1 June 2007

WP.11
ENGLISH ONLY

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

**UNECE Workshop on the Common Metadata Framework**
(Vienna, Austria, 4-6 July 2007)

# CASE STUDY - CANADA[1]

## 1. INTRODUCTION

### Organization Details:

1. The Integrated Metadatabase (IMDB) is Statistics Canada's (STC) corporate metadatabase initially put in place to allow proper interpretability of the data released to public.

2. The development and maintenance of the IMDB is the responsibility of Standards Division, which is part of the Informatics and Methodology Field (IMF). One section of the division, the Corporate Metadata Section (CMS), is dedicated to the IMDB activities; three other sections of the division look after the development and maintenance of the Agency's standards (i.e., statistical units, concepts, variables) and classifications.

3. The IMDB architectural design and systems development is provided by three officers of the Systems Development Division, which is also part of IMF. The chief of the CMS directs the work of three analysts responsible for the development of the content of the base, four officers responsible for the maintenance and update of the content, and one database administrator.

### Overall strategy and metadata management principles:

4. The IMDB finds its justification in Statistics Canada's *Quality Assurance Framework* (QAF) and in its *Policy on Informing Users on Data Quality and Methodology* (PIUDQM).[2] The QAF states that there are six dimensions to data quality: relevance, accuracy, timeliness, accessibility, interpretability and coherence. The PIUDQM addresses the *interpretability* dimension, which requires all statistical outputs to contain, or be linked to information on concepts and definitions, on methodology used to produce the data, and on data accuracy. The IMDB, whose development in phases started in 1999, is the primary vehicle for implementing the policy.

5. So far, for each of the 350 on-going surveys and statistical programs as well as for about 250 inactive ones (discontinued or one-time-only), the IMDB contains information on the purpose of the survey or statistical program, the methodology used to produce the data, the accuracy of the data and clean master datafile. It took a few years for survey and statistical program managers to incorporate in their data production cycle the provision of complete and up-to-date metadata on survey methodology and data accuracy. The main stimulus comes from division directors having to report, in their biennial report to the Chief Statistician, on the level of compliance of their IMDB survey record with the specifications of the PIUDQM.

[1] Prepared by Pierre-Paul Bellerose, Tim Dunstan, Carmen Greenough, Amie Lee and Alice Born of Statistics Canada.
[2] Definitions, data sources and methods

6. Completing the documentation of the variables, for which data are released, and their associated classifications is the next priority for the IMDB. The implementation of the metadata model for the documentation of variables is more complex than for the description of survey methodologies and the discussion of data accuracy, and therefore will take longer to complete.

**Lessons learned:**

7. The progress made during the variable documentation phase, as well as with the methodology and data accuracy documentation phase, leads us to conclude that it is more efficient to start documenting the metadata right at the outset of any new survey design or redesign, and at the data collection stage of surveys, instead of documenting them only after they have released data, which was our practice in the past.

## 2. THE STATISTICAL METADATA SYSTEMS AND THE STATISTICAL CYCLE

### How the IMDB fits into other organizational systems

8. One of the primary objectives of the IMDB is to inform users on concepts, methodologies and data accuracy. The IMDB provides the metadata to support, and that are linked, to the statistical products released by STC's Dissemination Division and the Communications and Library Services Division. These divisions are responsible for STC web site and most of its information modules, especially *The Daily* (STC news release bulletin), CANSIM (STC public database), STC Publications, Statistics by subject portals, and the Summary tables (close to 400 free data tables on many aspects of Canada's economy, land, people and government).

9. The responsibility for IMDB and its related *Definitions, data sources and method*s module on STC web site was given to Standards Division because it was designed as a means to enforce more coherence in STC data over the long term. *Coherence* is another dimension of STC's QAF. The coherence of statistical information reflects the degree to which it can be successfully brought together with other statistical information within a broad analytical framework, and over time. The use of standard concepts, classifications and target populations promotes coherence, as does the use of common methodology across surveys.

10. This long-term objective of enforcing standardization called for a metadatabase that would store information broken down into building blocks prescribed by a metadata model, where building blocks would be shareable among and linked to all surveys and statistical programs and for any reference period. Where the content of these building blocks evolves over time, storing and linking appropriately versions of them is required. As well, if the metadatabase is to play its role relative to coherence, it has to possess a rigorous registration process to track the validation steps of the information being stored. These are the functions performed by the IMDB.

### Overview of the IMDB metadata model

a) Defining concepts and variables

11. At the heart of any survey or statistical program, and throughout their lifecycle, are the concepts and variables one wants to measure. To promote coherence, these concepts and variables have to be standardized, as well as the questions used to measure them. For standardizing the naming of the variables, the IMDB applies a modified version of the ISO/IEC[3] 11179 *Metadata registries* standard. The basic components of the standard have been retained: object class, property, data element concept, data element and value domain. However, terminology that is more familiar to statistical users is used. The data element concept is known as a concept in Statistics Canada; the object class is known as a statistical unit or unit of observation; the property is known as characteristic; the data element is a variable; and the value domain of a variable corresponds to units of measure or classifications depending if the set of values that the variable can assume is continuous or not.

[3] International Organisation for Standardization (ISO) and International Electrotechnical Commission (IEC).

12. According to the metadata model used by the IMDB, a variable (data element) is comprised of a concept (data element concept) and a representation; the concept itself being comprised of a statistical unit (object class) and a property. A statistical unit is an agent (e.g., person), an event (e.g., birth) or an item (e.g., product) about which data are produced. A property is the characteristic being measured of the statistical unit considered. The representation is the form given to the resulting data (e.g., quantity, value, type or category). All three elements are used to create the name of the variables (e.g., value of sales of an establishment, type of occupation of a person, name of geographic location of a person).

13. The definition of a variable is provided by the joined definitions of its constituting components (i.e., the definition of the statistical unit involved and the property measured). The complete definition of a variable also requires the specification of the unit of measure or the classification used to depict the variable (i.e., the non-enumerated or enumerated value domain of the variable as it is known in the model).

14. Non-enumerated value domains correspond to continuous variables (i.e., variables that can assume any value in a set of rational numbers). Examples of such value domains are quantity and value. The sets of rational numbers are organized into units of measures, e.g. volume in cubic meters, Canadian Dollars, counts in years. In the case of statistical data disseminated through CANSIM in the form of continuous variables, data users need to know the unit of measure used to quantify the variable. The name of every unit of measure associated with CANSIM variables will be stored in IMDB.

15. The enumerated value domains correspond to categorical variables, for example: category of sex of person, name of geographic location of household, or type of industry of establishment. Often, data that are collected as continuous variables are converted to categorical variables for dissemination purposes. In statistical agencies, the sets of such categories are organized into classifications, standard or otherwise. Classifications are arrangements of classes of descriptors that are mutually exclusive, exhaustive in the universe of interest, and can be aggregated into successive levels in a hierarchy. In most cases, many classifications may be formed for the same categorical variable. The variable "type of industry of establishment", for example, has many classifications, including ISIC, NACE and NAICS. In the case of statistical data disseminated according to given classifications, data users need to know the meaning of each category or class within a classification and the hierarchical relationship between them. In IMDB, every classification will be named, the code and name of each of its constituent classes will be stored and the level and position of each in a hierarchy will be described.

16. To support the Survey planning and design phase of the survey life cycle, questions used in collection instruments and the associated choices of responses can be stored in the IMDB where they are linked to the related concepts, variables and value domains discussed above.

17. See, under section 6, a 17 slides PowerPoint presentation titled "Representing variables according to the ISO/IEC standard".

18. The loading of the variable components in IMDB is subjected to a registration process controlling authorization to use functions and assigning status levels to components relative to validation of the information and conformity to standards. This is illustrated by a 24 slides PowerPoint presentation titled "IMDB Registration of Survey Variables"; the link is provided under section 6.

**Describing the survey methodology steps and data accuracy**

19. To store information about the purpose of the survey or statistical program, survey methodology steps, data accuracy, and the IMDB uses a certain number of building blocks corresponding to phases of a statistical cycle. The content of these building blocks has a start and end date so it can be versioned over time; the descriptive texts are to be summaries not exceeding 4,000 characters, however, unlimited additional documentation may be attached to each version of each of the building blocks in order to provide more detailed information. The latter feature makes it possible for survey areas to use the IMDB to archive all the documentation related to a survey. Any piece of information stored in IMDB can be set for 'public' or

'internal only' display. Used that way, the IMDB also serves as a mechanism for "knowledge transfer" across generations within the agency.

20. The building blocks of a first layer are labelled: Statistical activity, Survey, Universe (Target population), and Instance; the latter is linked to a second layer of ten building blocks labelled: Instrument design, Sample design, Collection (Data sources), Error detection, Imputation, Estimation, Quality evaluation, Time series (Revisions and Seasonal adjustment), Confidentiality (Disclosure control), and Quality measure (Data accuracy). (The labels in parentheses correspond to the ones appearing on the public web pages describing the surveys.)

21. A 'Survey' is determined by its mandate, which is to measure certain characteristics of a target population or 'Universe'. Surveys can be grouped into 'Statistical activities' corresponding to study domains. The information stored in a specific version of a 'Survey' component –essentially: mandate, tombstone information, plus overview- applies to all 'Instances' the specific version of the survey is conducted. The frequency of instances varies from weekly to one-time-only. Each instance is linked to the applicable version of the applicable methodology step descriptions listed above as making up the second layer of building blocks. Again, the same building block content is stored only once in IMDB and it is shared through time and surveys wherever applicable. This is the feature that over time will enable standardization and hence higher level of data coherence.

**Relationship between the IMDB and other STC meta-information systems**

22. Currently, the IMDB intervenes at the analysis, dissemination, and post-survey evaluation phases of the statistical cycle. However, plans are underway to use the IMDB for the survey planning and design, and archiving phases of the survey cycle.

23. At the survey planning and design phase, IMDB can be consulted and used by survey managers for identifying existing variables for reuse and documenting new variables during questionnaire design; to store these variables and the related questions, which are then outputted to collection processes managed by other systems.

24. In the dissemination phase, IMDB is the source of summary texts describing surveys, their methodology, the data quality of the data produced and the definition of the variables they measure, as well as of the images of the questionnaires they used as data source. The metadata in the IMDB will also be part STC's archived datafiles and will be part of STC's Quality Management Assessment of its surveys and statistical programs.

- **Relationship with survey planning and design systems**

25. The IMDB has recently expanded its role as part of the Agency's Household Survey Content Harmonization (HSCH). The objective of this project is to define standard concepts and variables to be measured by new or redesigned surveys, as well as to create standard questions and standard question blocks to collect the data necessary to measure them. The naming of the variables is being done jointly by the IMDB and HSCH staff according to the IMDB naming convention, and the documentation of the related pieces of information as well as the related questions are being stored in the IMDB. These pieces of information will be outputted to a collection process system being developed by another division, Operational Research and Development.

- **Relationship with public information dissemination systems**

26. Utilisation of IMDB content by other public information dissemination systems can be categorized into four groups.

i) The main information modules accessible on STC web site containing statistical data

27. They include The Daily, CANSIM, Summary tables, STC Publications and Statistics by subject portals. STC has policies requiring a) that the data release must first be announced in The Daily, and b) that the CANSIM

tables and The Daily articles offer a link to the related IMDB survey records describing the survey methodology, the variables estimated and data accuracy. In addition, the Summary tables, the Publications and Statistics by subject portals provide links to the related IMDB survey records. As a result of these linkages, users finding a relevant product in any one module can easily and quickly find related products and associated metadata.

ii) The *Information for survey participants* module on STC web site

28. The module links to the IMDB to display the images of questionnaires as well the module is encouraged to re-use some of the surveys descriptive texts stored in IMDB.

iii) The Data Liberation Initiative (DLI)

29. For a minimal annual subscription fee, STC makes available to universities and colleges all its statistical products available in electronic form including Public Use Microdata Files (PUMFs). DLI uses the DDI standard to provide the documentation related to the PUMFs. The metadata already existing in IMDB are re-used instead of being re-created.

iv) Publications appendices on methodology and data quality

30. Authors are encouraged to re-use the content of the related IMDB survey records for these publication sections. More and more authors are doing so.

31. The metainformation associated with the phases between *Survey planning and design* phase and *Dissemination* phase are managed by various systems disseminated throughout the Agency. At that level, IMDB serves as the source for variables documentation and the classifications associated with them. This is particularly the case of datawarehouses, which support the *Aggregation / Analysis* phases of the survey life cycle. Datawarehouses are being developed in the major statistical areas: system of national accounts (SNA), education and health. The most advanced project is the SNA one whose architecture includes links with the IMDB.

32. Finally, the potential exists for metadata stored in the IMDB to be outputted to other metainformation systems in any output format that suit these other systems (i.e., SDMX, DDI, HTML, Wiki). This is illustrated by the PowerPoint slide titled Meta data outputs located under section 6.

## 3. STATISTICAL METADATA FOR THE PHASES IN THE STATISTICAL CYCLE

### Dissemination Phase

### Metainformation describing the statistical business processes

33. IMDB was designed to initially support data dissemination, that is, to provide users with the information they need to interpret the statistical data disseminated by Statistics Canada. As a result, the content of the database is largely driven by the requirements of the *Policy on Informing Users of Data Quality and Methodology*.

| Group | Description | Examples | Source (if reused) | Quality issues | Risks & challenges |
|---|---|---|---|---|---|
| Survey | A record is created in IMDB as soon a survey is prescribed by the Chief Statistician to be collected under the authority of the *Statistics Act*.<br><br>Survey characteristics and administrative details that remain unchanged from one cycle to the next. | Title and purpose of the survey, description of uses and users of the survey data, mandatory/voluntary reporting, type of data sources and collection methods, Division responsible and survey time frames, and target population. | Subject-matter area | Completeness, consistency and coherence with other on-line products | Subject-matter areas are not always cooperative in supplying the documentation required. |
| Methodology | Summary of methods used throughout the survey process. | Instrument design, Sampling, Data sources, Error detection, Imputation, quality evaluation, disclosure control, revisions and seasonal adjustments, Data Accuracy | Other IMDB records | Completeness, consistency and coherence with other on-line products | Subject-matter areas are not always cooperative in supplying the documentation required. |
| Questionnaires | Instrument image in PDF format | Electronic copies of forms mailed to respondents | Can be assigned to several surveys; reused over time | Conversion of CAPI/CATI scripts<br><br>Large numbers to track | Timeliness – ensure the images are publicly available on-line while in collection |

### Metadata for data elements

34. *Phase III* of the IMDB is presently collecting and outputting variable (data elements) level metadata in support of the Dissemination as well as the Survey Planning and Design phase of the statistical life cycle. In all cases, metadata objects are tracked over time for changes in content (versioning) and validity (registration). This metadata is outputted in CANSIM tables when ready for public release, and on Standards Division's webpages STC's internal WIKI for internal review.

| Group | Description | Examples | Source (if reused) | Quality issues | Risks & challenges |
|---|---|---|---|---|---|
| Variables | A variable is presently collected by a metadata analyst subsequent to analysing a CANSIM table in which the variable appears. The variable is linked to the appropriate surveys and tables, and made available to users on our internal webpages. From here, the variable is viewed by subject-matter and further analysis ensues. The validated variable is made available on the STC website linked to the appropriate survey and on the CANSIM site, linked to the appropriate table. | Type of Industry of Establishment,<br><br>Name of Geographic Location of Establishment | Subject-matter experts<br><br>Subject-matter documentation<br><br>STC-wide standard concepts and variables. | Completeness, consistency and coherence with other on-line products and across survey programs.<br><br>Reuse of standard concepts and variables | Understanding ISO 11179.<br><br>Volume of work.<br><br>Subject-matter areas are not always able to supply the documentation required. |
| Classifications | Documentation of the table dimensions in CANSIM in the form of classifications in the IMDB. The classification is mapped to the members in the CANSIM table. The classification is then linked to the appropriate variables, surveys and tables, and the information made available to users on our internal Web pages. From here the classification is reviewed by subject-matter and further analysis ensues. The validated version of the classification is documented and made available on the STC website linked to the appropriate survey and on the CANSIM site linked to the appropriate table. | North American Industry Classification 2002; Standard Geographic Classification 2001 | Subject-matter experts<br><br>Subject-matter documentation<br><br>STC standard classifications | | Understanding ISO 11179 standard<br><br>Number of variables to validate<br><br>Subject-matter areas are not always able to supply the documentation required<br><br>Dealing with high number of variant (non-standard) classifications in use |

**Survey planning and design:**

35. The metadata system to support *Survey planning and design* is currently being developed in the IMDB. It is represented by the Questionnaire model of the Corporate Metadata Repository (CMR) model.

| Group | Description | Examples | Source (if reused) | Quality issues | Risks & challenges |
|---|---|---|---|---|---|
| Questions | Questions will be collected for questions at the *Survey planning and development* phase and tracked through their development into *Standard Question Blocks* and *Standard Question Response sets.* They will be mapped to classifications in the IMDB and linked to variables and concepts as well as surveys. They will be assembled into collection instruments in another metainformation system outside the IMDB. | N/A | Subject-matter experts involved in developing survey design and development<br><br>STC standard questions | Completeness, consistency and coherence with other on-line products and across survey programs.<br><br>Reuse of standard question blocks and questions<br><br>Appropriate linkages | Understanding ISO 11179 standard<br><br>Volume of work.<br><br>An initial reluctance to input and manage metadata in support of questionnaire design in the IMDB by survey development teams |
| Response sets | Response sets will be collected for questions at the *Survey planning and design* phase and tracked through their development into *Standard Question Blocks* and *Standard Question Response sets.* They will be mapped to classifications in the IMDB and linked to variables and concepts as well as surveys. They will be assembled into collection instruments in another metainformation system outside the IMDB. | N/A | Subject-matter experts involved in survey design and development<br><br>STC standard classifications and response sets | Completeness, consistency and coherence with other on-line products and across survey programs.<br><br>Reuse of standard classifications as mapped to response sets<br><br>Appropriate linkages | Understanding ISO 11179 standard<br><br>Volume of work.<br><br>An initial reluctance to input and manage metadata in support of collection in the IMDB by survey development teams |

**Other Statistical Cycle Phases**

36. Metadata or paradata for data processing for Phases 2 to 5 of the statistical cycle (Survey preparation; Data collection; Input processing; Derivation, estimation, aggregation; and Analysis) are not part of the IMDB metadata model. These metadata systems are currently separate from the IMDB. For example, a centralized processing metadata system has been developed for STC's business surveys, which contains systems and related process metadata for questionnaires, edit and imputation, allocation and estimation, analysis and retrieval of the business data. While the business rules have not been finalized, the IMDB will be used as the source of metadata for archived data, an important and last phase of the statistical cycle.

## 4. SYSTEMS AND DESIGN ISSUES

### Development History

37. The IMDB project was initiated in 1998. The ANSI X3.285 Metamodel for the Management of Shareable Data and U.S. Bureau of Census's Corporate Metadata Repository (CMR)[4] model were chosen as the basis for the data model design. The ANSI X3.285 was later enveloped by the ISO 11179 standard. The IMDB development was partitioned into the following phases:

| | |
|---|---|
| Phase 1 | Consolidation of existing metadata stores into a central store; |
| Phase 2 | Metainformation describing the statistical business processes; and |
| Phase 3 | Metadata for data elements |

### Collection Systems[5]

a) Consolidation of existing data stores into a central store

38. The following systems were retired and their metadata consolidated into the IMDB:

- Statistical Data Documentation System (SDDS) consisting of primary textual descriptions of the statistical business processes.
- Meta Inventory of Data Assets Systems (MIDAS) consisting of metadata describing the confidential master data files.
- Thematic Search Tool and Paradox Meta System for Social Statistics
- Questionnaire Inventory

39. Custom extraction and load programs were built to move the data into the IMDB. This activity was completed in early 2000 and the four systems were retired.

b) Development of a collection system for Phase 2 Metainformation describing the statistical business process

40. An initial investigation was done by a development team to determine if there were already existing software tools both internally and externally to support collection. Existing software tools were not discovered, therefore, the decision was made for in-house development.

41. Oracle 8i was selected as the database and IBM Visual Age for Java was selected as the development tool. This system referred to as *MetaStat* was in development and production from 1999-2002. The development was ceased in 2002 because the IBM Visual Age for Java product was discontinued by the vendor and a migration path to another product was not supplied by the vendor. The data content collected and managed by the *MetaStat* system consists of Statistical Activity, Survey, Instance, Frame, Universe, Instrument, Data Files, Survey Methodology and Documentation. Supporting content also collected by MetaStat includes Organization, Contact, Keyword and Theme. The *MetaStat* system is still in current use for collection of Phase 2 information. New development for *MetaStat* was frozen in 2002. *MetaStat* support currently consists of ensuring the Oracle database drivers for Java and the Java classes (currently tested to support Java 1.3) will continue to behave as expected as we migrate to newer versions of the Oracle database. The current production version is Oracle 9i with the planned migration to 10g in July 2007. In the future, *MetaStat* will be retired and the functionality will be incorporated into architecture of the Phase 3 system described in the next paragraph.

[4] Daniel W. Gillman, 1999: Corporate Metadata Repository (CMR) Model; U.S Bureau of Labor Statistics.
[5] A Powerpoint presentation: **IMDB Phase 2-3 Architecture.ppt** provides an overview of the systems architecture.

c) Development of a collection system for Phase 3 Metadata for data elements

42. The decision was made by the development team to move towards open source development tools. The basis for this decision was to reduce the risk of vendor lock-in as was experienced during the development of the *MetaStat* system.

43. The development of the Phase 3 also provided an opportunity to enhance the data model to provide multilingual data support. In Phase 2, each language is represented in the Oracle tables as one column for English data and one column French data. In Phase 3, the English and French data are each represented each as a separate row in the Oracle table therefore allowing the ease of addition of other languages if required.

44. The system developed for Phase 3 is referred to as *MetaWeb*. It is a Java JSP and Servlet based solution. The data content collected by the *MetaWeb* system consists of Object Class, Property, Data Element Concept and Data Element. Conceptual Domain and Value Domain information is collected and populated into the IMDB database via a Microsoft Excel IMDB Extraction/Loader and an Oracle PL/SQL IMDB Loader. The decision to used Excel as a collection tool for the Conceptual Domain and Value Domain information was based on the functionality present in Excel for data manipulation (such as sorting), facilitation of presentation of complex multi-level information by using individual worksheets, familiarity of use of Excel in the organization and the ease of sharing of the Excel data with other applications (such as Datawarehousing) with in the organization.

45. Registration of all the Phase 3 items is managed by the *MetaWeb* system (including the Conceptual Domain and Value Domain information). The shared data content between Phase 2 and Phase 3 systems is primarily the administrative information (Organization and Contact) required for registration. The Organization and Contact information is managed solely by the *MetaStat* system. This Organization and Contact information is accessible to *MetaWeb*, but is not managed by *MetaWeb*.

46. Initial preparations for migration of the Phase 2 content into the Phase 3 architecture has started by creation of a bridge between the two systems which consist of a creation of Phase 3 system identifiers mapped to the Phase 2 system identifiers. Additional collection interfaces on the development schedule for *MetaWeb* include: Question, Question Response Choices, Question Block and a Value Meaning manager to support the *Survey planning and design* phase of the statistical cycle.

**Dissemination/Information Discovery Systems**

47. Web publication of the information from the IMDB is through HTML pages, which are generated dynamically using Perl scripts. The HTML pages conform to the Government of Canada's web page publication policy for common look and feel and accessibility. The IMDB web pages provide a Survey centric view of the IMDB information. The IMDB web pages on the internet contain the version of the information that is publicly disseminated.

48. Information discovery (available for internal use only) is through a wiki-based solution. Each Wiki page provides the context of the information and provides all the available links to the information. The wiki view provides a non-linear view to the information as the user can decide on the path to take. The wiki engine selected for use at Statistics Canada is MediaWiki 1.8 (which is the same wiki engine used by Wikipedia). The wiki pages are programmatically generated. The information from the IMDB Oracle Phase 2 and Phase 3 database is extracted using a VB .Net application. Specific wiki templates were developed for the IMDB and these are used to provide a consistent display presentation. Wiki tags and Wiki templates added to extracted IMDB information and this information is directly populated into the MySQL database of the MediaWiki engine. The IMDB wiki pages are refreshed daily.

49. Samples of the IMDB wiki pages for Data Element Concept, Data Element, Property, Question, Instance are found in the **SampleWiki*.pdf** files.

## 5. ORGANIZATIONAL AND CULTURAL ISSUES

### Responsibilities for data/metadata bases in the Agency

50. There are a variety of approaches to manage the development and maintenance of data/metadata systems their production units. Some production units have developed and maintain systems specific to individual surveys; others, like the units responsible for the many different annual establishment surveys, share a common system for data collection, processing and estimation. Statistics Canada has developed a number of generalized systems for data processing (i.e., BANFF for edit and imputation) and uses BLAISE for data collection. Overtime, there is a trend towards centralizing systems. To support data aggregation and analysis, STC is using the Common Datawarehouse Model such as the National Accounts Analytical Datawarehouse. In this case, the metadata to support the datawarehouse is "pulled" from the IMDB.

51. When it comes to dissemination, production units may have their own metadata support systems, but they are obliged to provide to Standards Division, for storing into IMDB, the minimal information prescribed by the *Policy on Informing Users on Data Quality and Methodology* (PIUDQM).

52. The development and maintenance of agency wide social, economic, and geographical variables and standard classifications is the responsibility of Standards Division.

53. The work of Standards Division is essentially geared by two STC policies: *Standards Policy* and *PIUDQM*. The application of these policies and their appropriateness over time is overseen by the Methods and Standards Committee which is co-chaired by two Assistant Chief Statisticians.

## 6. ATTACHMENTS & LINKS

Note: The links to attachments below will work once all files in the zip folder at http://www.unece.org/stats/documents/ece/ces/ge.40/2007/wp.11.e.zip are downloaded and extracted to the same location as this paper.

a) Variables – What they are.ppt

b) Registration of Variables.ppt

c) IMDBPhase 2-3 Architecture.ppt

d) AILayerF.ppt

e) SampleWikiDataElement.pdf

f) SampleWikiDataElementConcept.pdf

g) SampleWikiProperty.pdf

h) SampleWikiQuestion.pdf

i) SampleWikiInstance.pdf