The Human Resources Branch Data Strategy:

A Road Map for an Effective Data Environment

Contents

The Human Resources Branch Data Strategy:
A Road Map for an Effective Data Environment

Vision

An environment that uses data, advanced analytics and research to inform HR decision-making, not just for the Human Resources Branch, but for the entire Agency.

Background

Over the last several years, there has been exponential growth in the demand for data across all sectors. The Canada Revenue Agency (CRA) is no exception. On any given day, the Agency collects data and information to fulfill its mandate. In addition to tax and assessment related data, CRA, as part of its daily operations, also collects a wide range of data related to employees. The Human Resources Branch (HRB) is the owner of this data and has full responsibility over its accuracy and usage. To ensure the CRA and HRB are using HR-related data to its fullest potential, we must build the data environment and capacity. This is consistent with our HR2030 vision and will contribute to ensuring the CRA is a world class tax and benefit administration.

HRB is in the initial stages of its data environment development. The branch is aware of the existence of its data and understands that it has implicit value but perhaps more importantly, understands the potential of its data. HRB's senior management supports this analytical growth which is consistent with larger organizational changes to the CRA data environment such as the Business Intelligence (BI) Vision and the creation of the Chief Data Officer role.

In fact, the Service, Innovation and Integration Branch (the home of the Chief Data Officer), has created a Data Analytics Vision for the Agency and the HRB Data Strategy will fit within this approach. The Agency vision focuses on three areas: People, Data and Environment. Specifically, the vision is to:

> Position the CRA to become a data-driven, analytics enabled organization that optimizes business value from data. This includes:
>
> Having the right people with the right skills;
>
> Making our data clear and accessible; and
>
> Fostering a culture of innovation and collaboration.

HRB is in a unique position to contribute to this vision. The CRA of now and the future needs the right people with the right skill set at the right time. We need to understand how people move through their careers and through the Agency. We need to understand our most and least effective methods of training, assessment, staffing but we also need to understand the needs of our strategic business partners - the CRA program areas. In order to do this effectively, we need to fully understand our workforce, how it has evolved and continues to evolve, taking into account larger national and global drivers. In part, this means using data and research to help inform our decisions and direction setting. This document is meant to serve as a roadmap on how we will get there but it is meant to be evergreen. As priorities change, new information is known or tools become available – the strategy will respond and adapt. A timeline of activities, or a "to do" list can be found at the end of this document.

Guiding Principle

The Data Strategy will provide a comprehensive vision and actionable foundation that will enable the branch to strengthen its data and analytical capacity in order to maximize the ability to make informed decisions, based on evidence.

The Road Map

In moving along the road map outlined below, each objective will have the direct link to the Agency Data and Analytics Vision provided.

Tools

There is a wealth of information contained in data but converting the data into meaningful intelligence can be difficult without the proper tools. Technology has modernized the way data is collected, captured, manipulated and displayed. Currently, data users in the Branch primarily use Excel as their main tool and while this program has many benefits, it is inefficient for large data sets, merging data sets and conducting most types of data analytics. There are many ways tools can be used to enhance the Branch capacity and efficiency:

- o Where appropriate, purchase sophisticated analytics tools such as SAS® and SPSS® and work with the Information and Technology Branch to enable open-sourced statistical software such as R® and Python. In addition to offering opportunities for advanced analytics, these tools are essential to obtain efficiencies not available in Excel. [**Environment**]
- o Determine the Branch requirements in terms of text mining tools and software to address qualitative and unstructured data. [**Environment**]
- o Create an automated audit trail system to monitor data access to employee data and ensure that people who are viewing the data have a legitimate business reason to do so. This approach is consistent with the tax-side of the CRA and builds confidence in employees that HRB is managing their personal data with the utmost care and security. [**Environment**]

Skills

The workforce in the Branch possesses an array of skills. Some individuals have extensive knowledge of various types HR data while others have backgrounds in analytics and research. Several initiatives will help enhance the data analysis capabilities within the branch:

- o Assess the knowledge and the skills of existing employees working in the branch to determine whether or not they have a background that we can maximize for statistical analysis and research. This will allow us to assess whether we should be reallocating resources and where training resources and tools should best be allocated. [**People**]
- o Create learning paths for employees conducting statistical analysis to help them understand the different skills required for career progression within the ES community. In addition, given the attention data analytics is receiving, this will also allow employees in other parts of HRB as well in other parts of the Agency to clearly understand what is needed to work in this area. [**People**}
- o Create a list of courses, tutorials and sources of knowledge to enhance employees' skills and their ability to use statistical software effectively. [**People**]

- Create a peer learning group to ensure skills and knowledge are shared. This will reduce silos, ensure appropriate knowledge transfer and reduce the risk of key individuals being the sole holder of essential information. **[People]**

Data Quality Assurance

Data quality is essential to any data driven environment. Without the trust in data, all use of it is tainted by questions of accuracy. There are a number of actions that will ensure strong data quality assurance:

- A comprehensive listing of the HRB's data holdings is essential. This tells us where the data used resides and the tables being used within the larger systems. It is the essential first step for the development of any data environment. [**Data**]
- Develop a data dictionary that contains the metadata[1] for all data that is used in the Branch. This will not only ensure that all users and providers understand what the field or variable measures but will also contribute to knowledge transfer and faster learning for new employees.
- Conduct various data quality assurance testing on all data to ensure its accuracy. [**Data**]
- Create a listing of commonly made errors and the solutions for correcting. Monitoring this data is essential for data quality. [**Data**]
- Use bulletins to inform data users in the Branch of new fields/variables and their corresponding definitions. Ensure that these bulletins are available on HRB's Infozone page for ease of access. [**Data**]
- Conduct audits regarding the quality of the data that is entered into any of the main source systems owned by HRB. This will provide early detection of common errors being made as well as identify individuals entering the data who need additional training. [**Data**]

Streamline Requests and Reporting

HRB's existing data reporting and monitoring regime consumes significant resources. Across the branch, there is overlap and duplication in some aspects of reporting while in other areas reporting methodology differs so significantly that different data is produced for what appears to be the same measure. This leads to instability and confusion when reporting or monitoring. There are a number of actions that can improve reporting and monitoring:

- All providers of HRB data must abide by the Confidentiality Standard as developed by the Chief Data Officer of the CRA when releasing data that is required for any reason, outside of HR operations. Following this standard will ensure that employee information is protected and that it is not possible to identify any CRA employee. [**Data**]
- Implement the use of a tool to identify data that may result in a breach of confidentiality[2] for all data products and tables leaving the Agency. The tool should also be applied for data tables going to CRA employees who are not entitled to have access to this level of data. This would not apply to CRA employees requiring data for legitimate business purposes. [**Data**]

---

[1] Metadata is essentially all the pertinent information about a piece of data, such as a variable. For example, for a field or variable called "Start Date", the metadata should contain where the variable/field can be found, how long it has been in existence, the technical definition, a common language definition as well as any anomalies associated with it.

[2] G-Confid is a tool created for this and works within in the statistical software called SAS. It was created by Statistics Canada to allow for an objective review of data to ensue confidentiality is not breach in any given table.

- Create a policy for data requesters for when micro data, unsuppressed data and suppressed data are provided to ensure we apply the policy to the data and not the requester (e.g. ATIP, media requests). [**Data**]
- Assess all existing internal dashboards and reports for overlap and duplication as well as appropriateness of the frequency of reporting. [**Data**]
- Identify and maintain a list of data indicators that must be reported on to the Board of Management, the Agency Management Committee, the Human Resources Committee and other reporting requirements to ensure a full picture of HRB's reporting requirements. This list should be updated annually and available on our Infozone page. [**Data**]
- Identify areas of common interest for users and ensure data products are readily available to reduce the number of individual requests from clients. [**Data**]
- Develop a "common look and feel" to facilitate the assimilation of the information within and outside the Branch. This common look for all HRB data products will include proper referencing of sources, definitions, etc. and will ensure that all data products produced are easily defined and replicated. [**Data**]
- Use plain language when providing data products. [**Data**]
- Place as much data as possible on the Agency Data Warehouse to facilitate the appropriate sharing of data with our business partners. [**Environment**]
- Develop a shared data warehouse for all HR data holdings in order to gain efficiencies, reduce duplication of effort, and enhance consistency and reliability of reporting. [**Data**]
- Maintain a data clearing house in the Branch that is consistent with the clearing house for taxpayer data. This includes providing a quarterly report to the Chief Data Officer. All data that is leaving the Agency would be included here. [**Data**]

Data Literacy

As skills and knowledge of statistical analysis are essential for producing complex statistical analysis, so is a true understanding of the data being used. There are many ways to input and interpret data in the systems which can yield inaccurate data/conclusions/recommendations, thereby leading to decisions based on erroneous information. Data literacy is needed at all levels to ensure appropriate expectations of what the data can and cannot say, how data is processed, and the basic cleaning process.

- Provide easy and clear access to all data holdings and data definitions/metadata on the HRB Infozone page. [**Data**]
- Manage expectations by clearly, widely and routinely informing clients about the effort and time required to produce data, especially for ad hoc requests. [**Data**]
- Communicate the confidentiality standard to all users. [**Data**] [**People**]
- All data providers work closely with the content experts to ensure all data tables or products have the full context to the information. [**Data**] [**People**]

Conducting Research with HR Data[3]

Advanced analytics is essentially conducting research. It begins with a question that needs to be answered and a methodology on how best to answer it. While this seems simple, there are actually a

---

[3] Research and advanced analytics are very different from reporting. Research uses the scientific method of systematic observation, measurement and experimentation through testing a hypothesis or research questions. Advanced analytics are the tools used on the data to test different hypotheses/research questions. This is not used

number of complex factors that need to be considered prior to conducting the research if the answer is to be trusted and reliable. Many government departments have well-established research environments and to bring CRA's HR research environment in line, the following actions are needed:

- All proposed research using HR data will be presented and approved by the Branch Executive Committee (BEC). This will ensure that resources are being used to address key priorities of the Agency and government. It will also reduce the chance of research being conducted that is unethical, inappropriate or repetitive. [**Environment**] [**Data**]
- Create a research proposal template that must be filled out in order to bring a proposal before the BEC. [**Environment**] [**Data**]
- Ensure an ethics review is done for each research project to consider the impact of the research, confidentiality concerns as well as potential impacts to those involved in the research (including primary and secondary data collection)[4]. [**Environment**] [**Data**]
- Place all planned research activities on the Agency Research Plan[5] via the Agency Research and Innovation Committee for which HRB is an existing member. [**Environment**] [**Data**]
- Post all completed research on the HRB Infozone page and share with the HR community in the government. [**Environment**] [**Data**]

Data Governance Framework

All data environments must have a clearly articulated data governance framework that sets out the management and accountability of all aspects of the data owned by a given area. A data governance framework outlines all processes related to data integrity/quality, allowing access, definitions, collection and approval of use. In short, it will describe who is responsible for each of these aspects and the process to make any changes to the data environment.

- Write and publish a Data Governance Framework for HRB data [**Data**]

Open Data and Open Government

The government is committed to being more transparent and open with information. By posting CRA's data on the Open Data Portal, we would be contributing to that commitment, while at the same time, possibly reducing the number of ATIP's, media requests, etc. that are received.

- Create an open data plan for the branch so a broad range of aggregate administrative data becomes available. [**Environment**]

---

for data requests where someone is looking for the number or percentage of something but for more complex questions that require the examination of a number of different data sources using more complicated statistical techniques. For example, we can guess at the number of retirements in the next five years by using the percent and number of retirements in the last five years but if you truly want to predict retirements, you would need a clear research question, a number of data sources and then develop a statistical model that you could assess the role or influence of each of the different data sources on the decision to retire.

[4] Primary data collection involves informing participants that the research is being conducted, why, that they have the option of participating and that any information they provide will be kept confidential. Secondary data collection involves the use of data collected for other purposes. Depending on the nature of the research and data usage, consent may or may not be required.

[5] This plan is approved by the Agency Management Committee each year and ensures that all projects are reflective of Agency and government priorities.

- o Work with the Strategy and Integration Branch and the Chief Data Officer to place appropriate data sets on the Open Data Portal. [**Environment**]
- o Create a mechanism for Canadians to request HR data products from the Agency outside of the formal ATIP processes. [**Environment**]

Financial Implications

The Branch will incur costs from the implementation of the data strategy. Tools, training and time spent by the workforce on the implementation of the strategy will not be negligible. There are definite costs to purchasing and maintain software and sending employees on training. However, some of these costs for employee time will be counterbalanced by a reduction in unnecessary data reporting and manual efforts. Other costs, such as enabling advance analysis or ensuring proper use of data through an automated audit system, will not easily be associated with a tangible product but rather work towards the vision for data in the Branch. The bulk of the work in this road map will come from in-kind contributions by those already working with data in the Branch, but mostly from the Strategic Business Integration Directorate.

Data Technology Infrastructure

This road map does not address the technical aspects related to the IT infrastructure of HRB data systems. A separate road map written with this one in mind would be an important companion piece in our goal of creating a strong data environment for the Branch.

Next Steps

- o Communicate the strategy to all HRB employees
- o Report on progress against the action items in the road map on an annual basis at which time, changes or amendments may be made to the strategy to reflect priorities and other initiatives.

HR Data Strategy – Roadmap Action Items

Action Item	2018-19 Q1	Q2	Q3	Q4	2019-20 Q1	Q2	Q3	Q4	2020-21 Q1	Q2	Q3	Q4	2021-22 Q1	Q2	Q3	Q4	2022-23 Q1	Q2	Q3	Q4
TOOLS																				
Purchase sophisticated analytics tools (where appropriate) and examine open-sourced statistical software.	█	█	█	█	█															
Determine the requirements in terms of text mining tools and software for qualitative and unstructured data.						█	█	█	█											
Create an automated audit trail system to monitor data access to employee data.																	█	█	█	
SKILLS																				
Assess the knowledge and the skills of existing employees to maximize their potential for statistical analysis.	█	█	█	█																
Create learning paths for employees conducting statistical analysis.	█	█	█	█																
Create a list of courses, tutorials and source of knowledge to use statistical software effectively.	ongoing																			
Create learning group, increase knowledge transfer, reduce the risk of sole holder of essential information.																				
DATA QUALITY ASSURANCE																				
Create a comprehensive listing of the HRB's data holdings.	█	█																		
Develop a data dictionary that contains the metadata for all data that is used in the Branch.				ongoing																
Conduct various data quality assurance testing on all data to ensure its accuracy.									█	█	█	█								
Create a list of commonly made errors and solutions for correcting					ongoing															
Use bulletins to inform data users of new fields/variables and their corresponding definitions.																				
Conduct data validation regarding the quality of data entered into any of the main source systems owned by HRB.																				
STREAMLINE REQUESTS AND REPORTING																				
Disseminate and enforce the use of the Confidentiality Standard when releasing data outside of HR ops.	█	█	█	█	█															
Implement the use of a tool identifying confidentiality breach contained in data and tables produced.	█	█	█	█																
Develop data suppression standard applied to the data, not the requestee	█	█	█	█																
Assess existing internal dashboards and reports (for overlap and duplication) and frequency of reporting.	█	█	█	█																
Identify indicators that must be reported to the Board, AMC, HRC and others, ensure a full picture of HRB.	ongoing																			
Identify areas of common interest and make the data readily available to reduce the number of requests.	█	█	█	█																
Develop a "common look and feel" to facilitate the information assimilation.																				
Use plain language when providing data products.	ongoing																			
Place as much data as possible on the ADW to facility the sharing of data with our business partners.		█	█	█																
Create a shared data warehouse for all HR data holding					█	█	█	█	█	█	█	█								
Maintain a data clearing house and report biannually to the Chief Data Officer (CDO).																				
DATA LITERACY																				
Provide easy and clear access to data holdings and data definitions/metadata on the HRB Infozone page.					ongoing															
Manage expectations by informing clients about the effort and time required to produce data.	ongoing																			
Communicate the confidentiality standard to all users.	ongoing																			
Work closely with the content experts to ensure data tables or products have the full context.	ongoing																			
CONDUCTING RESEARCH WITH HR DATA																				
Create the Executive Research Review Committee via BEC to review and approve all planned research.					ongoing															
Create a research proposal template that must be filled out in order to bring a proposal before the RRC.	█	█	█	█																
Ensure an ethics review is done for each research project using proper submission template	█	█	█	█																
Place all planned research activities on the Agency Research Plan.	ongoing																			
Post completed research on the HRB Infozone page and share with the HR community in the government.													█	█	█	█	█	█	█	
DATA GOVERNANCE FRAMEWORK																				
Write and publish a Data Governance Framework for HRB data.	█	█	█																	
OPEN DATA AND OPEN GOVERNMENT																				
Create an open data plan so a board range of aggregate administrative data becomes available.					█	█	█	█												
Work with the Strategy and Integration Branch and the CDO to place data sets on the Open Data Portal.						█	█	█	█	█										
Create a mechanism for Canadians to request data from the Agency outside of the formal processes.									█	█	█	█	█							

Legend: ■ People (yellow) ■ Data (green) ■ Environment (blue)