BIG DATA PLATFORM POC

The Social Sciences and Humanities Research Council was interested in measuring the efficacy of their grants in terms of the benefits that result from various types of funding; for example:

| How many research products were created?

| How many Canadians benefited directly or indirectly?

| What types of beneficial outcomes were achieved?

SSHRC was faced with the task of gathering and integrating data sources from various disparate systems in differing formats in order to synthesize a holistic view of their grants and outcomes.

The SSHRC engaged Lixar to assist with these challenges. As a first step, exploratory discussions were held with keys stakeholders at SSHRC to review available datasets and understand the core use cases for the project. SSHRC provided three data sets comprising post-funding project results, plus an additional data set with project descriptions and additional details.

After importing the raw files into Azure Data Lake Store, Lixar began inspecting the data in preparation for analysis and visualization. Lixar cleaned up the data and performed and extraction of the data to make the values suitable for reporting. The scripts developed by Lixar aligned the relevant columns, joined the individual survey response files, and then incorporated the supplemental data from the file containing additional details.

This system, the result of only a handful of days' of effort by Lixar's engineering team, demonstrated the suitability of the "Data Lake" approach in terms of being scalable and flexible while providing a good solution for SSHRC's needs.

The resulting cleansed file was then imported into Power BI for analysis and demonstration. Lixar's data visualization team examined the data, and created two interactive dashboards aligned with SSHRC's original requirements of measuring the value of grant investments. Through feedback sessions the dashboard content and functionality was refined to meet SSHRC's needs and address their use cases.