From Acquisition to Access – Data Management at the Bank of Canada

Presentation at Beyond the Numbers, Federal Reserve Bank of St Louis
November 9, 2022
Michele Sura
Jennifer Trower
Our Organization

Knowledge and Information Services (K&IS)

Digital Library and Data Licensing Services

Information Quality and Governance

Information Management Lifecycle and Collaboration

Business Partners and Initiatives (e.g. data sensitivity assessment)

Contract Management and Licensing Compliance

Digital Scholarship, Library Systems and Metadata

Research and Reference / Alerting / Help Desk

Centralized data onboarding / Review and Governance

Research Data Management Service

Systems (Data Catalogue) and Metadata
Topics

Part 1: Data Acquisition and Licensing
- Centralized review and governance approach
- Data onboarding process

Part 2: Discovery and Access
- Data Catalogue
- Metadata schema and management
Topics

Part 3: Services for Researchers
- Research Data Management
- FAIR

Part 4: Next Steps
- Successes and Lessons Learned
- Open Access and Digital Preservation
Part 1: Data Acquisition and Licensing
Library’s Contract Management and Licensing Compliance team steer internal clients through the Bank’s data onboarding process, partnering with Data Management, Privacy, Legal, IT, Procurement.

- Market scans
- Demo, trial, samples
- Negotiation of scope of data asset, terms and conditions of use, number of users, delivery, and pricing
- Assignment of data/information security classification and privacy assessment
- Financial approval from review and governance committee
- Final procurement paperwork
- Establishing ingest methods (ideally API); internal access controls
- Provision of needed econometric software and related training
- Descriptive metadata record in Bank's Data Catalogue
A closer look at licensing compliance

- E-learning modules
- Participation in Bank’s mandatory Annual Compliance Exercise
- Ad-hoc licensing checks, pre-publication
- Focus on negotiating enterprise licensing, in particular with market data providers, when applicable and financially feasible
- Links to license agreements in Data Catalogue metadata records, along with summary of terms of use
Part 2: Discovery and Access
Data Catalogue: Vision

• One centralized, searchable catalogue of Bank’s data assets
  • Tables/charts/graphs
  • Files and sets, including 300K+ time series
  • Bases and aggregation services
  • Visualizations
• Encompassing monetary policy, research, market operations, banking operations, currency operations, corporate administration and more recently, retail payment supervision
• Discoverable via fulsome descriptive metadata – automated, when possible
• Useful search interface with facetted searching
• In late 2018, challenging to find a data catalogue platform which could accept a robust descriptive metadata schema

• Decided to use a digital library platform (TIND, commercial arm of open source Invenio), which could be expanded to include other content discoverability in the future (ILS, Archives, RDM)

• Automating the creation of descriptive metadata records a challenge due to inferior, legacy metadata
What’s in the Data Catalogue?

Resource Records
- Catalogued records for Data that has been:
  - Procurèd, Acquired, Created
- Curated Collections
  - (e.g. Subscriptions / Agreements, Securities (*internal name is Triton*), Statistics Canada, etc)
  - Across Bank data universe (research, policy, operations, corporate administration)
- Reproducibility Packages for Research Data

Authority Records
- Relational metadata records typically based on standardized metadata
  - STW Subject Headings
  - Personal/Corporate Names
  - JEL Codes
• Based on DDI (Data Documentation Initiative)

• Drew inspiration / cross-walked to DataCite, DCAT, SDMX, Dublin Core, Registry Interchange Format (Collections and Services)...and of course, FRED-S

• Ended up with a 65 field metadata schema
  • 21 mandatory fields (see appendix for list)
  • 28 controlled vocabularies / authority collections, which will be managed in the future by an Enterprise metadata management tool
### Categories of Metadata

#### Included in the Data Catalogue

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Descriptive Metadata</strong></td>
<td>- Curated human readable metadata&lt;br&gt;- Used for discovery and identification&lt;br&gt;- e.g. Title, description, access rights, location, keywords, subjects, time method, granularity, magnitude, ...65 possible fields&lt;br&gt;- links to reference materials</td>
</tr>
<tr>
<td><strong>Rights Metadata</strong></td>
<td>- Information to help users understand the conditions by which they are permitted to use the data&lt;br&gt;- e.g. Distribution, Redistribution, Republication&lt;br&gt;- With links to redacted license agreement</td>
</tr>
<tr>
<td><strong>Technical Metadata</strong></td>
<td>- Information on the technical properties or processes of the data, regardless of the focus/context of the files&lt;br&gt;- e.g. size of the incoming files, ingest method, frequency of ingest, technical contact</td>
</tr>
<tr>
<td><strong>Contextual Metadata</strong></td>
<td>- Primarily the structural, hierarchical or relational information between records and metadata&lt;br&gt;- e.g. Related resources, Collections, etc&lt;br&gt;- used by systems librarians</td>
</tr>
<tr>
<td><strong>Administrative Metadata</strong></td>
<td>- Information to help manage a resource, as well as support contacts&lt;br&gt;- e.g. Access support, data support, etc</td>
</tr>
</tbody>
</table>

*Fully indexed information about data resources...*
Data Catalogue Landing Page
<table>
<thead>
<tr>
<th>Sample Results - Haver Analytics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Creator:</strong></td>
<td></td>
</tr>
<tr>
<td>Bank of Canada</td>
<td>3</td>
</tr>
<tr>
<td>Bank of Montr...</td>
<td>2</td>
</tr>
<tr>
<td>Bloomberg LP</td>
<td>1</td>
</tr>
<tr>
<td>Board of Gover...</td>
<td>1</td>
</tr>
<tr>
<td>Bureau Of Eco...</td>
<td>1</td>
</tr>
<tr>
<td>Bureau of Lab...</td>
<td>1</td>
</tr>
<tr>
<td>Canadian Impe...</td>
<td>2</td>
</tr>
</tbody>
</table>

**Time Method:**
- Time series: 182

**Granularity:**
- Aggregated: 182
- Microdata: 1

**Unit of Observation:**
- Financial instr...: 2
- Financial instr...: 2
- Government: 3
- Object: 1
- Time unit: 182

**Observation Frequency:**
- 182

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Time series</th>
<th>Time unit</th>
<th>Aggregated</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPMED - Austria, Switzerland, Greece, Portugal</td>
<td>Contains detailed indicators from each country's national statistical office and central bank. Includes extensive data on Austria's national accounts, Switzerland's busin [...]</td>
<td>Dataset</td>
<td>Time series</td>
</tr>
<tr>
<td>AMECO - European Commission Macro Forecasts</td>
<td>Contains the European Commission Directorate General for Economic Affairs' (DG ECFIN) entire annual macroeconomic data set AMECO. These data cover a wide range of topics [...]</td>
<td>Dataset</td>
<td>Time series</td>
</tr>
<tr>
<td>ANZ - Australia &amp; New Zealand</td>
<td>Over 45,000 series covering industry detail on employment, national accounts, sales and inventories, prices and trade. Economic and business surveys are available from in [...]</td>
<td>Dataset</td>
<td>Time series</td>
</tr>
<tr>
<td>ASM - Annual Survey of Manufactures</td>
<td>U.S. statistics on 600+ industries covering 14 concepts such as spending, value added and cost of materials.</td>
<td>Dataset</td>
<td>Time series</td>
</tr>
<tr>
<td>ASMR - Annual Survey of Manufactures, States</td>
<td></td>
<td>Dataset</td>
<td>Time series</td>
</tr>
</tbody>
</table>
Sample Item-Level Record

Statistics Canada Database

### Description
CANSIM is a socio-economic database by which it shares Canadian information with the country and the world. This data is used by the Bank's economic departments and provides financial institution, financial market, alternative measures of inflations and commodity prices data.

### Details
- **Variant Title**: CANSIM
- **Resource Type**: Database
- **Status**: Active
- **Location**: FarmeMart
- **Access Note**: FAMEMART Access Required
- **Access Support**: Data and Statistics office
  \[DSOallrequests@bank-banque-canada.ca\]
- **Data Support**: Data and Statistics Office
  \[DSOallrequests@bank-banque-canada.ca\]
- **Republication**: Source Data Fully
- **Redistribution**: Bank Only
- **Related Resource**: Has Part: Cansim dataset
  
- **License Agreement**: Statistics Canada Open Licence
  
- **Information Sensitivity Rating**: Public
- **Time Method**: Time series
  - Cross-sectional
  - Longitudinal
- **Granularity**: Aggregated
  - Microdata
- **Unit of Observation**: Time unit
- **Generation Method**: Data feed
- **Subject**: Statistical data collection
  - Money and financial markets
- **Geography**:
  - Canada
  - Alberta
  - British Columbia
  - Saskatchewan
  - Manitoba
  - Ontario
  - Quebec
  - Newfoundland and Labrador
  - New Brunswick
  - Prince Edward Island
  - Nova Scotia
  - Nunavut
  - Yukon
  - Northwest Territories
- **Supplemental Resource**: How to Access Microdata (Link)
- **Creator**: Statistics Canada
- **Record Appears in**: Statistics Canada > CANSIM
  - Data Resources A-Z
  - All Data
Part 3:
Services for Researchers
Research Reproducibility - Goals

- **Manage**: Manage research data, and associated outputs, throughout its lifecycle
- **Support**: Support researchers in meeting to publishers' data submission requirements
- **Access**: Catalogue Reproducibility Packages
- **Expand**: Expand Open Access, alignment with FAIR principles
Reproducibility - Service and Support

**Curated Submissions**

- Ensuring complete and well documented packages
- Verification and confirmation with researchers
- Review Read Me, statements of availability
- Assess content uploaded by end-users

**Technical Infrastructure**

- Data catalogue
  - Metadata submission form
  - File-upload
- Repositories
- Preservation
<table>
<thead>
<tr>
<th>Reproducibility package - What is being captured?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read Me</td>
</tr>
<tr>
<td>Metadata for the Research Publication</td>
</tr>
<tr>
<td>Research Data and Software</td>
</tr>
<tr>
<td>Metadata for the Research Data and Software</td>
</tr>
</tbody>
</table>
Research Reproducibility Package

NB: access is still mediated via through RDM Team. In 2023 self serve access via RDM repository and IR

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>This data and software package contains a README file and two Excel data files on (1) the TVL of DeFi over time and (2) the decomposition in 2022 Q1.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Resource Type</strong></td>
</tr>
<tr>
<td><strong>Access Note</strong></td>
</tr>
<tr>
<td><strong>Access Support</strong></td>
</tr>
<tr>
<td><strong>Distribution Rights</strong></td>
</tr>
<tr>
<td><strong>Publication Year</strong></td>
</tr>
<tr>
<td><strong>Language</strong></td>
</tr>
<tr>
<td><strong>JEL Codes</strong></td>
</tr>
<tr>
<td><strong>Keyword</strong></td>
</tr>
</tbody>
</table>
| **Creator** | Chiu, Hing Lun (Jonathan)  
Kahn, Charles  
Koepppl, Thorsten |
| **Record Appears in** | Reproducibility > Research Data (MEDR)  
All Data |
Evolving FAIR

Source: National Library of Medicine

- PIDS
- Repository
- Preservation
- Classification and De-categorization
Part 4:
Successes and Lessons Learned
Success #1: Centralized management of external data/information resources used for research, policy and market operations

- Reduction of duplication of incoming content and effort
- Librarians connecting internal Bank clients to needed content; consolidation of data/information negotiation expertise
- Formation and facilitation of internal working groups to consolidate business need
- More efficient financial reporting
- Better prepped if vendor audit / review is requested, in particular for financial market data
Success #2:
Improvements to internal Bank comprehension of licensing considerations

- Open Access vs. free web (not really free!) vs. commercial subscription vs. no-cost data sharing agreement vs. legacy MOU
- Who can access? Share? Redistribute? Republish original/raw/source? Source/cite as inputs into a Bank of Canada calculation?
- Potential need to purge / delete when licensing/subscription does not renew
Success #3: Data Catalogue and its foundational metadata schema serve the Bank well

- Clients can see their standard / preferred data collections in the Catalogue
- Inclusion of Research Reproducibility Packages support submission to journals and future Bank research
- Ongoing consolidation of various internal lists of data resources within internal Departments
Lessons Learned

Centralized Data Onboarding process:
• Ongoing stickhandling of client expectations through Centralized data onboarding process
• Consolidated information/data budget does catch Sr. Executives’ attention and for non-economists, can be a concern

Data Catalogue and Metadata Schema:
• finding an external tool meant wading through shallow “data catalogues”
• Learning curve for some information professionals to “see” data assets sufficiently to allow full description
1. Migration of internal Research Reproducibility Package Records to external Bank website to improve the Bank’s Open Access approach

2. Need to determine: Bank platform vs. participation in Canadian national infrastructure for long-term preservation

3. Figure out best path to preserve, and maintain access to, the papers, readmes, associated files

4. Communicate and engage with several internal stakeholders on this digital preservation journey
Thank you and Questions?

Michele Sura, Assistant Director, Digital Library and Data Licensing Service, msura@bank-banque-Canada.ca

Jennifer Trower, Manager – Digital Scholarship, Library Systems and Metadata, jtrower@bank-banque-Canada.ca
Appendix: 21 mandatory metadata fields in Bank of Canada Data Catalogue

- Title
- Creator
- Description
- Resource Type
- Subject
- Keywords
- Location
- Access Note
- Access Support
- Access Support Contact
- Data Support
- Data Support Contact
- Republication
- Redistribution
- Bank Business Owner
- Time Method
- Granularity
- Unit of Observation
- Related Materials
- Geography
- Collection