Metadata management at Statistics Canada

Alice Born
Standards Division, Statistics Canada

alice.born@statcan.ca

FEDCASIC Workshop March 6-8, 2007

Canada

Outline

- Description of STC's Integrated Metadatabase (IMDB)

- Common metadata set

- Tools for entering metadata

- Registration – data elements, questions, etc.

- Integration

- STCwiki

Canada

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

Canada

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
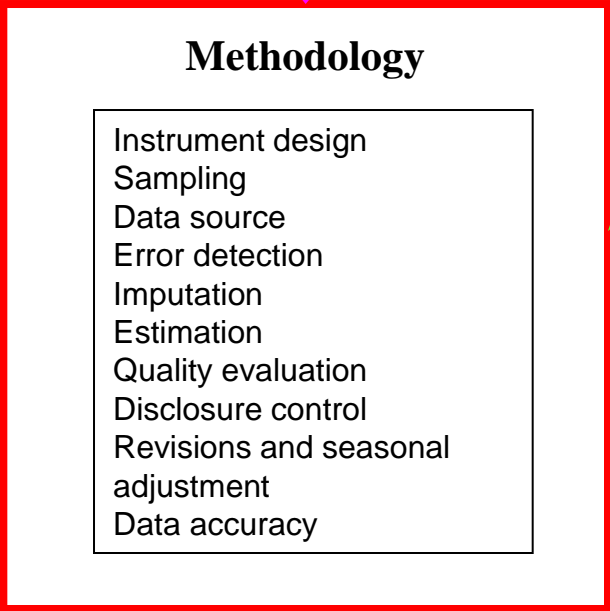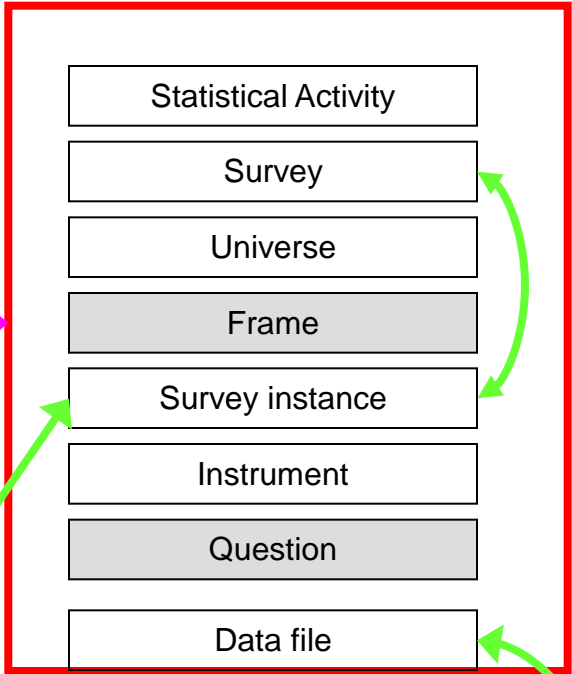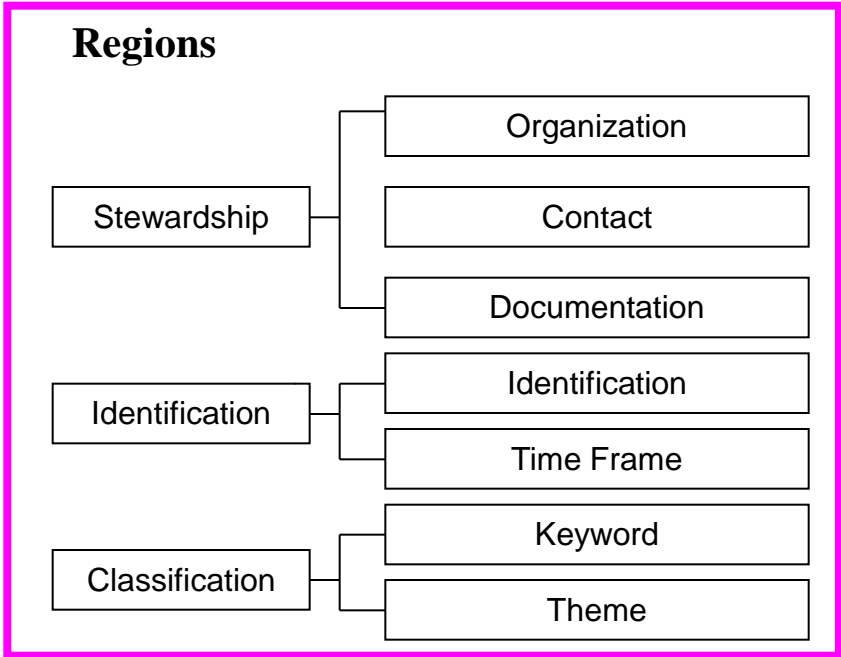
What is the IMDB based on?

- 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

Canada

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

Regions

Stewardship
- Organization
- Contact
- Documentation

Identification
- Identification
- Time Frame

Classification
- Keyword
- Theme

- Statistical Activity
- Survey
- Universe
- Frame
- Survey instance
- Instrument
- Question
- Data file

Methodology

Instrument design
Sampling
Data source
Error detection
Imputation
Estimation
Quality evaluation
Disclosure control
Revisions and seasonal adjustment
Data accuracy

Data Element *(ISO 11179)*

Data Element Concept
- Object Class
- Property

- Formula
- Conceptual Domain
- Value Domain

Canada

Common metadata set

Common metadata set – list of surveys

Common metadata set – survey instance

Common metadata set – list of variables

Common metadata set – data element

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)

Canada

Loading metadata into IMDB

CMR – Business dimension model

Survey Specification Manager
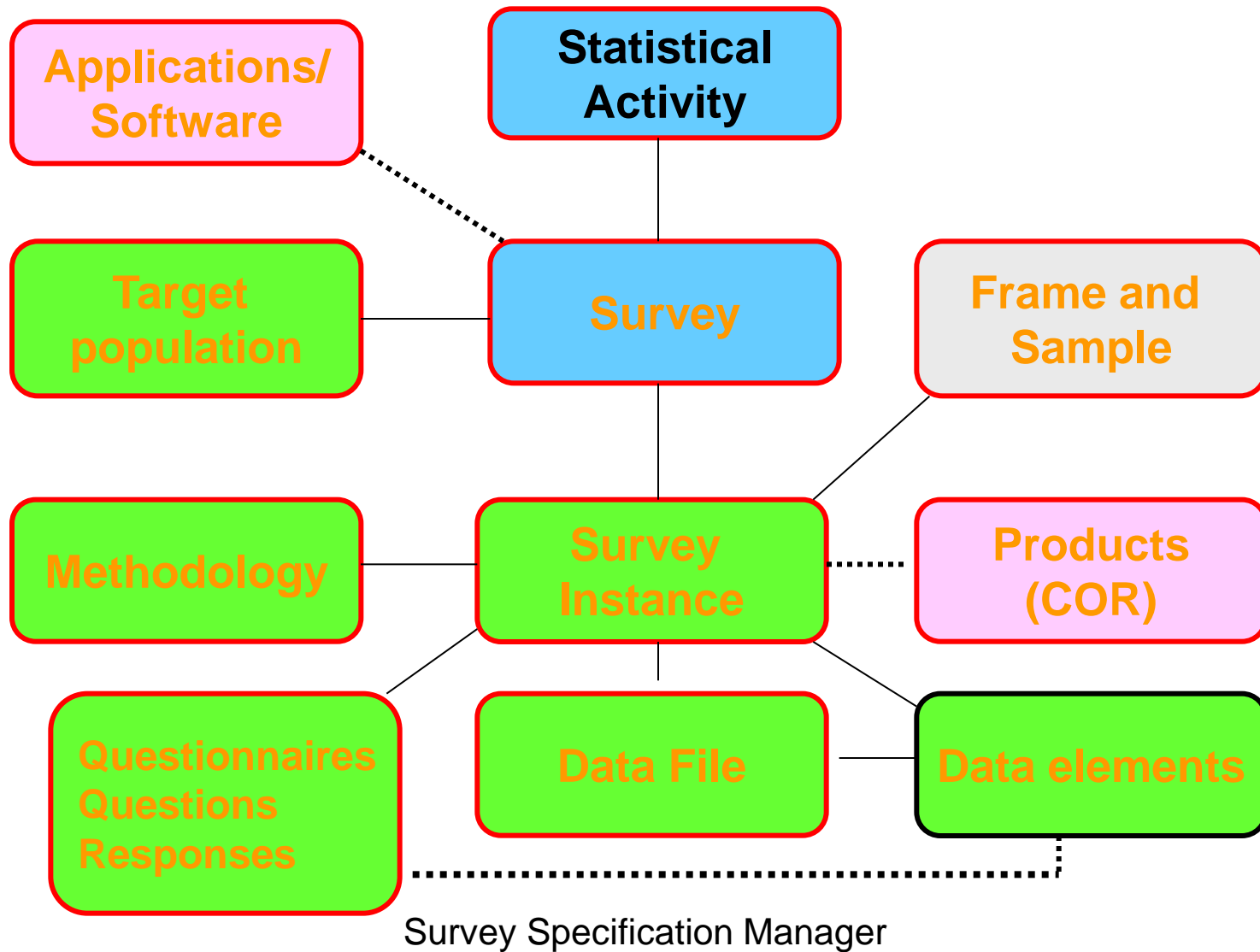
Registration of metadata

- 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

high

Progression

of

quality

low

| Preferred standard |
| Standard |
| Qualified |
| Recorded |
| Candidate |
| Incomplete |

| Retired |
| Superseded |

| Historical |
| Application |
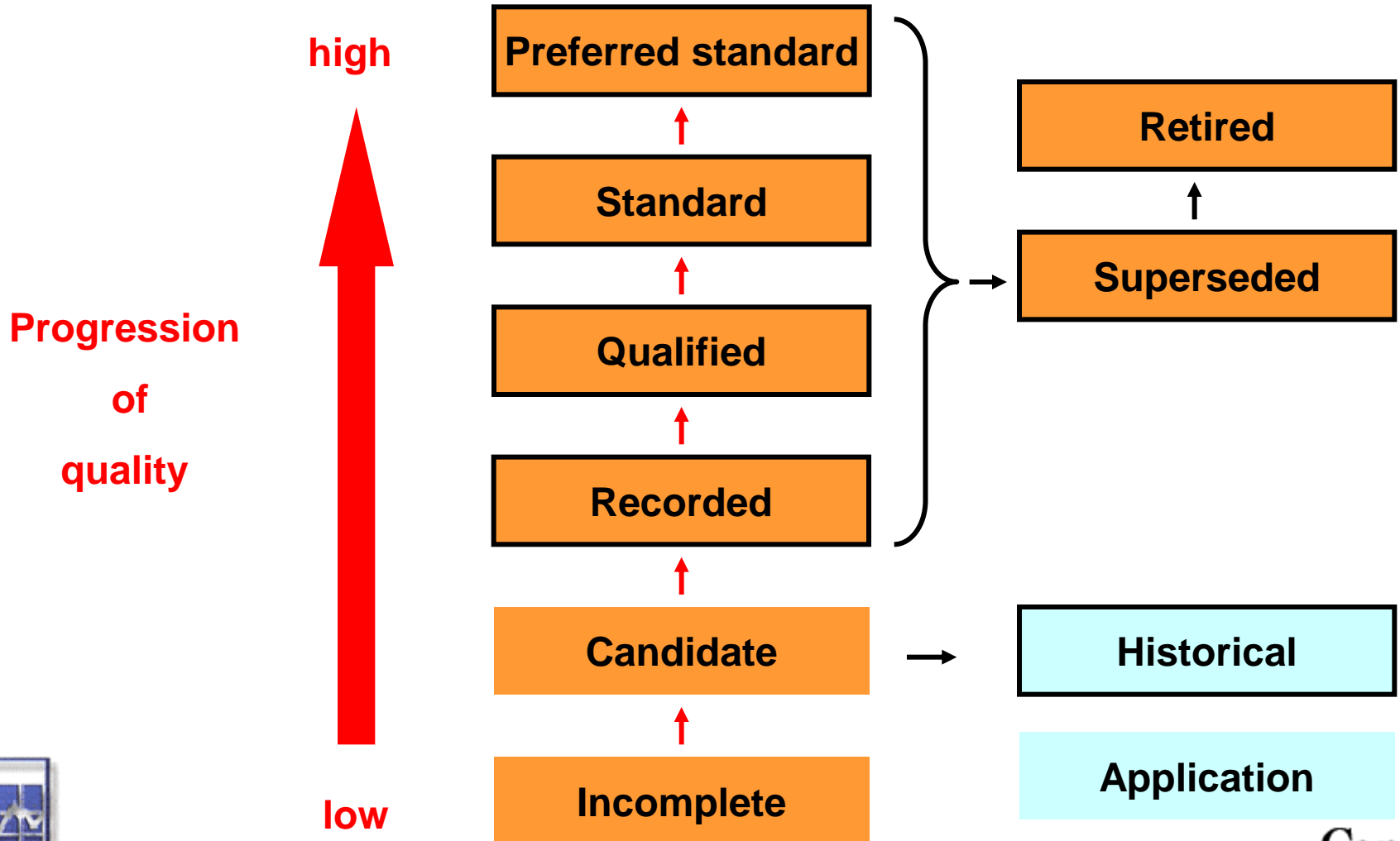
Management of Registration status

Registration Authority

(Completeness, accuracy, adherence to quality and terminological description standards)

Standards Division Registrar

Responsible Owner

(Content)

Steward

Submitter

Preferred standard

Standard

Qualified

Recorded

Candidate

Incomplete

Retired

Superseded

Historical

Application

Canada

Administrative status

Stages in registration process

Registration level

Level of conformance or harmonization

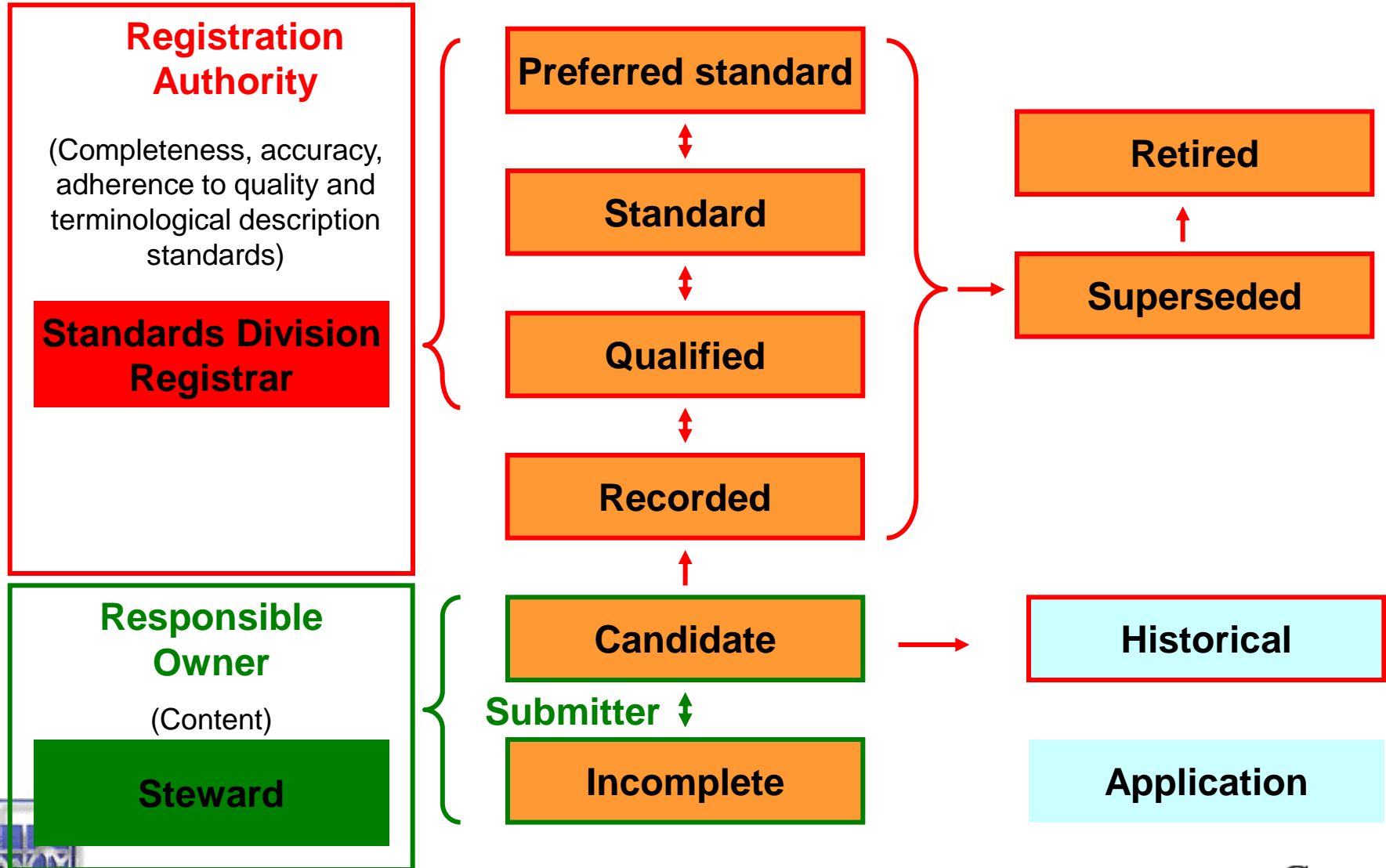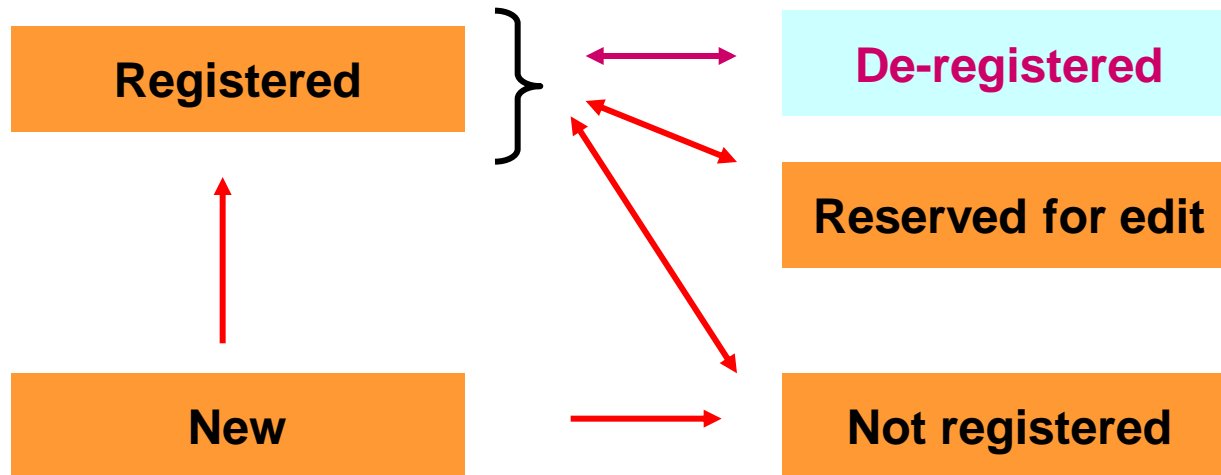|	
Agency	
↑	
Branch	**International**
↑	↑
Field	**North America**
↑	↑
Division	**Canadian**
↑	
Program	
↑	
Survey	

Data elements - registration
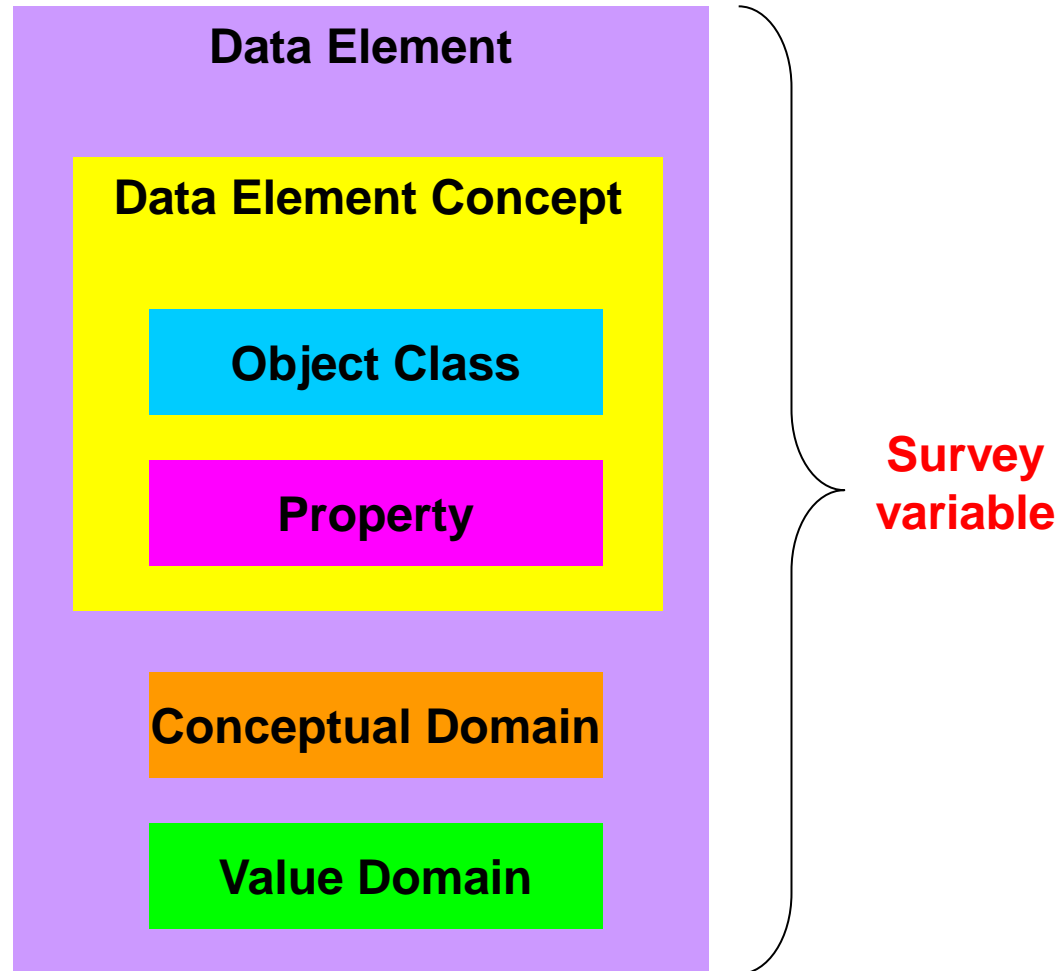
Integration – next steps
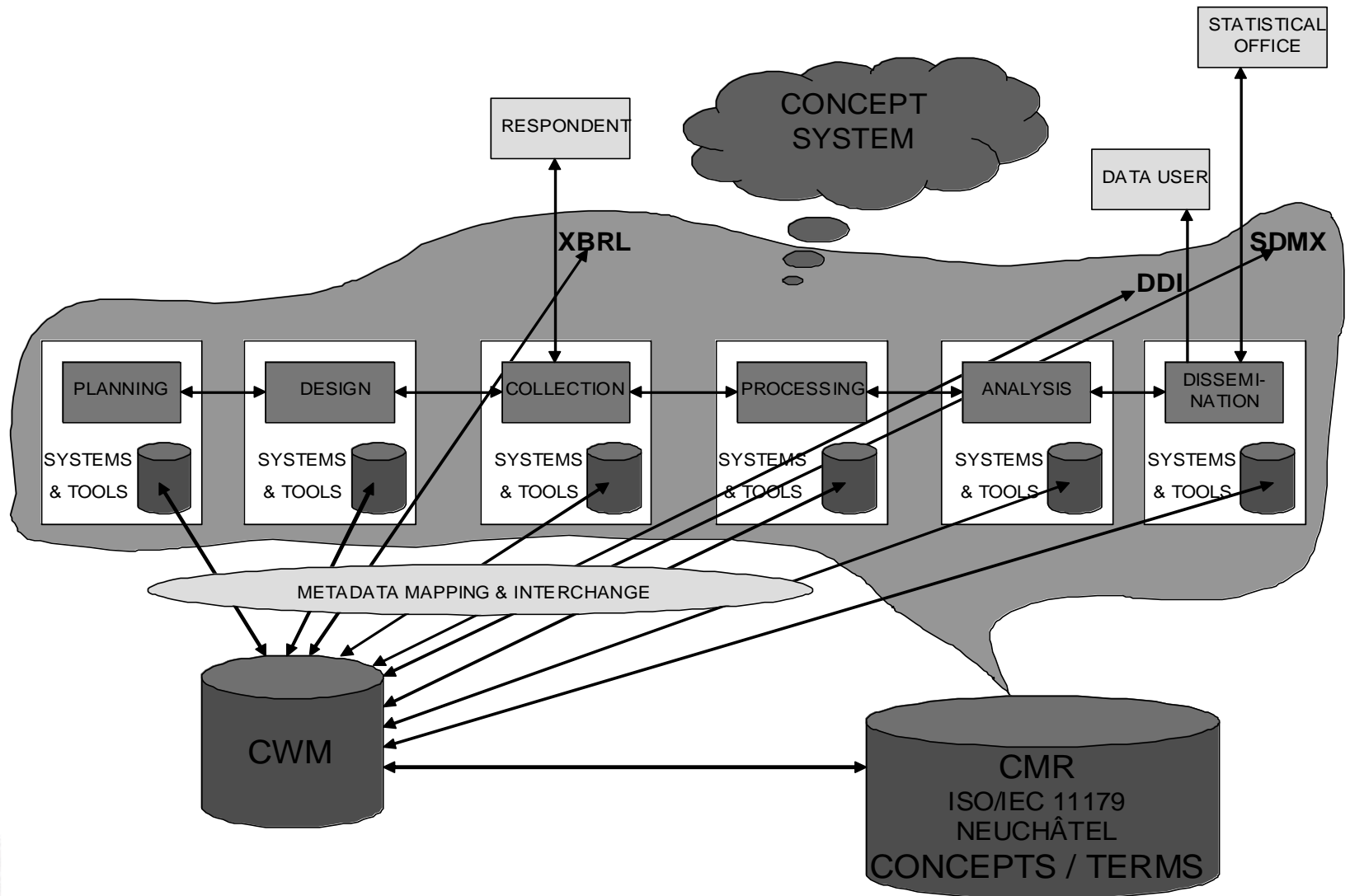
Metadata Schemes - STC

- 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