WP. 17
ENGLISH ONLY

**UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)**

**ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS)
(Geneva, 3-5 April 2006)

Topic (iii): Metadata and the Statistical Cycle

METADATA TO SUPPORT THE SURVEY LIFE CYCLE [1]

Invited Paper

Submitted by Statistics Canada,[2] Canada

[1] The opinions expressed in this paper are those of the author and do not necessarily reflect the official policies of Statistics Canada.

[2] Prepared by Alice Born, Statistics Canada. Contact: alice.born@statcan.ca The author wishes to acknowledge the contributions from Amie Lee, Carmen Greenough and Susan Ingram of Statistics Canada in the preparation of this paper.

I. INTRODUCTION

1. The Integrated Metadatabase (IMDB) is the corporate repository of metadata for each of Statistics Canada's current 567 surveys. These surveys are the Agency's core activities and the IMDB is the principal mechanism by which they are documented, providing a key information resource for data users and for corporate knowledge management. Under Statistics Canada's Policy on Informing Users of Data Quality and Methodology[3], data users are provided descriptions of the underlying concepts being measured, and the methodology and indicators of the quality of data Statistics Canada disseminates. Metadata and related documentation must conform to standards and guidelines issued under this policy. Also, Statistics Canada's Quality Guidelines[4] provide guidelines on how to describe the concepts and on indicators of data quality that should be reported as part of the metadata.

2. The content of the IMDB is organized around the survey entity. For the purposes of the IMDB, the term "survey" refers to the collection, analysis and reporting of data concerning characteristics of a population. These data may be collected directly from survey respondents, derived from other Statistics Canada surveys and/or collected from administrative files.

3. The IMDB contains statistical metadata and includes both reference and structural metadata.[5] Reference metadata describe key administrative characteristics, data sources, methodology, and measures of data accuracy of the survey. Structural metadata (or definitional metadata) refer to variables and their definitions, and related classifications. The objective of this paper is to describe, in detail, the common set of reference metadata related to the survey life cycle as presented in the IMDB. The paper illustrates one metadata model developed at a national statistical office in order to meet the requirements for disseminating statistical metadata to its data users. However, as the demand for statistical metadata, particularly from international organizations, increases, there is growing internal pressure to reuse existing metadata in the IMDB and add administered items to the model to fill these requirements.

4. The paper presents a description of the IMDB; details on the metadata supporting a survey life cycle; tools for entering the metadata; versioning rules in the IMDB; and additional administered items external to the IMDB and for future development.

II. DESCRIPTION OF THE IMDB

5. Statistics Canada has implemented a corporate metadatabase that stores metadata on its 566 current surveys and statistical programs. The IMDB contains another 312 records for surveys in various states (e.g., surveys with no publicly disseminated data, amalgamated, etc) for historical purposes. The content of the IMDB has been selected to suit it primary purpose, which is to provide users with information needed to interpret the statistical data that Statistics Canada disseminates. The type of information provided covers the data sources and methods used to produce the data published from surveys and statistical programs, indicators of the quality of the data as well as the names and definitions of the variables, and their related classifications. The metadata supports all of the Agency's dissemination activities including its online data tables, CANSIM and Canadian Statistics, publications and daily data releases. The IMDB has been built to facilitate the maintenance of historical statistical metadata as well as providing a snapshot of the metadata at any survey instance as far back as November 2000 – the starting point of the IMDB.

6. The IMDB model is based on the ISO/IEC 11179 specification and standardization of data elements, the Corporate Metadata Repository (CMR) from the U.S. Bureau of Labor Statistics, and an extension of the American National Standards Institute metamodel (ANSI X3.285). ANSI X3.285 was recently incorporated into the ISO 11179 standard. The CMR model consists of a data dimension, business dimension, administration

[3] http://www.statcan.ca/english/about/policy/infousers.htm
[4] http://www.statcan.ca/english/freepub/12-539-XIE/index.htm
[5] Terms taken from SDMX documentation. Structural metadata refers to the description and identification of statistical data and reference metadata describes and qualifies statistical datasets and processing more generally.

and document dimension, and terminology and classification dimension.[6] For purposes of this paper, Statistics Canada's application in the IMDB of the business dimension of the CMR, which supports the survey life cycle activities of statistical offices, is described in detail.

7. The database is implemented in Oracle 9i and is resident on a central server. Currently, the metadata is published on the Statistics Canada website[7] on HTML pages generated from Perl scripts and Oracle PL/SQL. These HTML pages are the basis for dynamically generated web pages that directly access the database. The database is kept up to date through a graphical user interface (GUI) tool, implemented in Java, and deployed over the Internet to desktops of the stewards of the metadata (e.g., Standards Division and selected survey divisions in Statistics Canada). Updates are quality assured and registered before they are made available for generation of the external HTML pages.

A. Metadata supporting the Survey Life Cycle

8. The IMDB model defines the entities for describing Statistics Canada's surveys and statistical programs, their content and their methodology, and the relationships between them. The model supports the metadata requirements of many of the phases of the survey cycle including survey design, data collection, input processing, derivation, estimation, aggregation, dissemination and post-survey evaluation.

9. The basic structure of the metadata in the IMDB is illustrated in Figure 1. Each entity is referred to as an administered item. Each of the administered items currently in the IMDB represents a part of the survey life cycle[8] and the Data Dimension of the CMR (i.e., data elements and value domains). Administered items are defined, and may be reused or shared; and they are also managed, tracked and organized. In order to complete the latter, each administered item is supported by the following "regions", outlined in red in Figure 1. The *stewardship* region (e.g., organization, contact and documentation) supports the administration aspects of the administered item such as the responsible division and information for registration as well as supporting documentation. The *identification* region (e.g., identification and time frame) manages the name of the administered item and the time context for the administered item. The *classification* region (e.g., keyword and themes) manages the classifications and keywords to which administered items are assigned. In Statistics Canada, some administered items (e.g., surveys and questionnaires), data tables, data releases and publications are organized around 27 top themes and 221 sub-themes. Those administered items shown in grey have not been implemented in the current version of the IMDB.

10. In Figure 1, the administered items have been grouped into items that support information about the survey and its "umbrella" statistical activity; the survey methodology; and data elements. The green arrows show some of the relationships between these administered items. In the model, all the administered items describing data sources and methodology (i.e., methodology box) are attached to the survey instance; survey instances are linked to the survey; and data elements (variables) and value domains (classifications) are linked to the data file.

11. The administered items in the current version of the IMDB related to the survey life cycle and covered in this paper are: 1. Statistical activity; 2. Survey; 3. Instance; 4. Universe; 5. Instrument; 6. Methodology; 7. Documentation; and 8. Data Files. These administered items reflect the mandatory requirements for reporting information on data sources, methodology and data accuracy for each survey as stated in the Policy on Informing Users of Data Quality and Methodology.

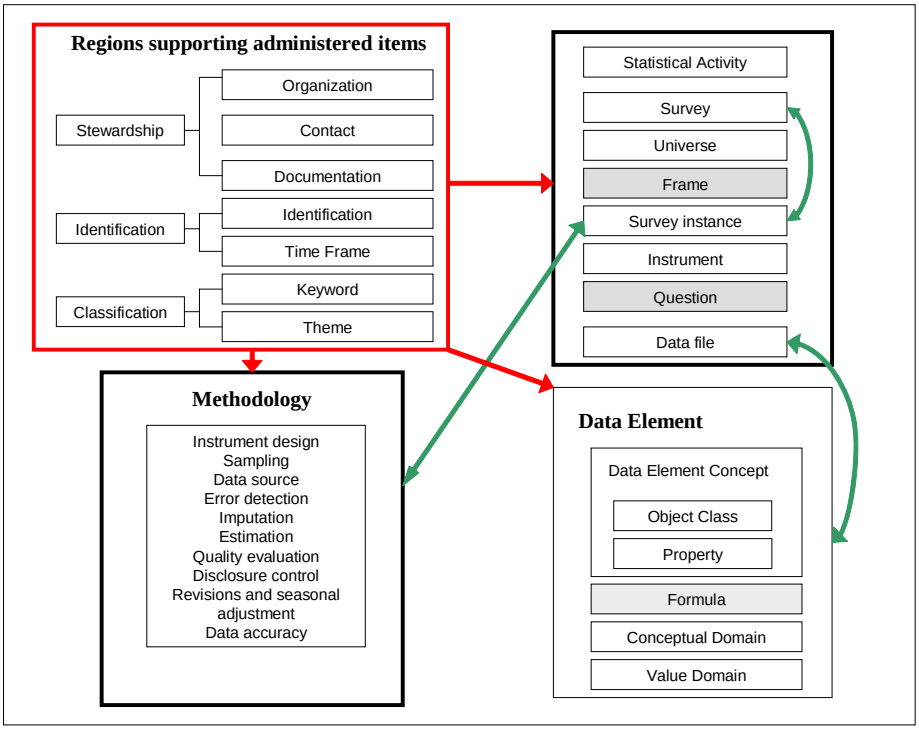
---

[6] Johanis, Paul and Dan Gillman, 2006: Metadata Standards and Their Support of Data Management Needs, Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS), Geneva, April 3-7, 2006.

[7] See www.statcan.ca and in particular, Definitions, data sources and methods module, http://www.statcan.ca/english/concepts/index.htm .

[8] The administered items supporting the survey life cycle in the IMDB match well to the proposed components in Part C of the METIS framework for statistical metadata (see Graeme Oakley, 2006). The IMDB also closely follows the administered items in the Business Dimension Model of the CMR.

Figure 1: Relationship of the survey process to the administered items in the IMDB.



B. Statistical Activity

12. The Statistical Activity administered item in the IMDB represents groups of surveys that share some common features and for which some common explanatory text would be useful to data users. For example, a Statistical Activity record was created for Statistics Canada's Unified Enterprise Survey (UES) Program, which contains general information applicable to 200 separate business surveys into a single master survey program. Another example is the Canadian System of National Accounts.[9] Not every survey needs to be linked to a Statistical Activity and is only created in consultation with subject-matter areas.

C. Survey

13. The content of the IMDB is organized around the survey entity as opposed to datasets as in other metadata models such as SDMX. A survey, in order to be considered as a record in the IMDB, is defined as a statistical activity that involves the collection, compilation and publication of statistical data measuring characteristics of a population. In the IMDB, surveys are defined as three types:[10]

- **Direct**: microdata are collected directly from a respondent with the use of a Statistics Canada collection instrument (e.g., Labour Force Survey);
- **Administrative**: microdata are extracted from administrative sources from an external organization, which were originally collected their own purposes (e.g., vital statistics from provincial and territorial governments); and

[9] http://www.statcan.ca/cgi-bin/imdb/p2SV.pl?Function=getSurvey&SDDS=1735&lang=en&db=IMDB&dbg=f&adm=8&dis=2.

[10] For purposes of presentation, surveys are referred to as "surveys and statistical programs" on the IMDB web pages as a way of representing all three types of surveys.

- **Derived**: data are derived from other Statistics Canada surveys or other data sources to produce datasets of new derived variables (e.g., national accounts, Gross Domestic Product, price indexes).[11]

14. The following guidelines are used to determine whether or not a "statistical activity" is a survey, and therefore requiring a record in the IMDB.

15. Activities producing clean microdata serving as data sources to surveys or to analytical studies, and for which no aggregated data are published, do not constitute surveys for the IMDB purposes. Direct surveys and administrative data can be active, discontinued or be conducted one-time only. Derived surveys can only be active or discontinued. One-time only derived statistics are considered as an analytical study and are therefore considered out of scope for IMDB. A compilation of selected data collected from direct surveys or administrative sources does not constitute a derived survey, even if it is produced on an on-going basis. In general, these statistical compendia, such as Statistics Canada's *Canadian Economic Observer*, are treated as a product and not as a survey.

16. Contrary to direct and administrative surveys, it is difficult to establish clear operational criteria for the designation of derived statistics. The extent of the transformation of the source data to produce new information is the critical factor, which cannot be quantified to establish an absolute rule. In the case of an activity drawing on data collected by others to produce a new dataset, one must ask oneself if referring the users to the metadata in the IMDB on the source surveys will adequately inform them on the quality and methodology of the product. If the answer is yes, the creation of a new derived survey and associated metadata is not called for; if the answer is no, then the statistical activity leading to the product should be designated as a derived survey.

17. The survey administered item contains the following information: the title and acronym of the survey, (i.e., Monthly Survey of Manufacture (MSM)); an overview of the survey that provides a description of the objectives of the survey, the survey population, for whom the data are intended and the use of the data; and the status of activity on the survey (e.g., active, discontinued, transferred or one time only). All surveys are assigned an identification number, known as the Statistical Data Documentation System (SDDS) number in the IMDB. Figure 2 shows the actual attributes as stored in the model.

Figure 2. Attributes of the survey administered item.

Survey	
Survey_AC_Id	INTEGER (PK1)
Survey_AC_Version	NUMBER(5, 2) (PK2)
Survey_AC_Name_en	VARCHAR2(1000)
Survey_AC_Name_fr	VARCHAR2(1000)
SurveyOverview_en	VARCHAR2(4000)
SurveyOverview_fr	VARCHAR2(4000)
Survey_Type	INTEGER
SDDS	INTEGER
Mandatory_Type	INTEGER
Longitudinal_Type	INTEGER
Census_Type	INTEGER
Direct_Type	INTEGER
Derived_Type	INTEGER
Administrative_Type	INTEGER
SurveyPurpose_en	VARCHAR2(1500)
SurveyPurpose_fr	VARCHAR2(1500)

[11] Derived statistics are referred to as "statistical programs" in the IMDB.

C. Other Administered Items

18. A Survey consists the administered items related to the survey life cycle that are grouped together through an Instance administered item for each reference period of the survey. Table 1 shows these administered items and their respective definitions. The indentations of each item in the table illustrate the hierarchical relationship between the entities. These administered items are reused or updated in successive survey instances or shared with other surveys and statistical activities in the generation of web pages. Through the Policy on Informing Users of Data Quality and Methodology, the IMDB has a **common metadata set**[12] that is reused for each survey. It is proposed that the administered items in the Statistics Canada metadata set be considered by other national statistical offices and international organizations as a "best practice" to help move towards a common set of metadata items to be used for various metadata exchange initiatives.

19. On the Statistics Canada website, users have access to metadata for each instance of each survey or statistical program for which data are disseminated. The administered items stored in the IMDB have been organized to present general information on the survey (survey title, status, frequency, record number and survey mandate) and metadata related to the survey life cycle for a survey instance (e.g., reference period, data release date, survey instrument (questionnaire), variables, survey description, data sources, methodology, data accuracy, documentation and data file (available internally only)). Quality metrics such as response rates and coefficients of variation are disseminated under Data Accuracy.

20. Figure 3 presents the web page for a survey instance, in this case, the Annual Survey of Manufactures. On the left hand side bar of the web page, there is additional information including links to Summary of changes over time and Other reference periods. The Summary of changes over time presents a chronology of changes to administered items as well as the start date of the survey and Other reference periods gives users access to metadata for other survey instances for which data have been released. We are currently developing a calendar of surveys in the field, that is, those surveys without disseminated data but with metadata for the some administered items including survey, instance, collection instrument, instrument design and collection method. In the Instance administered item, the time frame entity contains fields for the actual start and end dates for data collection of that survey instance. Using these dates, a list of surveys that are currently collecting data will be created and disseminated on both the Definitions, data sources and methods and Information for survey participants modules.

21. The release of data is announced in the *Daily*, the Statistics Canada's official release bulletin. This announcement triggers the need to update the metadata and create a new instance of the IMDB survey record. Release dates for "mission critical" surveys are posted a year in advance and other releases, two weeks in advance. Based on this information, the manager of the IMDB contacts the survey managers in advance for updated metadata.

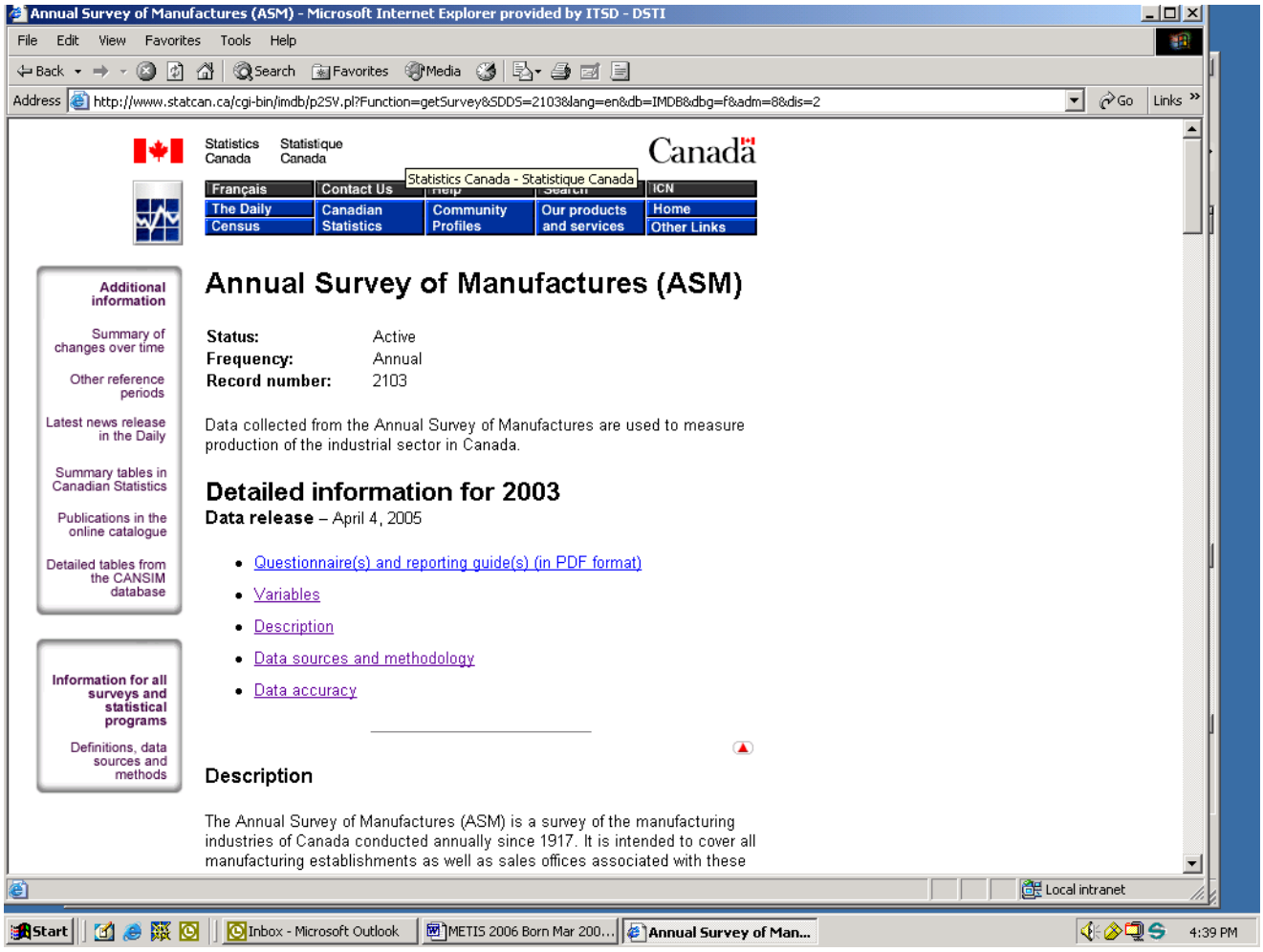
[12] The term metadata set is taken from the SDMX initiative. **Metadata set** is a set of information regarding almost any object that describe the maintainers of the data and structural definitions; describe the schedule on which data is released; describe the flow of a single type of data over time; describe the quality of the data, etc. In SDMX, the creators of reference metadata may take whatever concepts they are concerned with, or obliged to report, and provide a reference metadata set containing that information. Statistical Data and Metadata Exchange Initiative (SDMX), 2005: Framework for SDMX Technical Standards (Version 2.0), November, 2005.

Table 1. Administered items in the IMDB supporting the survey life cycle and their definitions.

IMDB administered items	Definition
Statistical activity	The statistical activity is groups of surveys that share some common processing system or conceptual framework. For example, a statistical activity was created for the Unified Enterprise Survey Program, which contains general information applicable to all surveys in the UES program. Not every survey needs to be linked to a statistical activity. A statistical activity record is created in consultation with subject-matter areas.
Survey	A survey is a statistical activity that involves the collection, compilation and publication of statistical data measuring characteristics of a population. Includes direct surveys, administrative surveys and derived surveys. Provides administrative details about the survey.
Target population (or universe)	The population of units that is actually covered by the survey. Identifies the statistical unit and any of its relevant characteristics that are used (e.g., Canadian population aged 15 and over and not residing in institutions; or, establishments in NAICS industry XXXXXX with revenues over a certain threshold). Where applicable, any differences between the survey population and the target population (i.e., population for which information is desired) of the survey are described.
Survey Instance	Refers to each time the survey process occurs (i.e., each cycle of a survey). For each reference period, a new version of the instance record is created. For example, for a monthly survey, the IMDB will contain one instance record for every monthly cycle of the survey.
Collection instrument	The vehicle used for the collection of data. For direct surveys, it is a questionnaire. For administrative surveys, it is the record layout of the input record. It is not applicable for derived surveys. A questionnaire can be in many forms such as paper version, electronic, etc. Each questionnaire is linked to the instance record to which it pertains. The questionnaire image is copied into IMDB in pdf format.
Methodology	A text description of each of the following aspects of the methodology of a survey.
Instrument design	Method used to design, test and implement survey instrument. It only applies to direct surveys. Description of the design; methods for testing the questionnaire (e.g., review committee, focus group, pilot survey, etc.); and date of last revision of the instrument.
Sampling	Description of the survey units, any stratification used and the sample selection methods. It does not apply to derived surveys.
Collection method	Details on the collection methodology and the type of instrument. Details on method of initial contact and follow-up. Also included is a description of the capture method. For administrative and derived surveys, this item can be used to describe data sources. Collection methods include: data collected directly from respondents with a use of a collection instrument include: electronic data interchange; respondent completed – paper format (mail or fax); respondent completed – touch-tone telephone; Computer Assisted Telephone Interview (CATI); or Computer Assisted Personal Interview (CAPI) methods; or data extraction from administrative files provided by an external organization; or data derived from other Statistics Canada surveys.
Error detection	Methods used to detect errors during collection, capture and processing of the data. Provides details on the types of edits used, ratios applied, etc. and identifies at which stage of the survey process it is done (i.e., during collection, or as part of processing.)
Imputation	Process used to replace missing microdata, and invalid or inconsistent responses identified during editing. Provides details on the type of imputation (e.g., manual, automatic, etc.), imputation rates, the method used (e.g., historical, hot deck, donor, etc.), and software used. Usually does not apply to derived surveys.
Estimation	Methods used to produce estimates for the survey population from collected data. It includes non-response macro adjustments, post stratification,

	calibration, weight-share methods, and variance estimation methods (e.g., direct, Taylor, Jackknife, Bootstrap, etc.). In the case of administrative and derived surveys, procedures and models used to produce the indicators are described.
Quality evaluation	Methods undertaken to evaluate the quality of the final data. Procedures include data confrontation with other published sources, re-interviews, reverse record checks or historical trend analysis.
Disclosure control	Measures taken to ensure that data from the survey does not disclose information concerning any identifiable respondent thus maintaining confidentiality of the respondent data. This summary can include for micro data, removal of respondent, content reduction and content modification; for tabular data, sensitive cells, correction methods such as collapses/suppress cells; and revisions by committees.
Revisions and seasonal adjustments	Methods used to adjust estimates in relation to same estimates for prior periods including benchmarking, calendarization or seasonal adjustments; and procedures for regular revisions to data.
Data accuracy	Data accuracy indicators for the survey. These measures include the coefficient of variation for the key variables in the survey, information on coverage error, response rates and any other relevant data accuracy indicators. Includes response bias and error, and processing errors.
Documentation	Documents useful for the users' understanding of the data can be linked and include user guides, data dictionaries, technical notes, etc.
Data file	Information on the location, format and content of clean data master files that are used as inputs to surveys or as outputs of surveys. The IMDB stores information on clean data master files that are produced for each instance of a survey.
Variables	Description of the meaning of a data point. Based on ISO 11179.
Statistical unit	Definition of the unit about which data are collected (e.g., establishment, household, person and births).
Property	Definition of the characteristic of the statistical unit.
Representation class	Describes the specific form of the representation of the property (e.g., type, name, category, value, area, index)
Classification or unit of measure	A set of allowed values that a variable may take. Classifications are used to represent categorical data and units of measure are used to represent quantitative data (e.g., dollars, tonnes)

Figure 3. Web view of the home page of a survey instance for the Annual Survey of Manufactures (2003).



D. Tools for Loading the Metadata

22. Description text for the administered items is entered into a set of input screens, which is internally referred to as Metastat. These input screens are used to capture information that is common to each of the administered items – identification, description, time frame, documentation, classification (e.g., key words and themes), organization and contact information – previously described as “regions” supporting administered items in Figure 1. Some of these have their own codeset. For example, there are different types of time frames built into the model including: effective period, reference period, collection period, data release and last update. Depending on the administered item, a subset of these time frame types is presented on the input screen (see Figure 5). The description text pertaining to each of the administered items is captured on its own set of input screens. Below are input screens for selected administered items (e.g., Survey, Survey instance (cycle) and Collection method (Data source)) with selected “tabs” (e.g., Identification, Time Frame and Description), respectively (Figures 4, 5and 6).

23. Every system built in Statistics Canada must provide both an English and French interface and editable (data entry) fields for both languages since the metadata published in both languages. When the application is started, the user selects the language of the interface. This allows coded text fields appear in the language of the interface. Every editable field is displayed allowing data entry to be done simultaneously in both English and French.

Figure 4. Input screen for the survey administered item in the IMDB – Identification tab.

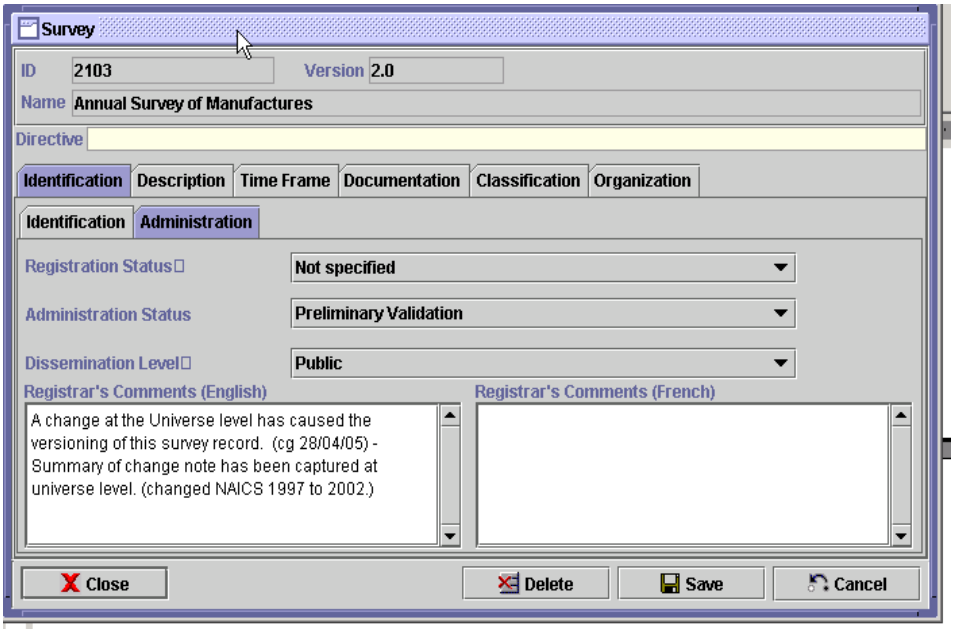


Figure 5. Input screen for survey instance (cycle) administered item – Time Frame tab.

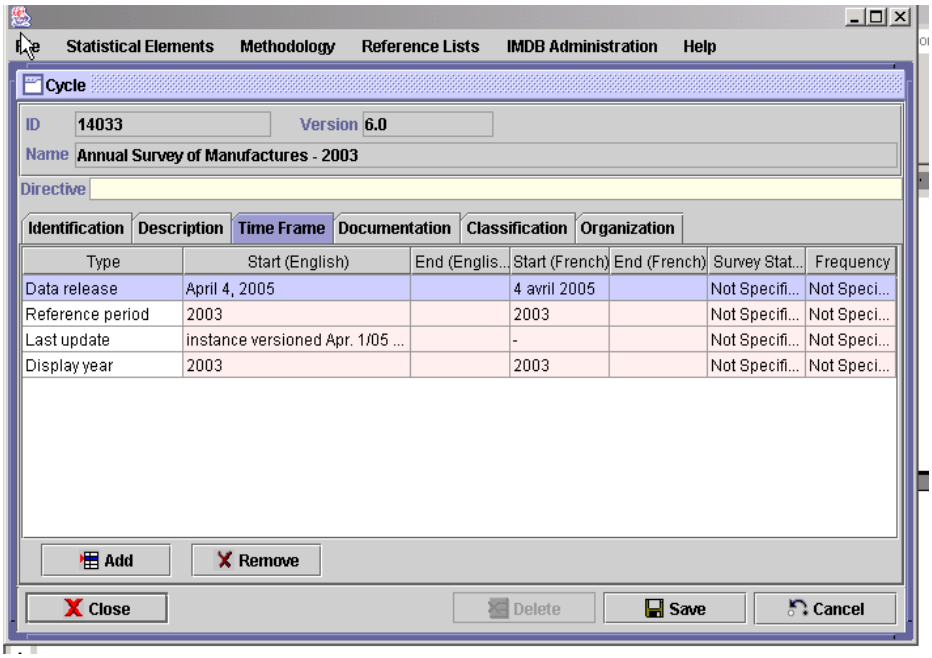
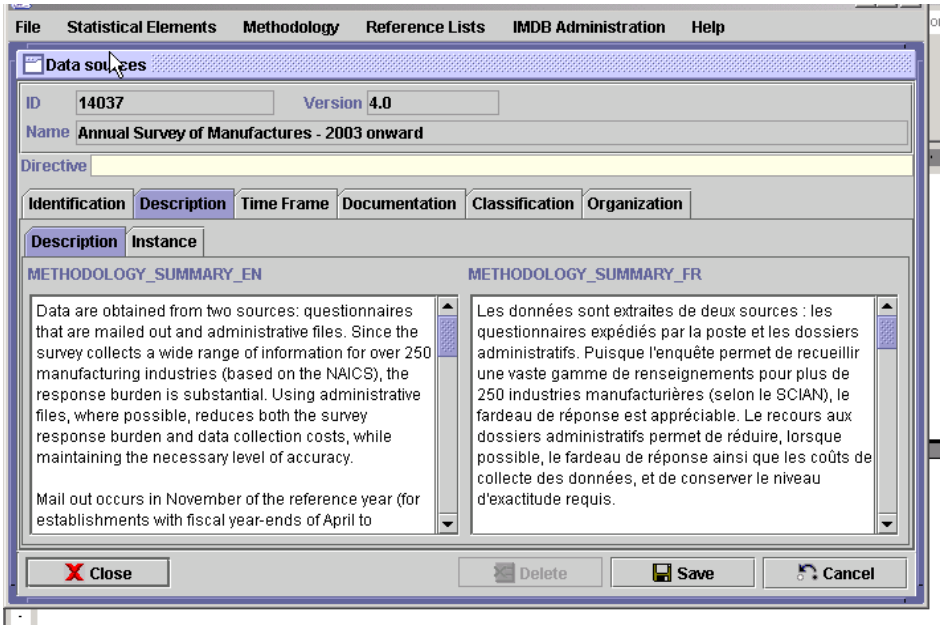


Figure 6. Input screen for Collection method (data source) administered item – Description tab.



III. VERSIONING RULES IN THE IMDB

24. Since metadata do not remain static over time, the metadata model must be able to accommodate these changes. For managing the metadata for each cycle of the survey and changes to the survey through its life cycle, versioning or "time travel" has been built into the IMDB. In this section, the procedures for revisions and versioning of administered items are presented.

25. When the content of an administered item requires revisions, several procedures for implementing them are available in the metadata model. These procedures include:

Create: creation of a new administered item (e.g., new survey);
Update: replacement of prior content with revised content (i.e., for errors in text and addition of more completed text); and
Version: new version of the same administered item while the previous version of the administered item is stored in the IMDB.

26. Although the Create and Update functions are relatively self-explanatory, the versioning function consists of different business rules for each administered item. Versioning is used when there is a need to retain changes over time for each item.

27. At the time an administered item record is created in the IMDB, the system automatically assigns an administered item identification (ID) and a version number (i.e., AC_ID 123 AC_Version 1.0 becomes AC_ID 123 AC_Version 2.0). While the administered item ID remains unchanged for the entire life cycle of the administered item, the version number changes each time the item is versioned. Versioning allows the descriptive text or elements of the administered item to be revised without overwriting the previous version of the metadata. Business rules for versioning of an administered item (in bold in Figure 7) have been developed for specific to the types of changes. Below is a more detailed description of the versioning rules.

A. Versioning of a Survey

28. Changes to the characteristics of the survey (e.g., name, status, frequency, type, statistical activity, Collection Registration Number (CRN), theme assignment and divisional assignment) result in the creation of a new version of the survey administered item in the IMDB. The only exception is changes to the survey

objective. A change in the survey objectives (i.e., Survey Purpose attribute) results in the creation of a new survey (i.e., assignment of a new SDDS number in the IMDB). In Metastat, the reason for the versioning of the survey is explained in the Version Revision Description field of the new version under the Identification tab. For example, in the case of the change in survey frequency, the note would state: *"In the past this survey was conducted on a monthly basis. It became an annual survey as of January 2004 because…..…"* The changes are recorded as part of the Summary of changes over time for the survey on the web version of the record.

B. Versioning of a Survey Instance

29. The only reason for versioning an instance record is a change in the reference period, which is generally triggered by the release of new data to the public. The instance record is versioned for each cycle of a survey and the links are established to the pertinent administered items that need to be updated for that reference period. An instance record is not versioned for reference periods during which there is no survey activity. The reason for the survey's inactivity is added to the Version Revision Description field under the Identification tab of next active instance. In cases where the survey was conducted but the data were not released, an instance record representing the reference period is still created in the IMDB. Text indicating that no data were released by Statistics Canada is displayed in the Data Release field.

30. For a monthly survey, the IMDB will contain one instance record for every monthly cycle of the survey, for an annual survey, one instance for every annual cycle, and so on. For each reference period, a new version of the instance record is created. This enables unchanged information to be carried over from one reference period to the next and only changed information needs to be updated.

31. Every instance record has various administered items linked to it (Figure 7). These include the collection instruments, the various methodology items and the data file. Some administered items may vary from one period to another, while others may remain stable. The new information in these changed administered items is captured separately and then linked to the instance record. For example, an instrument record is versioned each time a new image is produced even if the change is only the date printed on the questionnaire image. The instance administered item thus becomes a central location of all links to each administered item and provides the full picture of the entire survey process for each reference period. In general, changes to instrument, methodology and data files are released at the same time as a new version of the survey instance is released since changes are all triggered by the release of new data to the public.

C. Versioning the Target Population

32. A change to the target population of a survey results in the creation of a new version of this administered item as well as a new version of the survey. This also includes changes in the statistical unit and classifications. For example, under the Summary of Changes over time for the Annual Survey of Manufactures, the following is presented:

Target population – Prior to reference year 2000, the Annual Survey of Manufactures (ASM) provided estimates of principal financial statistics for all incorporated manufacturing businesses that had employees and had sales of manufactured goods equal to or greater than $30,000. With reference year 2000, the universe was expanded to cover all manufacturing units. This change added approximately 60,000 units to the ASM universe.

D. Versioning of a Methodological Administered Item

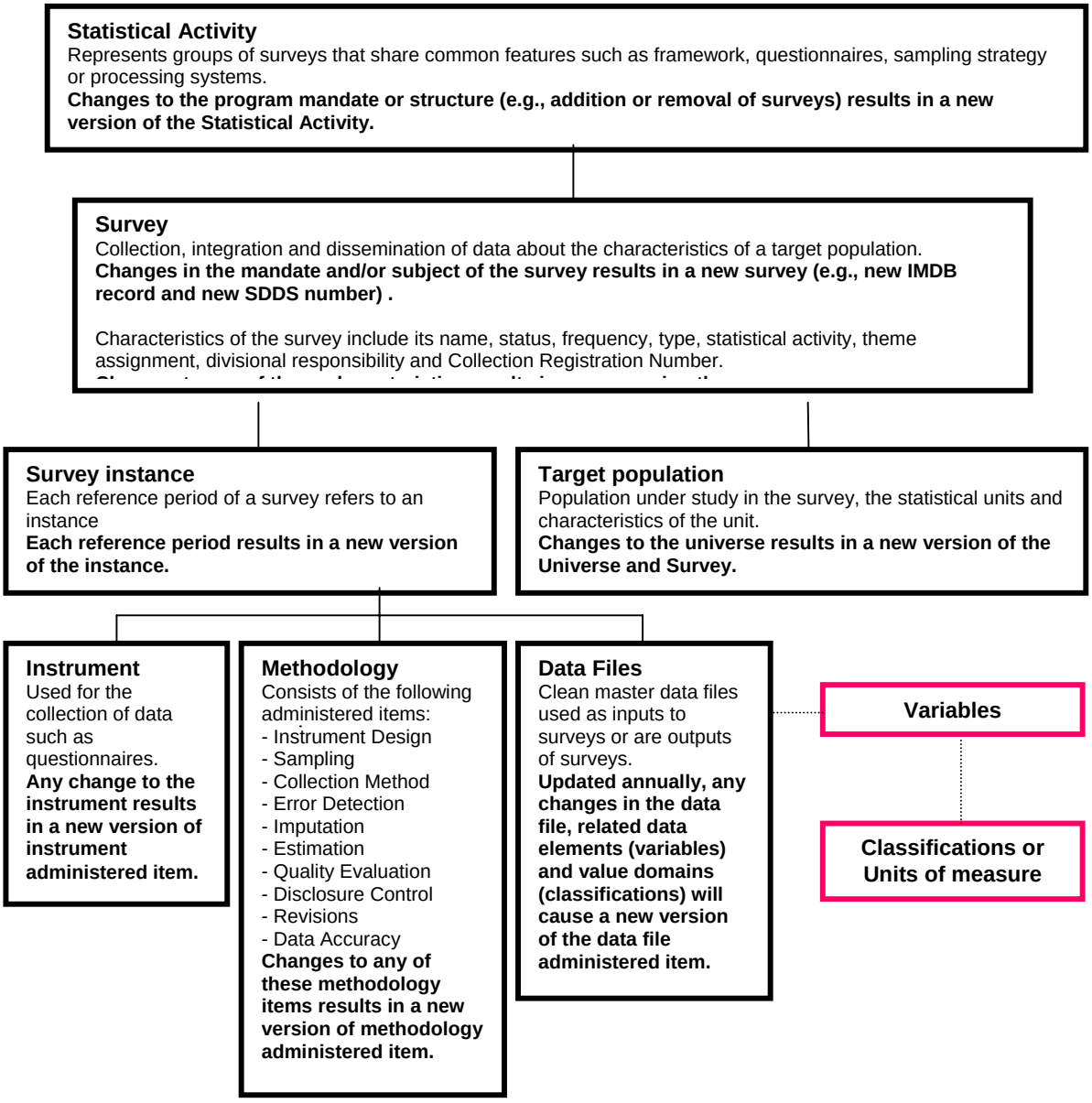
33. A methodology record is versioned only when there is a change of method. Improvement in the text does not constitute a new version of the methodology. For some items of the methodology group, the descriptive text contains reference to specific dates and/or reference periods. For these cases, the methodology records require versioning. In many cases, the sample size is reference-period specific, and the Sampling record is versioned for that reference period. Specific collection dates are also related to a specific reference period and Data sources, which contains these dates, is versioned that survey instance. Data accuracy is another

administered item that is often versioned for every instance of the survey since it contains measures that are reference-period specific.

E. Versioning of a Statistical Activity

34. Versioning of a Statistical Activity occurs when there is a change in the program mandate and/or structure, including the addition of surveys from other survey programs. Any changes to the description of the Statistical Activity or in the set of surveys that are linked to it will cause the creation of a new version of the Statistical Activity record.

Figure 7. Versioning rules for the administered items in the IMDB that support the survey life cycle.



IV. ADDITIONAL ADMINISTERED ITEMS AND FUTURE WORK

35. The Business Dimension Model of the CMR supports a number of additional administered items, which complete the necessary information to describe the survey life cycle. These are not directly stored in the IMDB but are connected through external links.[13]

A. Systems

36. In the CMR, there is an administered item for Systems, which provides metadata for both hardware and software systems. At Statistics Canada, metadata related to this administered item is stored in the Agency's Software Register (SR). The SR is a list of all application systems in use at Statistics Canada and the list of software products which are used in these applications. Although not directly linked to the IMDB at this time, the SR contains the survey identification number (i.e., Statistical Data Documentation System (SDDS) number) and the functional responsibility for managing the survey (i.e., FRC code) stored in the IMDB. However, it would be more useful to link the Agency's SR to the SDDS and its program element (PE) structure, which relates the Agency's expenditures and budget by programs or projects. This integration of the IMDB, the SR and corporate financial planning systems will permit analysis of application development and software diversity across the Agency and identify which surveys are dependent on what applications. The analysis can be used to monitor development (which are capitalized) and maintenance costs of applications; promote better coherence and reuse of software components across surveys and statistical programs; and identify surveys or sets of surveys that may be vulnerable because of dependencies on applications that are beyond their expected life or software that is no longer supported.

B. Products

37. The CMR also contains an administered item class called Products, which is linked to Data Sets. At Statistics Canada, this item is managed by the Common Object Repository (COR), which is external to the IMDB but contains links to the IMDB throughout the Statistics Canada website. Figure 8 illustrates the relationships between the metadata and the various disseminated products. For example, there is a direct link from the data release in the Daily to the survey instance (Figure 9). Surveys and attached metadata, and questionnaires are also organized by the subjects, thereby enhancing accessibility of the metadata for the user (Figure 10).

C. Corporate Planning and Post-survey Evaluation

38. With the IMDB, information on surveys can be reported by FRC (division), by subject, or other dimensions such funding source (cost recovery/base budget). As part of the further enhancement of the IMDB, links to corporate reporting systems are being evaluated particularly to support evaluation and audit of statistical programs.

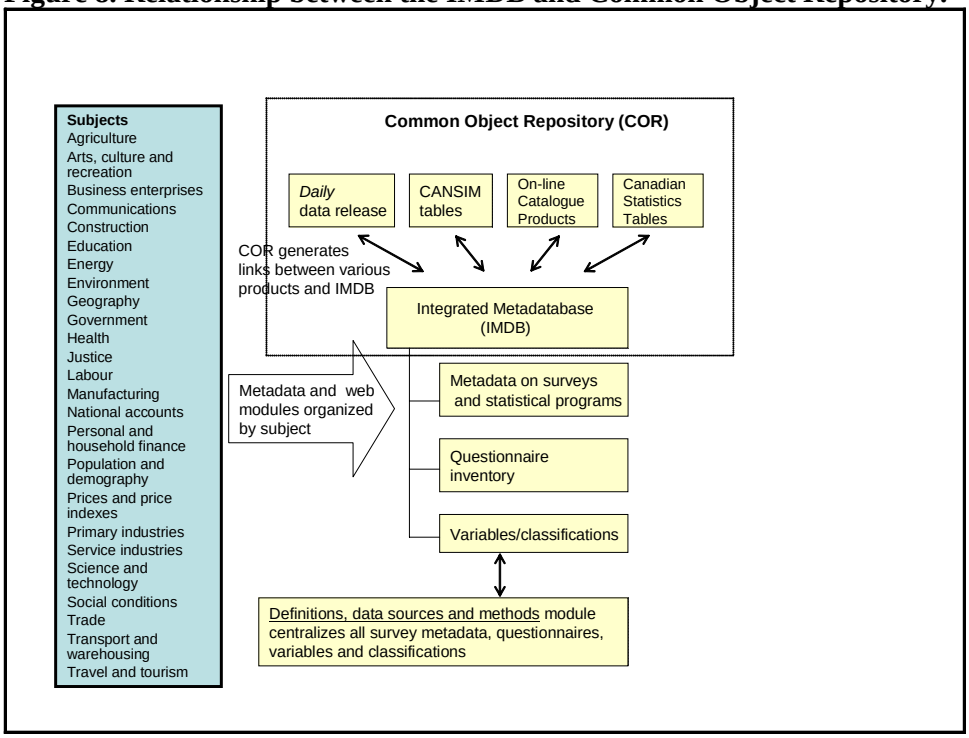
39. Statistics Canada is currently developing a systematic approach to conducting a post-survey evaluation. Referred to as the Quality Management Assessment (QMA), managers responsible for the delivery of a survey or statistical program will provide a description of the data quality management tools that are used in the various processes (e.g., questionnaire testing, interviewer monitoring); an assessment of the impact of changes (i.e., as the result of a survey redesign) to these processes on quality, costs, response burden and confidentiality; and a description of how accuracy is managed (i.e., how sampling and non-sampling errors are prevented, identified, corrected and evaluated). The QMA will be developed within the Agency's Quality Assurance Framework.[14] It is anticipated that the results of the QMA will provide valuable input to divisional quadrennial

[13] See Figure 2. Business Dimension Model in Working Paper 7 (Paul Johanis and Dan Gillman, 2006).
[14] http://www.statcan.ca/bsolc/english/bsolc?catno=12-586-X&CHROPG=1

and biennial program reviews and identify the need to add administered items and quality indicators in the IMDB to meet the needs for reporting quality management practices in the Agency.[15]

40. As part of the QMA project, there will be links to the relevant metadata on data sources, methods and data accuracy stored in the IMDB. Here we can use the IMDB to analyze and evaluate the practices across surveys and statistical programs. For example, we could produce information on how many and what surveys have moved from X-11-ARIMA to X-12-ARIMA for seasonal adjustment, or on the use of generalized systems for imputation or estimation. Other measures of data quality can be assessed across surveys and over time. They may include time lapsed from data collection to data release (timeliness), time required for respondents to complete a questionnaire, response rates, and refusal rates. All of this information is stored in the IMDB.

Figure 8. Relationship between the IMDB and Common Object Repository.



[15] Julien, Claude, 2006: Quality Assessment Management at Statistics Canada, European Conference on Quality and Methodology, Q2006.

Figure 9. Link in the Daily to the survey instance.

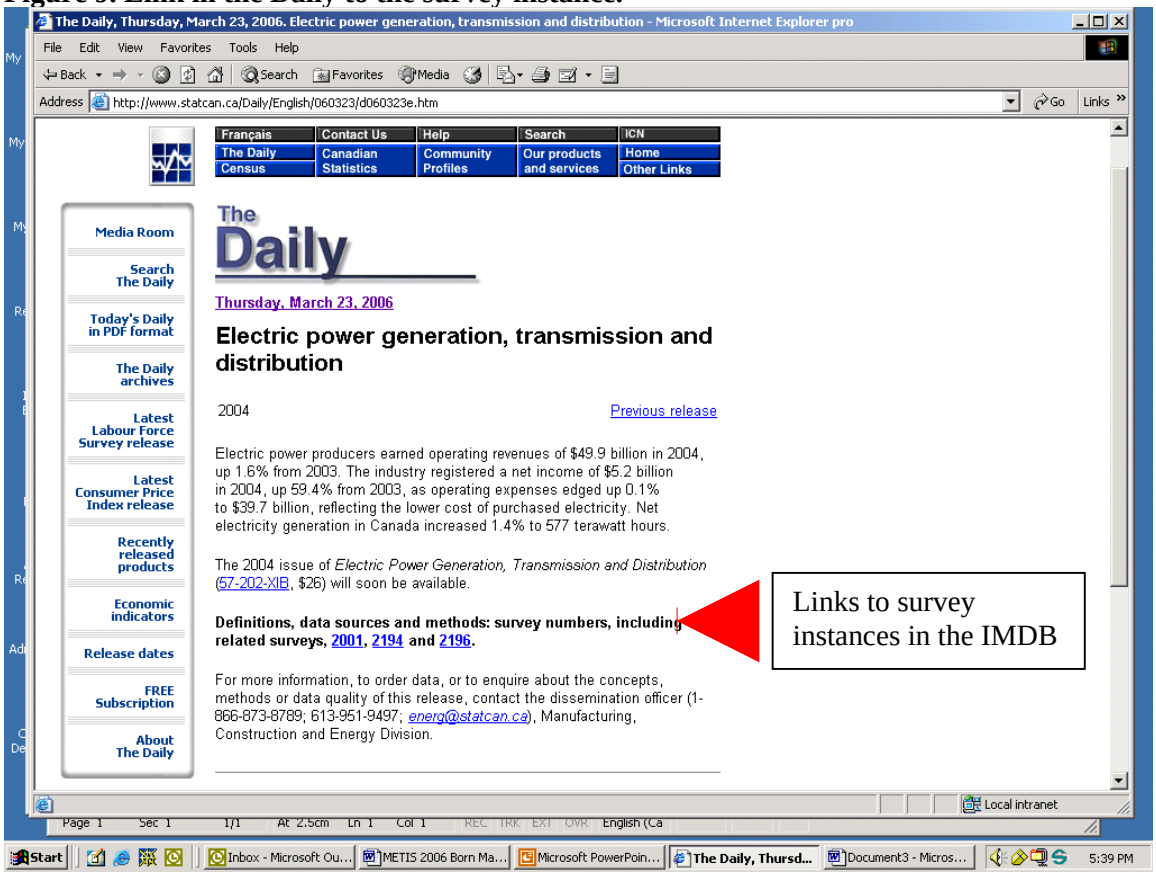
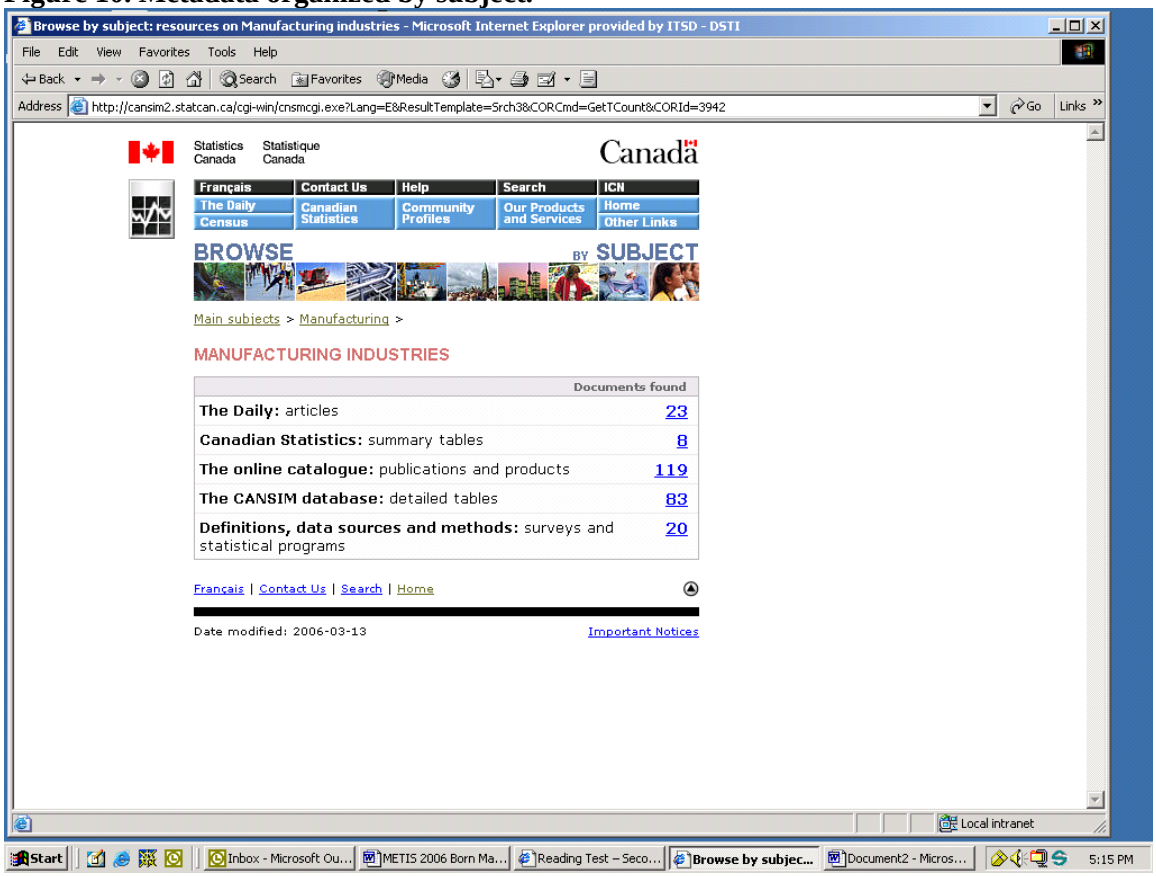


Figure 10. Metadata organized by subject.



V. CONCLUSIONS

41. The IMDB is becoming the single source of metadata for describing Statistics Canada's surveys and statistical programs. This means that survey managers have to supply metadata on their individual surveys only once to the stewards of the metadata, Standards Division. Since IMDB is built on a common metadata set with reusable administered items and attributes, survey managers can reuse the descriptions for different survey cycles and across other surveys they might manage. Also, the use of a common metadata set presents a "common look and feel" to data users accessing the metadata through our website.

42. While the most of the content in the IMDB was determined by the Policy on Informing Users of Data Quality and Methodology, both internal and external users have indicated other requirements when it comes to statistical metadata. The IMDB has been designed and continues to be developed to meet these needs. Now that the metadata is complete for each survey, other users can access those administered items that meet their requirements. In additional to supporting the information requirements for disseminated data, the IMDB is being used as a source of information for standardizing survey processes and content, corporate and financial planning, quality management at the survey level, survey respondents, international data exchanges and data researchers.

43. The purpose of this paper was to provide an example of a set of statistical metadata that supports the survey life cycle as a contribution to the METIS manual on statistical metadata, currently in development. Hopefully, this has been achieved.

VI. REFERENCES

Johanis, Paul and Dan Gillman, 2006: Metadata Standards and their Support of Data Management Needs, Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata (METIS), Geneva, April 3 to 5, 2006.

Johanis, Paul, 2005: Documenting Data Elements in Statistical Agencies, JSM 2005, Minneapolis, August 7, 2005.

Johanis, Paul, 2000: Statistics Canada's Integrated Metadatabase: Current Status and Future Plans, Work Session on METIS, November 28-30 2000, Washington, D.C., United Nations Statistical Commission and Economic Commission for Europe Conference of European Statisticians.

Julien, Claude, 2006: Quality Management Assessment at Statistics Canada, European Conference on Quality in Survey Statistics, Q2006, ……

Lee, Amie, 2003: Integrated Metadatabase (IMDB) Architecture, Statistics Canada, Ottawa, November 5, 2003.