

# *Metadata management at Statistics Canada*

---



Alice Born  
Standards Division, Statistics Canada

[alice.born@statcan.ca](mailto:alice.born@statcan.ca)

FEDCASIC Workshop March 6-8, 2007

# Outline

- Description of STC's Integrated Metadatabase (IMDB)
- Common metadata set
- Tools for entering metadata
- Registration – data elements, questions, etc.
- Integration
- STCwiki



# *What's needed for metadata management*

- Disciplined approach to data and metadata management - standards
- An agreed enterprise-wide metadata architecture
- Standard, generic E2E processes: collection, processing and dissemination of statistical data and metadata



# *Statistical metadata at Statistics Canada*

- Integrated Metadatabase (IMDB)
  - Information about each of Statistics Canada's 560+ current surveys
  - Aimed at helping users interpret statistical data
    - Survey description
    - Survey instrument – questionnaires, questions
    - Methodology
    - Data accuracy
    - Variables, classifications
  - Also repository for all variables, classifications and questionnaires



## *What is the IMDB based on?*

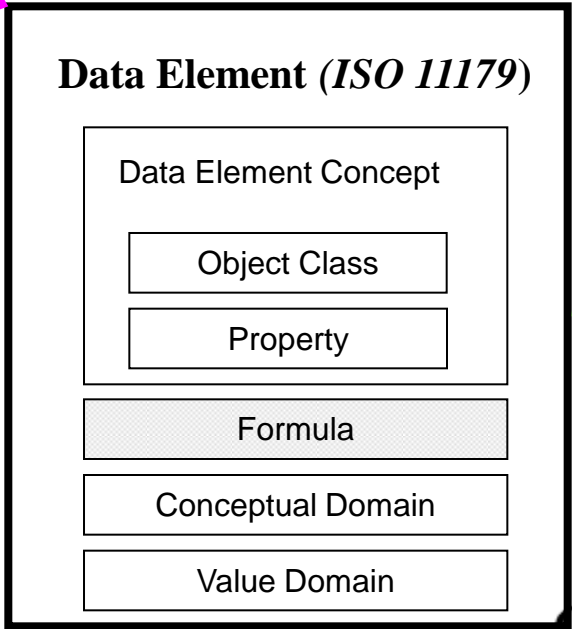
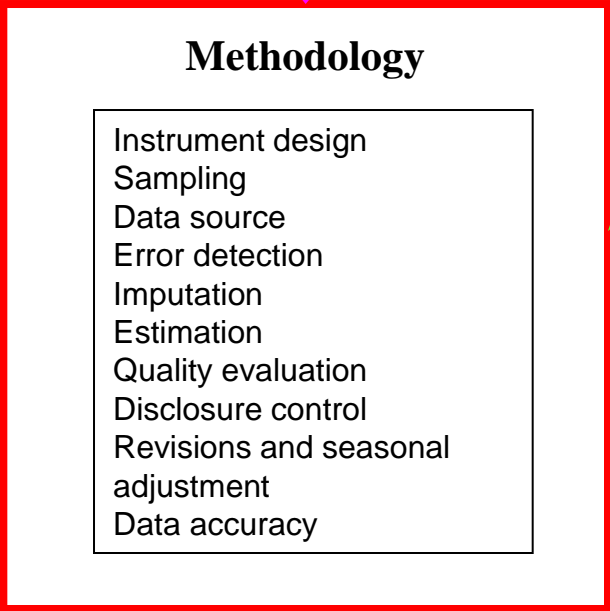
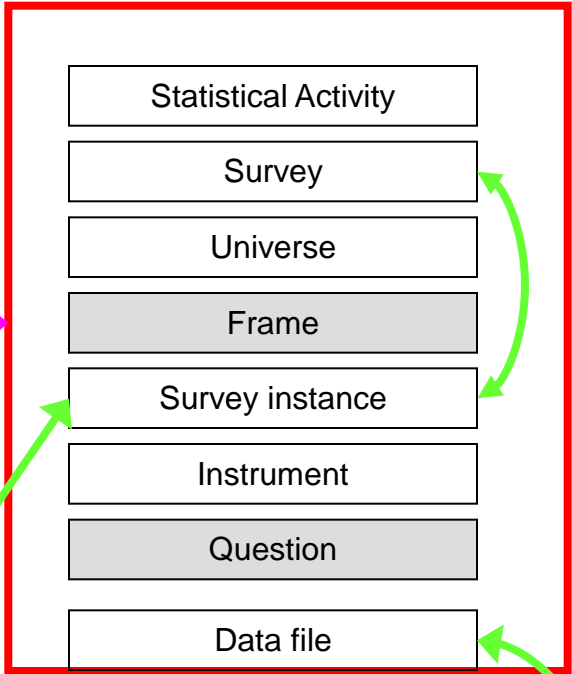
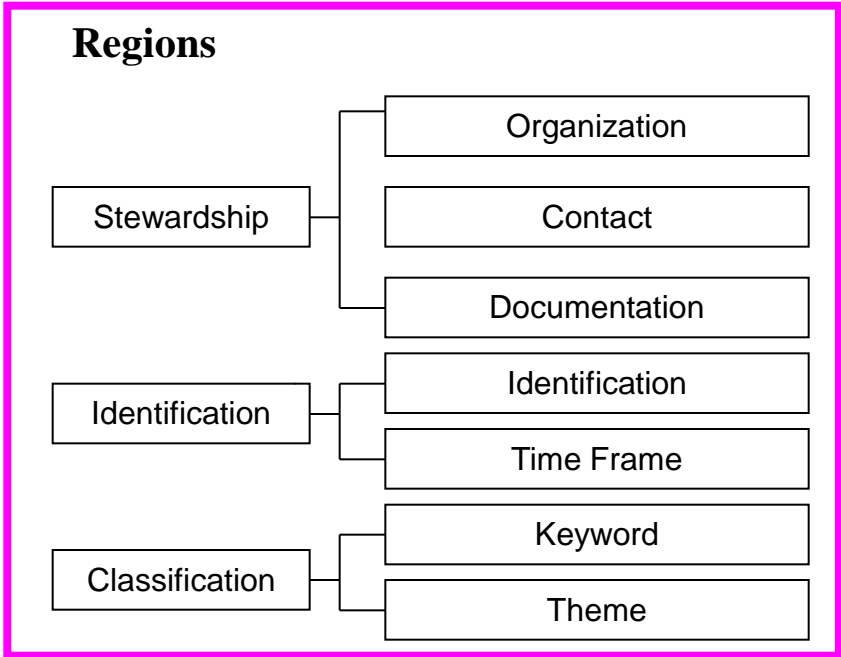
- *ISO/IEC 11179 Metadata Registries*
- Corporate Metadata Repository (CMR) – USBC (D. Gillman)
  - Business dimension – supports survey life cycle
  - Data dimension – data elements and value domains
  - Questionnaire model
  - Administration – registration
- Initially built to support data dissemination but can be linked to other parts of the survey life cycle



## *Surveys and statistical activities- definition*

- Metadata in the IMDB is organized around the *survey* entity
- Three types of surveys:
  - Direct – paper surveys, CAPI, CATI
  - Administrative
  - Derived – GDP, CPI
- *Statistical activity* is a group of surveys that share common feature, common explanatory text
  - e.g., System of National Accounts, Health





# Common metadata set

Definitions, data sources and methods - Microsoft Internet Explorer provided by ITSD - DSTI

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Refresh Print Mail

Address <http://www.statcan.ca/english/concepts/index.htm> Go Links English Spanish

Canada

Français Contact us Help Search ICN  
Site map About us Privacy Accessibility My account

Search this section

**HOME**

Definitions, data sources and methods

What's new

Standard Geographical Classification 2006

Chart of Accounts 2006

New surveys

Upcoming releases

Consultation on classifications

NAPCS 2007

NAICS 2007

Reference

Classification coding tools

Concordances to other classifications

About Definitions, data sources and

**STATISTICS CANADA**  
CANADA'S NATIONAL STATISTICAL AGENCY

## Definitions, data sources and methods

Our purpose is to provide information that will assist you in interpreting Statistics Canada's published data. The information (also known as metadata) is provided to ensure an understanding of the basic concepts that define the data including variables and classifications; the underlying statistical methods and surveys; and key aspects of the data quality. Direct access to questionnaires is also provided.

### Surveys and statistical programs

- [List by subject](#)
- [Alphabetical list](#)

### Questionnaires

- [List by subject](#)
- [Alphabetical list](#)

### Definitions

- [Statistical units](#)
- [Concepts and variables](#)

### Standard classifications

- [Industry](#)
- [Occupation](#)
- [Product](#)
- [Geography](#)
- [Instructional Programs](#)
- [Chart of Accounts: Financial Position and Performance](#)

### Other classifications

- [List by subject](#)
- [Alphabetical list](#)

Local intranet

start

Inbox - Microsoft ... M:\

2 Microsoft Pow... Welcome to the I... Definitions, data s...

4:00 PM



# Common metadata set – list of surveys

Statistical methods: List of surveys and statistical programs by subject - Microsoft Internet Explorer provided by ITSD - DSTI

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Refresh Print

Address <http://www.statcan.ca/english/sdds/index.htm> Go Links English Spanish

Statistics Canada Statistique Canada

Canada

Français	Contact Us	Help	Search	ICN
The Daily Census	Canadian Statistics	Community Profiles	Our products and services	Home Other Links

Home

[List of surveys and statistical programs by record number](#)

## List of surveys and statistical programs by subject

**Aboriginal peoples** [\[Complete list\]](#) ( 3 )

- Aboriginal society and community
- [Business and finance](#)
- Education, literacy and skills
- [Health and well-being](#)
- Households, housing and environment
- Justice issues
- Languages and cultures
- [Population characteristics](#)
- [Work, income and spending](#)

**Agriculture** [\[Complete list\]](#) ( 49 )

- [Crops and horticulture](#)
- [Farm financial statistics](#)
- Farms and farm operators
- Food and nutrition
- [Land use and environmental practices](#)
- [Livestock and aquaculture](#)

**Business performance and ownership** [\[Complete list\]](#) ( 17 )

- Business adaptation and adjustment

<http://www.statcan.ca/start.html> Local intranet

start | Inbox - Microsoft ... | M:\ | 2 Microsoft Pow... | Welcome to the I... | Statistical method... | 4:03 PM

# Common metadata set – survey instance

Annual Survey of Architectural Services - Microsoft Internet Explorer provided by ITSD - DSTI

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites RSS Print Mail

Address <http://www.statcan.ca/cgi-bin/imdb/p25v.pl?Function=getSurvey&SDDS=2420&lang=en&db=IMDB&dbg=f&adm=8&dis=2> Go Links English Spanish

Statistics Canada Statistique Canada

Canada

Français Contact us Help Search ICN  
Site map About us Privacy Accessibility My account

Search this section

**STATISTICS CANADA**  
CANADA'S NATIONAL STATISTICAL AGENCY

**Annual Survey of Architectural Services**

**Status:** Active  
**Frequency:** Annual  
**Record number:** 2420

The survey objective is the collection and dissemination of data necessary for the statistical analysis of the Architectural Services.

**Detailed information for 2005**  
**Data release - December 11, 2006**

- [Questionnaire\(s\) and reporting guide\(s\) \(in PDF format\)](#)
- [Variables](#)
- [Description](#)
- [Data sources and methodology](#)
- [Data accuracy](#)

HOME

Definitions, data sources and methods

Additional Information

Summary of changes over time

Other reference periods

Links to related products:

The Daily

Summary tables

Publications

CANSIM (\$)

Done Local intranet

start [Taskbar icons] 4:07 PM

# Common metadata set – list of variables

The screenshot shows a Microsoft Internet Explorer browser window. The title bar reads "List of variables for: Annual Survey of Architectural Services - 2005 - Microsoft Internet Explorer provided by ITSD - DSTI". The address bar contains the URL: [http://www.statcan.ca/cgi-bin/imdb/p3Variable.pl?Function=getDEInstance&SourceId=2420&sAC\\_Id=2420&sAC\\_Version=1](http://www.statcan.ca/cgi-bin/imdb/p3Variable.pl?Function=getDEInstance&SourceId=2420&sAC_Id=2420&sAC_Version=1). The page content includes the Statistics Canada logo and navigation menu with links for Français, Contact us, Help, Search, ICN, Site map, About us, Privacy, Accessibility, and My account. A search box is present with the text "Search this section". The main heading is "STATISTICS CANADA CANADA'S NATIONAL STATISTICAL AGENCY" followed by "List of variables for: Annual Survey of Architectural Services – 2005". The text below states: "This list consists of variables currently disseminated through CANSIM including links to their definitions and classifications." A list of links follows: [Expenses of Establishment, Type](#), [Expenses of Establishment, Value](#), [Industry of Establishment, Type](#), [Occurrence of Establishments, Count](#), [Revenue of Establishment, Type](#), and [Revenue of Establishment, Value](#). A sidebar on the left lists "HOME" categories: "Definitions, data sources and methods", "Additional Information", "Summary of changes over time", "Other reference periods", "Links to related products:", "The Daily", "Summary tables", "Publications", and "CANSIM (\$)". At the bottom, the date "Date modified: 2005-11-28" is shown, along with "Top of page" and "Important notices" links. The Windows taskbar at the bottom shows the Start button, several open applications, and the system clock at 4:05 PM.

# Common metadata set – data element

Type of Expenses of Establishment - Microsoft Internet Explorer provided by ITSD - DSTI

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites RSS Mail Print

Address [http://www.statcan.ca/cgi-bin/imdb/p3Variable.pl?Function=assembleVariable&Item\\_Id=1986&Repclass=586&EntryType=](http://www.statcan.ca/cgi-bin/imdb/p3Variable.pl?Function=assembleVariable&Item_Id=1986&Repclass=586&EntryType=) Go Links English Spanish

Statistics Canada Statistique Canada

Canada

Français	Contact us	Help	Search	ICN	
Site map	About us	Privacy	Accessibility	My account	

Search this section

**STATISTICS CANADA**  
CANADA'S NATIONAL STATISTICAL AGENCY

## Type of Expenses of Establishment

'Expenses' refers to decreases in economic benefits or service potential, during the reporting period, in the form of outflows or consumption of assets or incurrence of liabilities that result in decreases in equity, other than those relating to distributions to owners.

'Establishment' refers to the level at which the accounting data required to measure production is available (principal inputs, revenues, wages). The establishment, as a statistical unit, is defined as the most homogeneous unit of production for which the business maintains records from which it is possible to assemble all the data elements required to compile the full structure of the gross value of products or shipments, and inventories), the cost of materials and services, and labour and capital used in production.

'Type' refers to the reporting of 'Expenses of Establishment' using the following classification(s):

- [Expense Categories Services](#)

**CANSIM Links**

[Summary statistics for architectural services \(all establishments\), by North American Industry Classification System \(NAICS\) \(3600004\)](#)

start | Local intranet | 4:18 PM

## *Data elements and value domains*

Following ISO 11179 approach to defining variables and reusing items:

- 85 object classes (statistical units)
- 290 characteristic items (properties)
- 506 data element concepts
- 1,059 value domains (non-enumerated and enumerated - classifications)
- 1035 data elements (variables)



# Loading metadata into IMDB

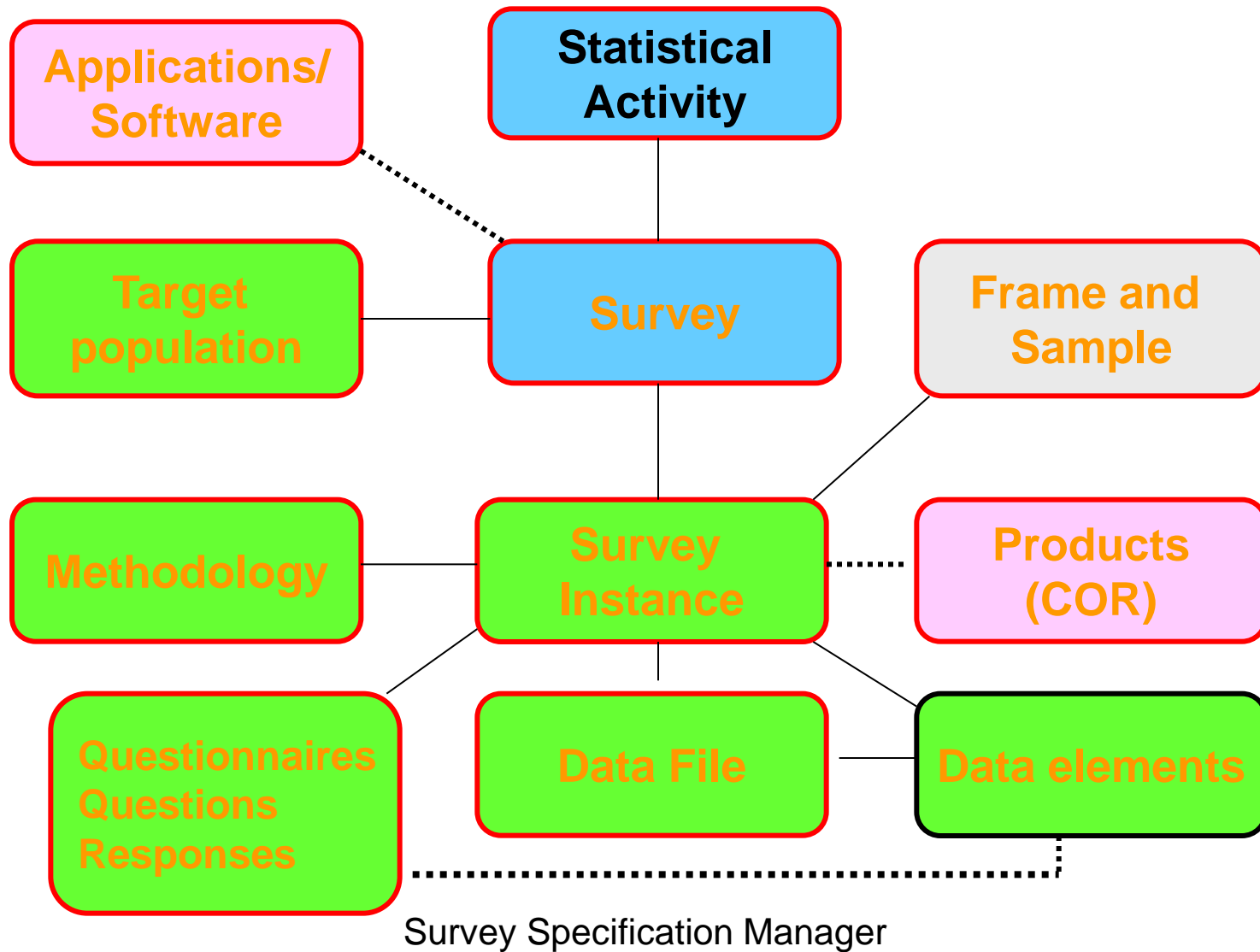
The screenshot shows a software window titled 'Statistical Activity' with a menu bar containing 'File', 'Statistical Elements', 'Methodology', 'Reference Lists', 'IMDB Administration', and 'Help'. The main form area is divided into several sections:

- Header:** ID: 10293, Version: 1.0
- Name:** Health and Activity Limitation Survey
- Directive:** (Empty field)
- Navigation Tabs:** Identification, Description, Time Frame, Documentation, Classification, Organization, Context
- Sub-sections:**
  - Identification Administration:**
    - Name (English): Health and Activity Limitation Survey
    - Name (French): Enquête sur la santé et les limitations d'activités
    - Acronym (English): HALS
    - Acronym (French): ESLA
  - Version Revision:**
    - Revision Description (English): (Empty)
    - Revision Description (French): (Empty)

At the bottom of the window are four buttons: 'Close', 'Delete', 'Save', and 'Cancel'.



# CMR – Business dimension model



# *Registration of metadata*

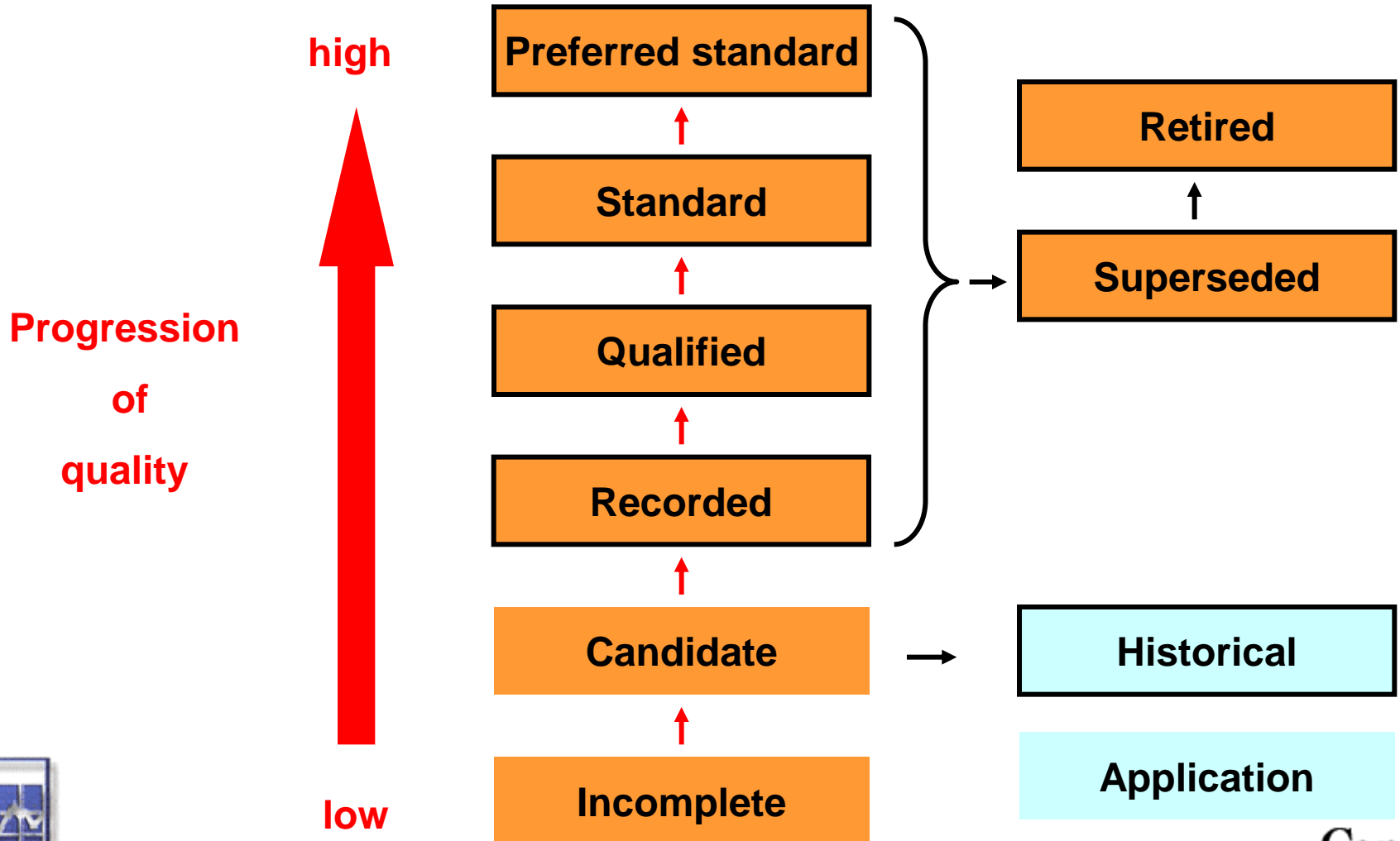
- Set of rules that apply to metadata registry
  - Monitor source of metadata and quality of the metadata
1. **Registration status** – identifies the quality or progression of quality
  2. **Administrative status** – stage in the registration process
  3. **Registration level** – level of conformance or harmonization



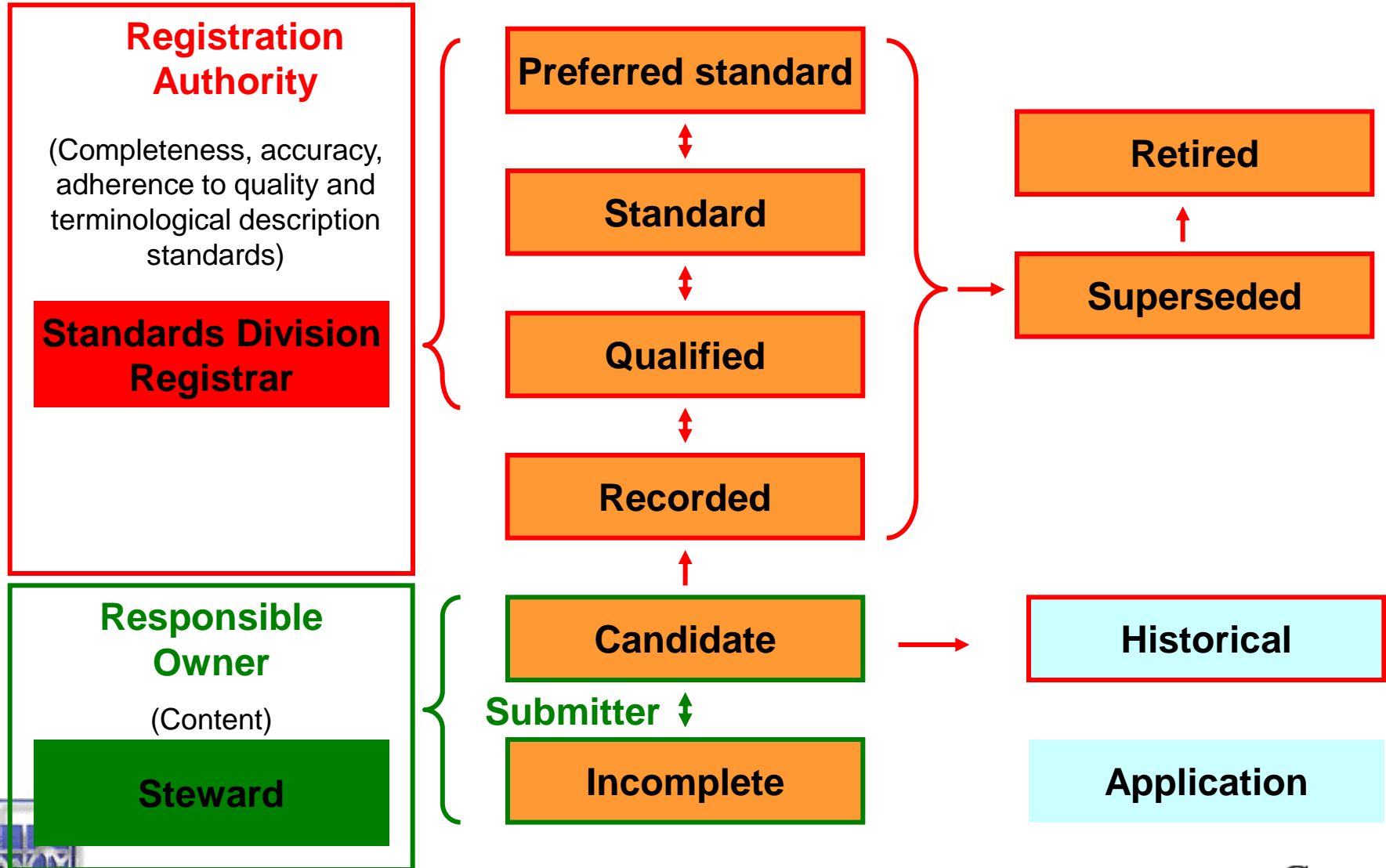


# Registration status

Identifies the quality or progression of quality

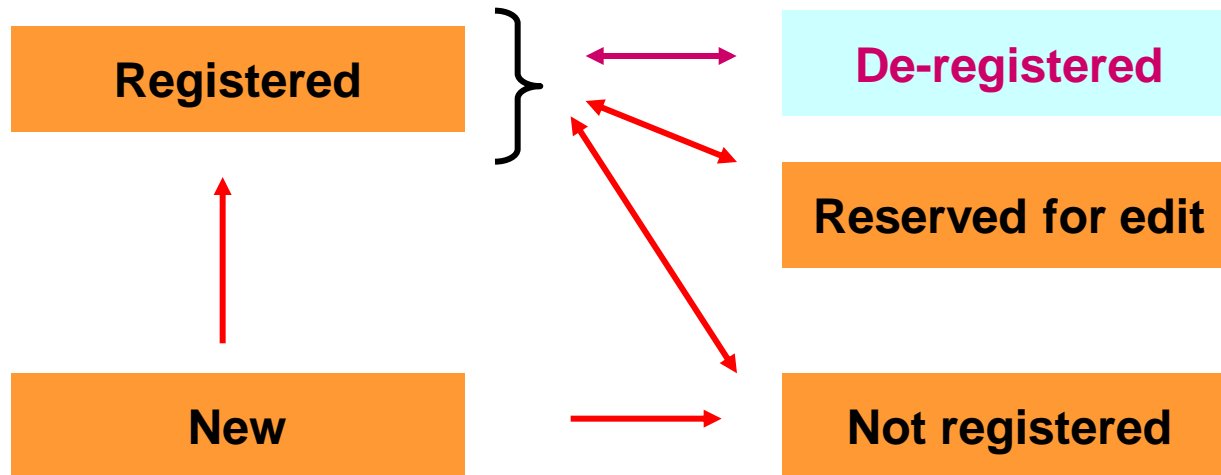


# Management of Registration status



# Administrative status

Stages in registration process

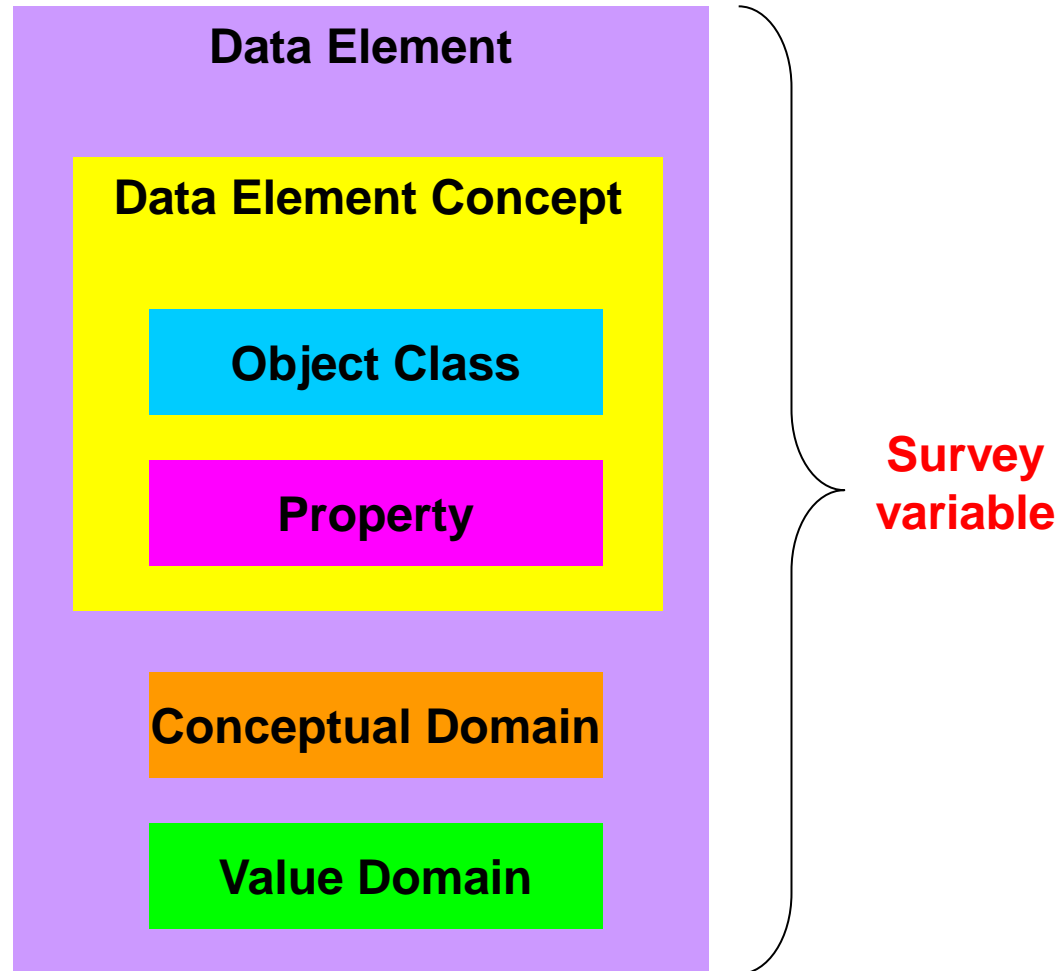


# Registration level

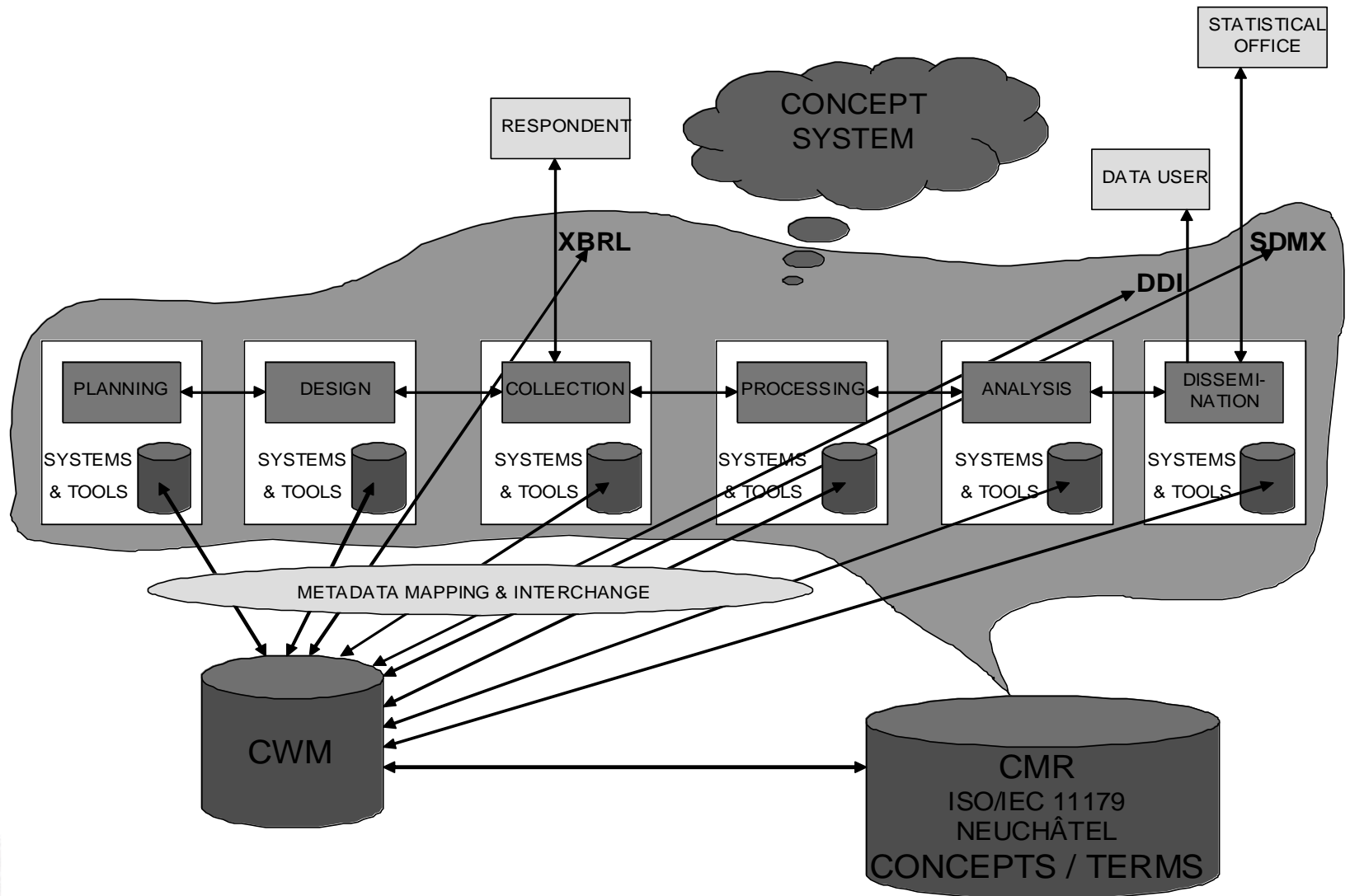
Level of conformance or harmonization



# *Data elements - registration*



# Integration – next steps



## Metadata Schemes - STC

- XBRL (extensible business reporting language) – financial data for business surveys
- DDI (data documentation initiative) - DLI
- SDMX (statistical data and metadata exchange) – OECD's NAWWE, education statistics, external debt
- CWM (common warehouse metamodel) – System of National Accounts - IMDB



## *Linking questions to IMDB*

- Linking the Survey Specification Manager (question blocks for household surveys) to IMDB
- Potential for:
  - Reuse of administrative layer in the IMDB
  - Linking to other metadata (questions to data elements and value domains, surveys)
  - Populating the questionnaire model in the IMDB from the SSM





# STC Wiki

mhtml:file://F:\files\ImdbPortal.mht

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites Refresh Mail Print

Address F:\files\ImdbPortal.mht Go

Log in / create account

portal discussion edit history

## Portal:IMDB

Social • Business • National Accounts • Analysis • Communications • Informatics • Management Services • All Portals

### The IMDB Portal edit

The Integrated Metadata Base (IMDB) is a corporate repository of information on each of Statistics Canada's nearly 400 active surveys (and a roughly equal number of discontinued 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 corporate knowledge management and for data users.

Metadata can support at least three broad functions within a statistical agency: data dissemination; data production, including collection; and management of the statistical system. IMDB was designed to initially support data dissemination, that is, to provide users with the information they need to interpret the statistical data we disseminate. As a result, the content of the database is largely dictated by the requirements of the [Policy on Informing Users of Data Quality and Methodology](#).



[Read more...](#)

#### Selected article edit

### Registration Status

The registration status category identifies the quality or progression quality of administered items (AI). It identifies the AI's completeness, accuracy, and performance to context and format.

#### Related portals edit

[Informatics](#) | [Systems Development Division](#) | [Data Warehouse Centre](#) | [Data Warehouse Centre](#)

navigation

- Main Page
- Community portal
- Current events
- Recent changes
- Random page
- Help
- Donations

search

marital status Go Search

toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link

start

FEDCASIC presentati... F:\files mhtml:file://F:\files\I...

6:20 PM