Knowledge Representation and breaking down silos at the Bank of Canada

InterPARES Trust AI International Symposium

Alex Richmond
ENTERPRISE INFORMATION ARCHITECT KNOWLEDGE AND INFORMATION SERVICES

Marielle St-Germain
METADATA AND STANDARDS SPECIALIST KNOWLEDGE AND INFORMATION SERVICES
1. The problem(s)
The business problem(s)....

The Bank is a big and smart place, with lots going on in every function we carry out....

*Sometimes we don’t know what we know!*

*And...it’s no longer good enough to just know things...*

So why is that?

- Because even with the application of cloud-based technologies, data lakes, virtualization etc. the Bank suffers from information silos

What causes these silos?

- Application-centric rather than data-centric thinking
- Not enough consistent use of data management standards
- Resource contention when it comes to cross-functional support to rectify architectural issues
- And the thinking that there are too many “immediate” issues to deal with that we can’t address new approaches
What’s happening behind the scenes?

The challenges....

We have difficulty consolidating our information silos because in many cases they were developed to support very specific business requirements. The data models, use of metadata etc. have been applied differently in each application. And this means???

**Cross platform communication and reporting is hard!**

Even with cloud-based infrastructures, and improved data staging services, the traditional relational data-base architectures make it hard to keep up with the high-paced requests for changes and improvements. Combining data and tables is time consuming.

**We can’t easily create new data mashups and trust the data quality. Especially if we want to automate the process!**

Our data doesn’t always mean the same thing across the Bank. What CEA calls a region may not be the same as COM. What HR calls position may not be how a manager understands the role.

**We’re not always sure we are comparing apples to apples!**
2. Towards a Knowledge Graph
What is a Data Fabric?

The Data Fabric Architecture

The Data Fabric

Think data drawn together from across the Bank, and industry – the whole becomes greater than the sum of its parts!

Source: *Data Fabric as Modern Data Architecture*, Alice LaPlante, 2021
A few key capabilities are needed to meet our challenges:

- We need to be able to seamlessly connect the data and allow for cross-functional and application analytics
- We need to ensure we are talking about – and meaning – the same things (concepts)
- We need to trust, when we do combine data, that the results are accurate

This is where enterprise knowledge graphs enter the scene...
Components of the knowledge graph

- Distributed data (and actually the more distributed the stronger the use case for the enterprise knowledge graph)
- Semantic metadata - what our data and connections between them mean (this comes from the taxonomies and ontologies)
- Connected data - meaning of data comes in part from its connection to other data (this comes from the technical frameworks we utilize behind the scenes)
- Algorithms – graph algorithms extract structure and infer behavior

How to Build a Knowledge Graph
Implementing LEI and MDM

The two-pronged composable architecture....

Let’s look at LEI – Legal Entity Identifier

LEI for Bank of Montreal

NQQ6HPCNCCU6TUTQYE16

Primary Key [PK]

Master List – Cities (ISO)

Master List – Countries (ISO)

Entity Legal Name
Legal Address First Line
Legal Address City
Legal Address Country
Legal Address Postal Code

Master Data Repository (MDR)

Enterprise Metadata Management Tool

Enterprise Metadata Management Tool
The knowledge graph meets these Bank requirements we’ve identified....

- Flexibility in the face of complex and changing data
- Description in terms of business concepts
- Ability to deal with unanticipated questions
- The knowledge graph is data-centric (as opposed to application-centric)
- It supports data as a product
- It supports FAIR
Ontologies

• "Formal, explicit specification of a shared conceptualization" (Guarino, Oberle & Staab, 2009)
• Machine- and human- readable
• Formal knowledge representation languages
3. Challenges
• Learning curve
• "Graph thinking"

Use Cases, Competency Questions and Documentation

- Identification
- Verbalization
- Priorization
- Foreseeing
- Resources
• Bottom-up/Top to bottom
• Level of expressiveness
  • Tool’s functionalities and limitations
• Building blocks
  • From PoC to enterprise-wide
## Semantic Modelling (Alexopoulos, 2020)

<table>
<thead>
<tr>
<th>General Elements</th>
<th>Common and Standardized Elements</th>
<th>Semantic and Linguistic Phenomena</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Entities</td>
<td>• Lexicalization</td>
<td>• Ambiguity</td>
<td>• Semantic Accuracy</td>
</tr>
<tr>
<td>• Relations</td>
<td>• Synonymy</td>
<td>• Uncertainty</td>
<td>• Completeness</td>
</tr>
<tr>
<td>• Complex axioms, Constraints, and Rules</td>
<td>• Relation Subsumption</td>
<td>• Vagueness</td>
<td>• Consistency</td>
</tr>
<tr>
<td>• Terminology</td>
<td>• Part-Whole Relation</td>
<td>• Rigidity, Identity, Unity and Dependence</td>
<td>• Conciseness</td>
</tr>
<tr>
<td></td>
<td>• Semantic Relatedness</td>
<td>• Symmetry, Inversion and Transitivity</td>
<td>• Timeliness</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Open- and Closed-World assumptions</td>
<td>• Relevancy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Semantic Change</td>
<td>• Understandability</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Trustworthiness</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Availability, Versatility and Performance</td>
</tr>
</tbody>
</table>

**Quality**
- Semantic Accuracy
- Completeness
- Consistency
- Conciseness
- Timeliness
- Relevancy
- Understandability
- Trustworthiness
- Availability, Versatility and Performance
4. A simple example

The Functional Classification Scheme and the Corporate Retention Matrix
### Functional Classification Scheme

#### Functions
- Major responsibilities to fulfil the organization’s mandate
- Core functions (set forth in legislation)
- Standard functions (managerial and enabling)

#### Activities
- Actions or processes undertaken to accomplish a Function
- Unique activities (specific to a Function)
- Common activities (common across many Functions)
Functional Classification of Corporate Records

- **FUNCTION**
  - Highest level
  - Bank mandate
  - Static

- **Activity**
  - Actions or processes to accomplish a function

- **Records Series**
  - Group of records created, captured and managed as a result of an activity
Access Control

<table>
<thead>
<tr>
<th>Name</th>
<th>Access Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numbering</td>
<td>005</td>
</tr>
<tr>
<td>Term Type</td>
<td>Activity</td>
</tr>
<tr>
<td>Scope Notes</td>
<td>Information related to ensuring that necessary physical and electronic measures are in place to ensure appropriate access for authorized employees and clients in order to safeguard Bank of Canada assets. See: SECURITY - MONITORING for information on performing surveillance to and for an organization.</td>
</tr>
</tbody>
</table>

Source of Term

<table>
<thead>
<tr>
<th>URL</th>
<th>Inactive</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date From</td>
<td>25/02/2010</td>
<td></td>
</tr>
<tr>
<td>Date To</td>
<td>28/09/2011</td>
<td></td>
</tr>
<tr>
<td>Date Created</td>
<td>25/02/2010</td>
<td></td>
</tr>
<tr>
<td>Date Modified</td>
<td>28/09/2011</td>
<td></td>
</tr>
<tr>
<td>Developer Notes</td>
<td>Carolyn Holmes - Sept 1, 2011 - ensure standard scope note Carolyn Holmes - Sept 28, 2011 - remove &quot;procedural&quot; from scope note and added cross-ref - Validation prep C. Holmes - August 17, 2017 - numbering and ensure generic text &quot;information related to&quot;</td>
<td></td>
</tr>
</tbody>
</table>
Useful Queries

• Query activities that are in use (linked to a Function or Functions)
• Query activities that are not in use (not linked to any Function)
• Query all activities and see relationships (if any) to a Function or multiple Functions and all attributes of the activity (scope, numbering, etc)
• **Full schema query (all Functions and Activities) and all metadata props**
• Function query (all functions and scopes, identifiers etc)
• Audit query – (changes to the schema)
• To check wording consistency, I’d like to check “Client Support” activity in the schema
• To check nomenclature consistency, I want to verify if “110” is used only for “Committees and Meetings” activities
• FUNCTION1
  • Activity1A
  • Activity1B
  • Activity1C
• FUNCTION2
  • Activity2A
  • Activity2B
  • Activity2C
• ...

Schema
- FUNCTION1
  - ActivityA
  - ActivityB
  - ActivityC
- FUNCTION2
  - ActivityA
  - ActivityD
  - ActivityE
- FUNCTION3
  - ActivityB
  - ActivityF
  - ActivityG
• FUNCTION1
  • ActivityA
  • ActivityB
  • ActivityC
• FUNCTION2
  • ActivityA
  • ActivityD
  • ActivityE
• FUNCTION3
  • ActivityB
  • ActivityF
  • ActivityG
• FUNCTION1
  • ActivityA
    • RecordsSeries1A1
    • RecordsSeries1A2
    • RecordsSeries1A3
  • ActivityB
    • RecordsSeries1B1
    • RecordsSeries1B2
• FUNCTION2
  • ActivityA
    • RecordsSeries2A1
    • RecordsSeries2A2
    • RecordsSeries2A3
  • ...

Schema
Model

Function contains Activity results in Record hasSeries RecordsSeries
Model
Model
• Function
• Activity
• Records Series
• Scope Notes (containing way more than a scope note)
• System of Record
• Office or Primary Responsibility
  • X
  • Y
• Legislative Requirements
• Limitation Periods
• Statutory Retention
• Business Requirements
  • 2 years for X
  • 3 years for Y
• Final Retention
• Method of Disposition
• Records Classification
• Comments
• Security Categorization
• Default Decategorization
• File Type
• Pre-1946 Records
• LAC Appraisal Framework Citation
• RDA
• BoC Archives Appraisal
Model
References


Thank you