Chief Information Officer's Speaking Points

Big Data and Analytics Canada 2020 Summit

Version 1.0
2020-02-11

Shared Services Canada
Services partagés Canada

Canada

What was the origins of the data strategy

About Digital Government

- The Government of Canada (GC) is re-imagining its relationship with Canadian citizens
- With the view of ensuring the needs of Canadians are prioritized; the GC is streamlining and securing service delivery
- We want to ensure a consistent experience for all Canadians removing the need to navigate multiple institutions to complete a service transaction

About the Data Roadmap and call for departmental data strategies

- Our ability to re-imagine our service delivery depends on our ability to leverage data (old adage: you can't improve what you don't measure)
- Federal departments are handling increasing amounts of data with opportunities to leverage a variety of digital solutions; business models are changing as are resource and expertise requirements;
- With the landscape becoming increasingly complex and the call from Canadians for user-centric service delivery becoming increasingly louder, federal departments must shift from data as an output of program functions and systems to data as an enterprise strategic asset
- The call for departments to produce data strategies was born out of this need. These strategies are to address how each department will make more strategic use of data while protecting citizen's privacy

How we are moving from spreadsheets to data lakes

What is a data lake and how it's different from other approaches

- A data lake is usually a single store of all enterprise data including raw copies of source system data and transformed data used for tasks such as reporting, visualization, advanced analytics and machine learning. A data lake can include structured data from relational databases (rows

and columns), semi-structured data (CSV, logs, XML, JSON), unstructured data (emails, documents, PDFs) and binary data (images, audio, video).
- Shared Services Canada's (SSC) Enterprise Business Analytics Program (EBAP) "data lake" (also known as the Enterprise Data Repository (EDR)) is currently a database architecture containing only structured and semi-structure data using Oracle
- EBAP has different layers within its architecture to support data flow including swamp (raw copy of all required data from the source), staging (technical cleansing of the data), lake (aggregation and integration of data) and curated lake (user-centric data displays)

How the EDR is different from traditional data warehouses

- We are taking all the source data once to satisfy current but also future reporting needs
- We are not focused on fulfilling specific reporting requirements but rather delivering corporate data as an asset where the consumer can explore and report as required
- All data is open by default (and only unprotected data is brought into the EDR)
- We are not correcting/cleaning data in the EDR – unlike data warehouses, the Data Management Program's Data Quality Framework will address data quality issues and have remediation of quality performed in the source

What is our objective

Improve data-driven decision making at SSC by building a central repository where all corporate SSC data is housed to:

- Decrease our reliability on spreadsheets and "stitching the data together on the fly"
- Support and improve repeatable, consistent and timely reporting and analysis

How will we get there

- Onboard all corporate data of enterprise value into the EDR to deliver data as an asset

- Allow open by default access to the EDR for SSC consumers to support their decision making efforts
- Build in relationships across data domains to allow data consumers to easily "stitch" two domain of data
- Build consistent reference data to allow for a consistent approach to reporting – talking apples with apples.

How are we "upping" our data literacy

- The goal is to demystify data: make it real, understand its importance.
- We are devising and implementing an Organizational Change Management (OCM) strategy including data-related awareness and educational programs.

What technology is our EDR built on

Data sets from:

- Business Intake Tracking System (BITS) - SSC system
- Enterprise Control Desk (ECD) - SSC system
- Information Technology Service Management – Enterprise Control Desk (ITSM-ECD) - Other departmental extracts of ECD data
- Government of Canada Information Technology (GCIT) Strategic Plan - TBS extract
- Application Portfolio Management data (APM) - TBS extract

Technology to support onboarding and delivery of EDR data:

- Oracle Database
- Managed secure file transfer (MSFT) to securely transport data sets in from outside SSC to the EDR infrastructure
- SQL Server Integration Services (SSIS) to extract, transform and load (ETL) the data through the EDR
- SAS to shape, report and analyze the EDR data

What is our end-state, and what is our path to get there

- End state is based on Gartner recommendations

- Data Analytics Centre of Excellence (partnership of business and Information Technology/Information Management) – with a hub and spoke model.
- Analytics as a Service
- Data as a Service
- Goal in year 3 and beyond is to move from 'reporting' to 'predictive' and 'prescriptive' analytics in order to enable business decisions. Path to get there is to continue to build the EDR, continue to foster and build data literacy and analytics expertise, ensure funding post year 3, define Chief Analytics Officer and Chief Data Officer roles, continue to build the technology capabilities (including Machine Learning, Artificial Intelligence); and continue to define and breed the culture of data driven, data literacy and open by default (while ensuring the privacy and security of data)

Our approach for Data stewards

What do we mean by "Data Stewards"

- Data Stewards are those people in an organization responsible for the security, quality and access to their program area's data
- They establish, implement and enforce data policies, processes, procedures and standards.
- They also manage data quality issues while safeguarding against data loss.
- They are critical to the provision of trusted data assets that enable executives to make informed decisions.

Who makes a good data steward

- Familiarity with database concepts
- A good handle on data integration and management
- Specific knowledge of platform(s) in place
- Working technical knowledge of systems and solutions in place
- Strong understanding of data entry/update best practices.
- Strong client service orientation.
- Proven analytical and problem-solving abilities.
- Good written, oral, and interpersonal communication skills.
- Ability to conduct research into data issues and as required.

- Ability to present ideas in business-friendly and user-friendly language

What are our expectations of them

- They apply sound data stewardship, as per their Role and Responsibilities
- They actively participate in SSC's Data Steward Community of Practice
- Share and promote common data standards, models, and protocols to facilitate sound stewardship and use of data across SSC and with external organizations (e.g., API standards, open data standards, data exchange models)
- Guide culture change across the GC by raising awareness about existing tools, guidance and standards and advocating for their use
- Identify opportunities for data exchange driven by departmental and GC business value
- Collaborate with other GC groups and communities (e.g. GC Enterprise Data Community of Practice, Digital Academy and Canadian Digital Services) to ensure alignment on efforts to leverage data as a business and strategic asset
- Align with GC Digital Standard "Be a good data steward"

Open Government

Open Government (OG) is a means to safeguard our democracy and fuel solutions to complex problems.

Benefit

- Respond to growing demand for government information
- Increasing need for both an engaged public service that is responsive to the needs of the public and citizen participation
- Rapid development in digital opportunities transforming everyday lives of citizens, public and private sectors
- Growing recognition of OG's potential to reduce costs, create efficiencies and boost productivity both within and outside of government

Examples

- Toronto's non-profit Swim Guide and Water Guide

- They integrate data from municipal, provincial and federal jurisdictions to maintain up to data information on safety levels of Ontario's beaches and water sources.
- Recall apps
- Apps to provide recall information both on a connected smart fridge by Samsung and voice assistants using Google Home and Amazon Alexa.
- Open Government Portal API on Recalls and Safety Alerts provided by Health Canada. Re-using information from multiple sources, including the Canadian Food Inspection Agency and the Healthy Canadians website, would allow users to search for recalls, warnings, and alerts for food, consumer and medical products.
- Nepal 2015 earthquake – data driven disaster relief
- Massive amounts of data published to the Humanitarian Data Exchange cutting through jurisdiction issues related to international mobilization and multilateral organizations involved.
- Data pointed to water sources, hospitals, need for immediate aid
- Helped direct efforts where most needed

Self-serve with other departments

- Using the cloud solution we are currently testing, SSC can provide unclassified data and analytics tool for other departments to use our data for their own analytics. Allows co-creation of dashboards and visualizations, to allow SSC and partners to gain insights from our data.
- Because of the multiple toolset in our cloud solution, will allow any user regardless of skillset to consume our data for analytics. Toolsets include tools for non-expert users with a google-like experience (Thoughspot) , tools for visualizations resources (SAS), and tools for advanced data scientists (R, Python).
- In addition to the technical aspects, a roadmap and promotion strategy is being put together in order to advertise the business value.

Analytics capabilities

- Building an enterprise data repository with the focus of loading key data sets into the EDR, in order to do reporting and analytics of our services
- Defining a single source of truth and trusted data for analytics
- Developing standard push button reports that can be used for multiple service lines

What are we looking for

Using key enterprise data in our EDR, looking for answers and insights on our services such as: how well are we doing as a service provider, what are our areas of weaknesses, where are our bottlenecks, how long does it take to provide a service, time to resolve incident, etc.)

What have we learned from a technical / analytics perspective

- Need to be data driven first – start with the right data
- Need to include data management (data quality, data governance, reference data, data literacy program)
- One tool will not do it all. Requirement to diversity the analytics toolset in order to provide the full spectrum of capabilities within the department (different tools based on different roles, and different needs). However, do not want a free for all, so need to determine user profiles, and assign tools accordingly
- Design for analytics capabilities and infrastructure is complex, time to deploy lengthy in an on-premises solution. Cloud faster to deploy, however solution required to determine which data can be in the cloud, connectivity issues with moving large sets of data from on-premises systems to cloud.
- Cloud excellent for proof of concepts of analytics capabilities/tools.

What to know from a CIO perspective

Resources difficult to find – particularly Full Time Equivalents (FTEs). Important to have an innovative Human Resources strategy that now only focuses on acquiring resources, but also building the expertise needed with existing resources.