

4. Statistical Metadata Systems (Statistics Canada)

4.1 Metadata system(s)

(a) Integrated Metadata Base (IMDB)

The IMDB is based on the ISO/IEC 11179 Metadata Registries and the Corporate Metadata Repository model (CMR). The metadata layer extends across all phases of the statistical business process and can support disseminated data, analysis, archived datafiles, and the planning and design of surveys. Metadata in the IMDB is beginning to be linked to some data warehouses, which hold both micro- and aggregate data; and can be potentially used for data analysis including data benchmarking and data confrontation.

(b) Integrated Business Surveys Project (IBSP)

The business statistics program includes approximately 250 surveys and administrative-based programs. The IBSP was initiated in April 2010 to make use of shared and generic corporate services and systems for collecting, processing, disseminating and storing statistical information. Content for business surveys is to be harmonized wherever possible^[1] and the approach to data analysis streamlined across programs.

(c) The Common Tools Project

The goal of the Common Tools Project is to implement a harmonized set of processes and tools to support social surveys. The project is divided into two primary environments, the Social Survey Metadata Environment (SSME) and the Social Survey Processing Environment (SSPE). SSME uses four tools to feed information to and from a metadata repository. This allows metadata to be documented once and reused throughout the process to improve quality and generate efficiencies.

[1] A project to harmonize content for business surveys will be undertaken in 2013.

4.2 Costs and Benefits

Detailed costs associated with implementing these projects are not yet available. Benefits from establishing the IMDB include more rigorous information and metadata management, harmonization and standardization of concepts, knowledge sharing and reuse of information assets.

The IBSP has incorporated good metadata management and is starting to integrate systems that are metadata-driven and that optimize the use of corporate services such as collection and methodology. Reuse of content modules, increased use of administrative data and the adoption of electronic questionnaires are expected to reduce micro-editing, standardize methods and processes and lead to operational efficiencies.

Creating a metadata environment (SSME) within the Common Tools project has facilitated the transfer of information between business processes through a suite of tools for survey documentation.

Overall, these projects are reducing system maintenance, development and training costs while improving sharing of standards and best practices.

4.3 Implementation strategy

(a) Integrated Metadata Base (IMDB)

The IMDB has been implemented using a "step-wise" approach with three development phases and future opportunities to re-use metadata and expand the IMDB metadata model to link to other information systems in the Agency.

Phase 1 produced a set of static web pages displaying data sources and methods for each statistical program and survey. These were accessible to external users through hyperlinks from data tables and publications on the Statistics Canada website. Internal users could browse the full inventory through the Agency's Intranet site.

In 2000, Phase 2 began collecting reference metadata including survey methodology and data accuracy. Information was formatted and validated by subject matter areas before being loaded into the IMDB. Like the initial phase, Phase 2 information is available to external users through hyperlinks on the website and an internal version on the Intranet site. Updates are triggered by new data releases so that metadata accompanies every release.

The past 10 years have seen improved quality of Phase 2 content and a push to include more information in the IMDB. Phase 3 has been initiated to add definitions of concepts, variables and classifications for all subject matter areas. This work is expected to be completed by spring 2015.

(b) Integrated Business Surveys Project (IBSP)

The IBSP has produced a semantic model with a proposed IBSP metadata classification, along with an agreed-upon vocabulary of terms and definitions. Details of metadata by system components are being identified. Current plans are to develop one single portal to make accessible overarching metadata, the data it describes and the processes it controls. This tool will enhance reporting capabilities and built-in quality

assurance.

(c) The Common Tools Project

Systems have been developed using the Rational Unified Process (RUP), an iterative approach. Using this method, several tools are in development concurrently. The priority is to deliver basic functionality as quickly as possible and to combine all tools into an integrated processing and metadata management environment. The initial phase of the project created a processing environment and associated standards including naming conventions and directory structures. Subsequent phases cover individual processing steps.