

Statistics Canada's approach in managing statistical classifications

Presenter: Stéphane Crête, Assistant Director -
Centre for Statistical and Data Standards, Statistics
Canada

31 January 2024

Addressing fundamental gaps in our Data Management and Metadata capabilities

Many statistical organizations face challenges in **harmonizing** their data and metadata content and structure: **Statistics Canada is no different. Our Agency was lacking** enterprise data and metadata capabilities. For example, our stored classifications were difficult and error-prone to work with, especially when using statistical packages and programming languages.

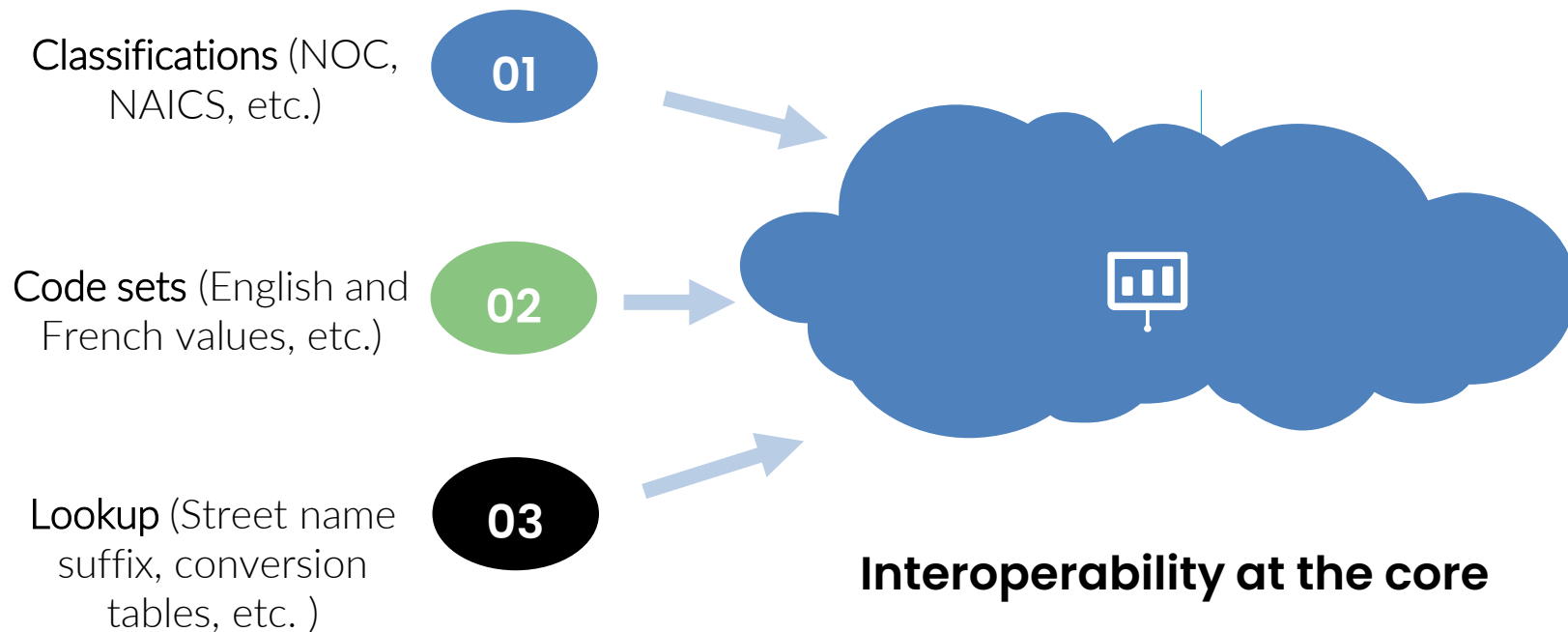
Little to no approach had been taken to provide external-facing Data Stewardship capabilities.

Technology is helping us pave the way forward for metadata management.



Reference Data as a Service (RDaaS)

Reference Data is defined by the set of permissible values used when describing data



Interoperability at the core

Sharing of the repository of **high quality reference data** for **use by all**

Department Consistency

Without a common vocabulary we can't talk to each other

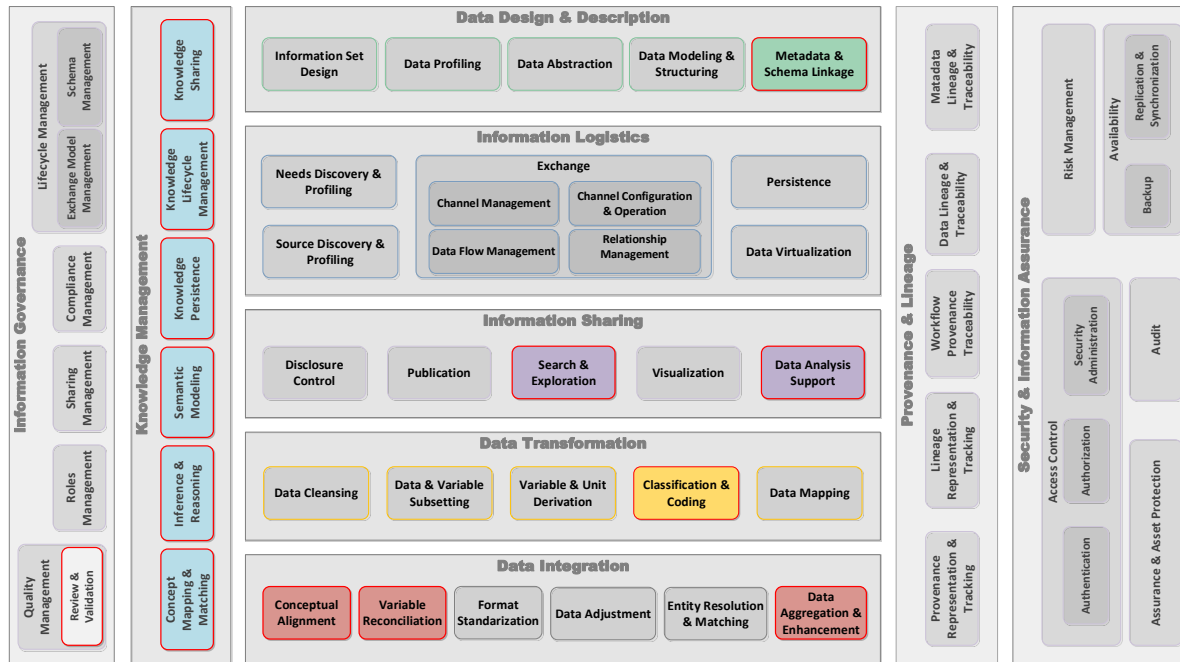
Data Value

Without **consistent Reference Data** across Open Data, data cannot be accurately connected and value is low

Data Sharing

Inconsistent Reference Data is a **barrier for interdepartmental sharing**

Statistical Classifications – Data Capabilities



Statistical classifications will help satisfy the needs/requirements of many official statistical-related data capabilities such as:

- Knowledge management
- Classification & coding
- Search & exploration
- Metadata & schema linkage
- Conceptual alignment
- Variable reconciliation
- Data aggregation & enhancement

Ariā – Manage Statistical Classifications



International Standard Industrial Classification of All Economic Activities Rev. 4 V1.0.0

- ▼ **A** Agriculture, forestry and fishing [3]
 - > **A01** Crop and animal production, hunting and related service activities [7]
 - ▼ **A02** Forestry and logging [4]
 - ▼ **A021** Silviculture and other forestry activities [1]
 - A0210** Silviculture and other forestry activities
 - > **A022** Logging [1]
 - > **A023** Gathering of non-wood forest products [1]
 - > **A024** Support services to forestry [1]
 - > **A03** Fishing and aquaculture [2]
- > **B** Mining and quarrying [5]
- > **C** Manufacturing [24]
- > **D** Electricity, gas, steam and air conditioning supply [1]
- > **E** Water supply; sewerage, waste management and remediation activities [4]

Levels	
Levels	4 (766 Codes)
Section	21
Division	88
Group	238
Class	419

- Lifecycle management approach for statistical classifications
- Ability to manage index terms, inclusions and exclusions
- Ability to extend categories with user defined attributes

Ariā – Manage correspondence tables and crosswalks



North American Industry Classification System - Canada V2007.1.0 to V2012.1.0 V1.0.0

Source Map Target Discussion

Found 28 matches in 245ms

Source	Map Types	Count	Target
33322 : Rubber and Plastics Industry	Merger (---/---)	0	33324 : Industrial machinery manufacturing
454113 : Mail-Order Houses	Merger (---/---)	8	454110 : Electronic shopping and mail-order houses
44313 : Camera and Photographic	Merger (---/---)	0	44314 : Electronics and appliance stores
31132 : Chocolate and Confectionery	Merger (---/---)	0	31135 : Chocolate and chocolate confectionery manufacturing
Beans	Merger (---/---)	0	315249 : Women's and girls' cut and sew clothing manufacturing
315234 : Women's and Girls' Cut and Sew Clothing Manufacturing	Merger (---/---)	2055	44314 : Electronics and appliance stores
Jacket and Skirt Manufacturing	Merger (---/---)	0	72251 : Full-service restaurants and limited-service eating places
44311 : Appliance, Television and	Merger (---/---)	9	7225 : Full-service restaurants and limited-service eating places
72211 : Full-Service Restaurants	Merger (---/---)	0	7225 : Full-service restaurants and limited-service eating places
7222 : Limited-Service Eating Places	Merger (---/---)	0	7225 : Full-service restaurants and limited-service eating places
7221 : Full-Service Restaurants	Merger (---/---)	0	454110 : Electronic shopping and mail-order houses
454112 : Electronic Auctions	Merger (---/---)	0	

- Managing changes to statistical classification versions
- Managing crosswalks between statistical classifications
- Manages simple and complex mappings (1 to 1, many to 1, 1 to many, many to many) with the use of distribution ratios
- GSIM-driven approach in documenting changes between versions of a classification

Ariā – Manage Statistical Classifications



Aria API 1.1.0-SNAPSHOT OAS3

[aria.swagger.json](#)

Ariā is an innovative platform for managing and leveraging knowledge in support of data production, publication, or analysis. Aligned with current initiatives promoting modernizing of data production and analysis, Ariā delivers a powerful web based collaborative environment for maintaining concepts, classifications, concordances and related entities across their lifecycle, and deliver information as user friendly reports, actionable scripts/programs, or standards based XML. Flexible search capabilities enable rapid and efficient discovery and navigation, and powerful serializers deliver content in diverse formats for immediate reuse by users and applications. Ariā further complies and facilitates the use and adoption of global standards and best practices.

The Ariā API is designed not only as a way to retrieve the metadata that is managed in the Ariā platform, but also to perform specialized queries about specific resources. This allows both a comprehensive look into the metadata, but also saves users time by answering questions without the need to dig through the metadata by hand.

Servers

/aria-api

Search

Search

Ariā's search functionality provides users with the ability to search for the resources they are looking for based on a variety of parameters. Each type of resource will have an endpoint where the actual search is performed, along with the potential for supplemental endpoints that can be used to retrieve user specific parameters to use in the search. These parameters may change based on the user's role.

GET /api/search/classifications Search classifications.

GET /api/search/classifications/filters Gets classification search filters.

GET /api/search/concordances Search concordances.

GET /api/search/concordances/filters Gets concordance search filters.

- Comprehensive set of modern RESTFul APIs
- Open API Documentation
- Great Internationalization features
- Ability to generate Linked Data output for SKOS/XKOS



Statistics
Canada

Statistique
Canada

Canada

Lessons Learned

- Interoperability is gaining popularity
 - Difficulties in accessing official statistical classifications
 - Lacking technical specifications for APIs
- Challenges in integrating to other standards
 - Generic Statistical Information Model (GSIM) versus Linked Data (SKOS/XKOS)
 - Statistical Data and Metadata eXchange (SDMX)
- Difficulties in querying semantic knowledge graphs
- Difficulties in using version-based artifacts



Moving Forward

- Increased international collaboration is needed
- Official registry of statistical classifications
- Develop technical specifications for APIs
 - Stepping stone for interoperability and open-source solutions
 - Leveraging the potential of metadata for machine learning and other AI applications
- Set expectations for the role of data standards
 - Develop mappings between data standards
 - Revise/develop best practices/guidelines



Stay connected!

StatsCAN app

Eh Sayers podcast

StatsCAN Plus

The Daily

Website

Surveys and statistical programs

Data service centres

My StatCan



Questions? Contact us: infostats@statcan.gc.ca

Statistics Canada— Your National Statistical Agency



Delivering insight through data for a better Canada

