

Data sources are the key to understanding your systems and helping in determining scope. The more detail, the better. Are you able to provide a full list of data sources beyond the examples provided?

We are still in the process of finding new data sources for some of our metrics, but the following are key contributors to our insight generation:

- *Statistics Canada*
- *Conference Board of Canada*
- *Lightcast (previously Burning Glass)*
- *CBRE*
- *CMHC*
- *Colliers Canada*
- *Crunchbase*
- *D&B Hoovers*
- *City of Calgary Open Data*
- *Province of Alberta Open Data*
- *FP500 Database*
- *Open Data Canada*
- *Surveys and studies*
- *Dynamics 365 (CRM) – Future possible need*
- *Dealroom*
- *LinkedIn Talent Insights*
- *Various reports: PWC, RBC*

Can you provide the major business domains captured from public sources and third-party data sources that are listed in Appendix 20.A?

- *The following is not exhaustive, but provides some examples:*
 - *Real Estate Data – Rent, vacancy rate etc.*
 - *Labour/Workforce – skills, sizes, gains/loss, unemployment, demographics, participation, employment by industry, job openings, credentials, NAICS/NOC monitoring etc.*
 - *Calgary Market Information – GDP, house prices, taxes, head office counts, small business counts, businesses by industry, airport connectivity, demographic changes, immigration, graduates, post-secondary statistics, including awarded degrees etc.*
 - *Comparisons with other cities in Canada and the US are common*
 - *Examples of output reports:*
 - https://www.calgaryeconomicdevelopment.com/assets/Reports/WhyCalgary_Our-Economy-in-Depth-2022-06.pdf
 - https://www.calgaryeconomicdevelopment.com/assets/Reports/Labour-Force-Survey-Results_February-2023.pdf
 - <https://www.calgaryeconomicdevelopment.com/insights/demographics/>
 - <https://www.calgaryeconomicdevelopment.com/invest/economic-indicators/>

Is there a data governance and stewards' program that is in place?

- *Not at this time. We look to establish data governance practices during the course of this project.*

What are the dashboarding tools that are used? Are there any tools that are considered as part of future implementation?

- *Today, mostly visuals available in PowerPoint and Excel. Only recently have we started working with Power BI.*

Have you considered a data lake as an end-state architecture?

- *We would look to vendor guidance on this approach.*

Object storage? Will S3 be suitable?

- *We would look to vendor guidance on suitability of this suggestion.*

What is the total volume (GB) of your data? Do you know your expected rate of growth?

- *Today, there is about 100GB for research/insights data stored. Growth rate cannot be determined at this time.*

What does your data team look like today? Data Scientists? Data Engineers?

- *There are not currently any specialized roles related to Data at CED. We look to upskill or augment roles as part of this project.*

Do you have the in-house skill sets required to write ETL jobs for data ingest and data transformation? If yes, what languages?

- *There is entry-moderate level knowledge of ETL at CED. Some knowledge exists in SQL, Python and DAX. Background in ETL is in Power Query only.*

How often does the data need to be ingested?

- *Most data will need to be refreshed monthly.*

What is the lifecycle for how long data should be stored and/or moved to colder storage options?

- *As part of this project, we look to establish these parameters within our data governance practices.*

Is it important for this environment to be running all the time? Would you benefit from a self-service, ephemeral environment? (e.g., Data Science Workbench)

- *We will need more context for this question, but the environment is likely only to be accessed during business hours from Monday to Friday.*

Are any licensed tool for creating the framework for storage, ingestion, integration, and processing tools being considered?

- *We look to vendors for suggestions, but initial considerations have included Power BI, Tableau and Alteryx. Alteryx primarily being used if CED is able to leverage / re-use already available storage and compute resources from a Canadian telecom provider.*

What is the project management methodology followed for the implementation of architectures? Is the Agile approach to project management preferred?

- *Agile would be preferred.*

What is the future team formation you envision? Tooling used by the team? Do you see your current technical team running all future data pipelines?

- *CED is aware that we may be required to upskill or augment existing roles to successfully operate and maintain our data platform. We are open to suggestion on various ways of operating in the future, plus roles that should be considered. It would be beneficial if CED teams can establish API connections for new sources and manage most of the platform in a self-service way.*

As a result of this RFP, would you award a contract to:

(a) a vendor who can provide a plan for an initial architecture/platform based on the objectives/deliverables and also do the implementation of the Data Modernization project

or

(b) a vendor who can provide a plan only for an initial architecture/platform based on the objectives/deliverables. If so, would you release an RFP later for the implementation of the project specifically?

- *CED's preference is to work with a Vendor who can guide both architecture and implementation of the recommended platform to ensure continuity between the proposal and what's delivered.*

Are US based firms eligible to respond to this RFP? Are there any prerequisites to respond to this RFP?

- *Yes, US firms are eligible to respond to the RFP.*

Can you please confirm that per Section 13. Deliverables what is intended by ‘...coordinate and execute delivery of initially agreed upon strategy, technologies and platform to achieve MVP’. Does that mean the successful proponent would be responsible for delivering all items listed under Section 12. MVP Phase and Initial Deliverable including implementing storage, ingestion, integration and processing tools, plus at least one visualization tool? And if so, then are these all to be included in the budget of \$100,000?

- *It is desired by CED that the selected Vendor would be responsible for delivering all items in Section 12 within the \$100,000 budget. However, Vendors are welcome to submit a proposal whereby the implementation/delivery portion of the suggested solution is CED's primary responsibility. This assumes that the Vendor provides a clear plan and guide for all platforms/technologies that CED should procure and integrate to successfully achieve the MVP scope.*

What volume and frequency of data is coming from each respective source system listed?

- *In most cases, data from each source is updated monthly as it becomes available. Some PDF and other reports are quarterly. Most CSV files are between 1-2MB and about 200-300 would be updated monthly. Most are CSV/XLS. We look to grow this number as we access more sources of data. PDF documents range in size from 1-5MB and there would be less than 20 per month at this time.*

Is ingestion of all data from the listed source systems in scope for the MVP or a subset? If so, can you confirm data volumes per source to be ingested?

- *CED would like to begin ingesting from these sources as part of the MVP. The expectation is that systems would be in place to begin ingesting data from these sources over time, but not all needs to be ingested for MVP. As an example, Stats Canada could be a first source to begin ingesting. An example Stats Canada dataset and files can be found [here](#).*

How are PDF data sources being used today? Are they processed for downstream analytics/reporting (e.g., through a document processing pipeline)?

- *Today, PDFs are manually scrubbed for data and then entered into a spreadsheet. There is no automation at this time.*

Will a pilot use case be selected for the MVP to validate the platform / analytics tool selected? If so, has this use case and its respective data sources been defined?

- *Yes, CED will likely focus on one or two data sources for the MVP. Stats Canada is a likely contender, as is Conference Board of Canada data. There will be a further focus on a specific theme of data such as labour and workforce statistics.*

Is any reporting/analytics being done today that will need to be migrated to the platform? If so, is this migration in scope for the MVP and can you provide details on the existing analytics solutions?

- *CED plans to treat this program as a blank slate and an opportunity to start fresh with new reporting and analytics mechanisms. As such, there is no anticipation of migration.*

Are there a large number of predefined reports/extracts which are based on similar sets of data?

- *Yes, Stats Canada would be an example of a key data source with predefined extracts that are frequently used.*

Do you expect that Master Data Management will be required for matching and merging entities from various data domains, and if so – what are the expected data domains (customers, addresses, etc.)?

- *MDM will be desired but may not be an MVP requirement. CED largely matches and merges data based on geography (city, province/state, country), NOC codes, NAICS codes and other industry-standard identifiers.*

In the FAQ, it is mentioned that CED may upskill/hire to support the implementation, operation and maintenance of the platform. Are any CED resources expected to support the MVP build, and to what extent?

- *Yes, CED has a Director specifically assigned to this initiative with project/program management background and would also be expected to play a pivotal part in change management across the organization. They can assist where needed, as this platform will be their key focus area. Additionally, CED's Economic Intelligence team will assist with data sources, plus has some background in coding and data-analyst type responsibilities. This team will be the primary user of ETL and visualization tools deployed at CED.*

We are assuming we can leverage our [delivery resources] in Europe and Asia. Please confirm if there are any location requirements for delivery resources.

- *CED does not have any location requirements to complete this work.*

Are there any applications that are currently running on AWS, Azure, GCP that should be taken into consideration?

- *CED primarily uses MS365 and Azure today.*

Is implementation of CI/CD pipelines and deployment to higher environments in scope for the MVP, or will the vendor be working within one environment?

- *We are currently planning to work within one environment.*

Are you looking for the vendor to be engaged beyond the initial MVP for future efforts?

- *Possibly, depending on resource availability and CED's needs to expand platform capabilities over time.*

Is the requested response one that includes only overall approach and services, or is it expected to include a technology platform as well?

- *Our desire would be that the Vendor can cover approach, services and at a minimum, a high-level recommendation for a technology platform in the response. We realize that through additional discovery the platform and architecture could evolve. It would be beneficial if the Vendor can also assist with some portions of delivering/implementing the recommended technology platform, so ideas around that would also be welcome in the response.*

Section 11 states the objectives of the project. Can you describe why this project is important to your organization? Section 9 does state the background, but we are interested in the thinking of your organization.

- *CED looks to elevate their overall capabilities to become a leader in economic data management and analytics. As part of this project, CED aims to:*
 - *Reduce time spent on manual data ingestion and processing*
 - *Implement a source-of-truth that everyone in the organization has access to and is working from*
 - *Improve speed to insight. As new data becomes available, immediately focus on synthesizing and analyzing that data, rather than spending time integrating it.*
 - *Integrate from multiple sources – sometimes not possible today due to time required*
 - *Improve efficiency of generating visuals with data for PowerPoints, reports and other presentations*
 - *Implement new tools for deeper analysis and trend monitoring*
 - *Reduce leads times to respond to data requests from external parties*

After the core data platform gets built for the MVP, what are the immediately high priority use cases which are powered?

- *CED aims to begin automating ingest of data from frequently used sources such as Statistics Canada, Lightcast and Conference Board of Canada, which will dramatically impact today's manual workloads. Additionally, as data is ingested, we plan to begin building dashboards that are accessible throughout the organization as a starting point to pivot to data-focused operations. CED will establish a focus during the MVP on specific data subjects to work with first, such as labour/workforce, and will build dashboards and automate manual existing reports with these new capabilities.*

What is the long term goal 2 years from now?

- *The following are high-level goals within a 2-year period:*
 - *The majority of data sources are automated and ingested as new data becomes available.*
 - *The entire organization has access to dashboards that leverage all data sources CED has access to. Team members will have access to a source-of-truth and the latest data, plus they can build comparisons and view historical trends in the data.*
 - *Integrate additional data sources that aren't used today (TBD what these are) due to reduced time spent on collecting and processing data manually.*
 - *Have faster access to insights and key changes in data from platform intelligence*

What is the long term goal now 5 years from now?

- *Integrating foresight and AI technologies to add context to current and historical data. Being able to see relationships and trends that otherwise would have gone unnoticed and provide context for them within global signals of change.*
- *Cement our position as a leader in the community for data and insights related to Calgary*

When looking at a data platform is there a preference in going 100% cloud native (meaning everything is developed in AWS with cloud native services), versus using low-code tooling like Alteryx?

- *CED endeavors to operate this platform with limited augmentation of data-specific and specialized roles. As such, we would look to the vendor to recommend how to achieve that. Overall, we would prefer low-code or no-code solutions throughout.*

Is it expected that the target MVP design will be finalized as we respond to the RFP? Or can it be adjusted during deep-dive discovery meetings prior to contract signing?

- *CED anticipates that the MVP and accompanying deliverables will evolve during conversations with vendors. Design does not need to be finalized during response, but a high-level vision of how to achieve MVP would be desired*

Are there any Canadian Provincial restrictions on data residency?

- *No, but data storage should reside in Canada.*

Are there any current regulatory compliance requirements that CED has? If so, can you please list them?

- *Primarily that any stored data resides in Canada.*

Do the self-service visualizations/insights need to be accessed by the public/unauthenticated users?

- *Yes, CED intends to integrate select visuals and backgrounds on our public website.*

How many users will be responsible for creating the insights/dashboards/visualizations?

- *To begin with we estimate about 5, with that number growing to a maximum of 10.*

How many users will need to access insights/dashboards/visualizations?

- *We anticipate that 40-60 users will require access to view dashboards and visualizations.*

Do we need to import all existing data resides in Sharepoint storage, excel files to new data platform?

- *No, CED plans to use this opportunity as a fresh start.*

What could be Statistics Canada API invocation frequency we need to plan - daily / weekly / monthly or on demand?

- *Most Stats Can data that is used by CED is updated monthly.*

Statistics Canada offers 15 date methods. Do we have to integrate all of them?

- *We would anticipate selecting one to use ongoing.*

We are not sure but Statistics Canada may have some data access threshold restrictions either per minutue / hour or day. In such case, if that may cross the limits, are you looking to opt-in for paid subscription, if any, so that we can always retrieve data. Please let us know.

- *We don't believe that this will be a concern for CED given that we will mostly update monthly with refreshed data available on different days.*

Lightcast is offering dataset from multiple countries. We assume that we will pull data for Canada only. Please let us know.

- *Yes, we would primarily be using data from Canada and possibly the US.*

We assume that Conference Board of Canada data require manual download and submit to ETL processing pipeline. Please let us know if you have any suggestion to automate (like web crawl etc).

- *Yes, this process is manual. We would look to the vendor to suggest ways that this can be automated (if possible).*

CVCA reports are in PDF format and we would like to understand what are the information you are planning to extract from it. Please share as much details you can as it may require to introduce Machine Learning to the extend to automate process.

- *We would primarily be looking to extract data from tables within the PDF. We understand that it is difficult to extract from visuals and text.*

Additionally, the RFP mentions project data. I was curious to know how project data is currently structured and what tools are being used to manage it. This information would help me to ensure that the proposed solution is aligned with current practices and can provide an improved solution for managing project data.

- *CED leverages various project management tools that may or may not store that we would like to report on. Any data would be exported from the software (Monday.com, Trello etc.), so it is assumed that it would have some structure.*

Furthermore, I would like to gain a better understanding of the current use case for prep and process. Could you please describe the current process for preparing and processing data and why it may be problematic? Is it because the data is currently in Excel and there are limitations to Excel's capabilities?

- *All data processes at CED today are highly manual. There is no normalization of date formats, geography or establishment of relationships between multiple datasets. Each dataset is loaded to Excel and then analyzed using built-in visualization tools, or in PowerPoint. As an example, if data from Stats Canada and Conference Board is being compared or merged, just values may be manually copy/pasted to ease visualization creation.*

In addition, I would like to know if there is a database of any kind currently in use at CED, such as MySQL, Postgres, MongoDB or NoSQL.

- *CED does not currently leverage any of these databases.*