities. InSight® powered by AI/ML can enhance citizen and employee experiences, drive mission outcomes, increase employee productivity, and digitize core business processes.



ABOUT IRON MOUNTAIN

Iron Mountain Incorporated (NYSE: IRM) is the global leader in innovative storage and information management services, storing and protecting billions of valued assets, including critical business information, highly sensitive data, and cultural and historical artifacts. Founded in 1951 and trusted by more than 225,000 customers worldwide, Iron Mountain helps customers CLIMB HIGHER™ to transform their businesses. Through a range of services including digital transformation, data centers, secure records storage, information management, secure destruction, and art storage and logistics, Iron Mountain helps businesses bring light to their dark data, enabling customers to unlock value and intelligence from their stored digital and physical assets at speed and with security, while helping them meet their environmental goals. Visit www.ironmountain.com for more information.

© 2021 Iron Mountain Incorporated. All rights reserved. Iron Mountain and the design of the mountain are registered trademarks of Iron Mountain Incorporated in the U.S. and other countries. All other trademarks and registered trademarks are the property of their respective owners.

USFED-AIDATAWHITEPAPER-1121

