

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

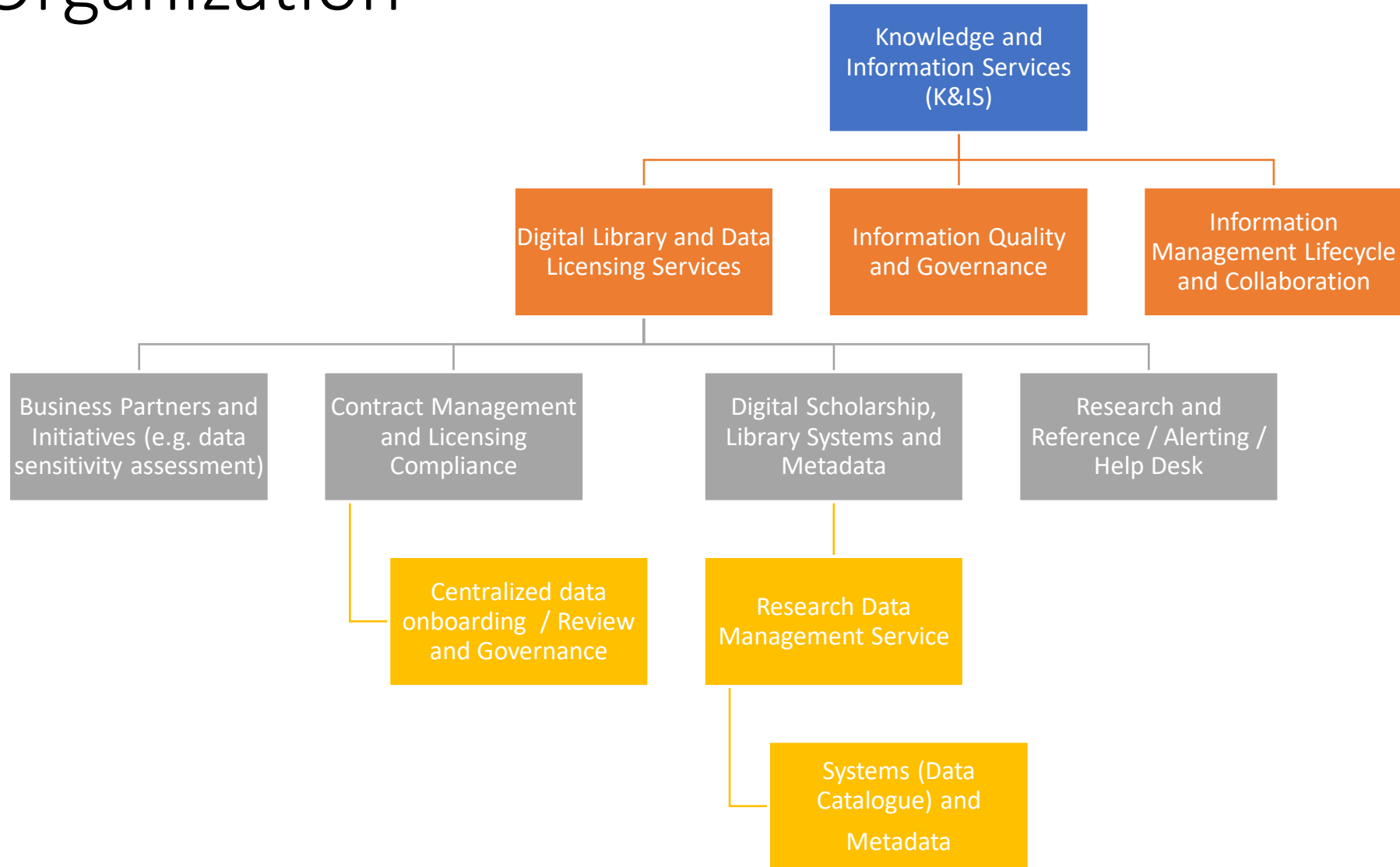
Presentation at Beyond the Numbers, Federal Reserve Bank of St Louis

November 9, 2022

Michele Sura

Jennifer Trower

Our Organization





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


Bank of Canada's Data Onboarding Process

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 faceted searching
- 



Implementation experience

- 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:
 - **Procured, 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



Metadata Schema

Descriptive Metadata

- Curated **human readable** metadata
- Used for **discovery** and identification
- e.g. Title, description, access rights, location, keywords, subjects, time method, granularity, magnitude, ...65 possible fields
- links to reference materials

Rights Metadata

- Information to help users understand the **conditions** by which they are permitted to **use the data**
- e.g. Distribution, Redistribution, Republication
- With links to redacted license agreement

Technical Metadata

- Information on the technical **properties or processes** of the data, regardless of the focus/context of the files
- e.g. size of the incoming files, ingest method, frequency of ingest, technical contact

Contextual Metadata

- Primarily the **structural, hierarchical or relational** information between records and metadata
- e.g. Related resources, Collections, etc
- used by systems librarians

Administrative Metadata

- Information to help **manage** a resource, as well as support **contacts**
- e.g. Access support, data support, etc

Categories of Metadata Included in the Data Catalogue

Fully indexed information about data resources

...



Data Catalogue Catalogue de données

Search 507,986 records for:

Search

Narrow by collection:

- ☒ All Data (507,986)
- ☒ Regulatory Returns (42)
- ☒ Bank of Canada (1,450)
- ☒ Statistics Canada (457,178)
CANSIM (452,609) Non-CANSIM (4,569)
- ☒ Triton (41)
- ☒ Reproducibility (937)
Research Data (MEDR) (937)

[Search Tips](#) :: [Advanced Search](#) :: [Browse](#)

Data Catalogue Landing Page

Sample Results - Haver Analytics

Creator:

<input type="checkbox"/> Bank of Canada	3
<input type="checkbox"/> Bank of Montr...	2
<input type="checkbox"/> Bloomberg LP	1
<input type="checkbox"/> Board of Gover...	1
<input type="checkbox"/> Bureau Of Eco...	1
<input type="checkbox"/> Bureau of Lab...	1
<input type="checkbox"/> Canadian Impe...	2

[show all](#)

Time Method:

<input type="checkbox"/> Time series	182
--------------------------------------	-----

Granularity:

<input type="checkbox"/> Aggregated	182
<input type="checkbox"/> Microdata	1

Unit of Observation:

<input type="checkbox"/> Financial instit...	2
<input type="checkbox"/> Financial instr...	2
<input type="checkbox"/> Government	3
<input type="checkbox"/> Object	1
<input type="checkbox"/> Time unit	182

Observation Frequency:



[ALPMED - Austria, Switzerland, Greece, Portugal](#)

[Haver Analytics](#)

Contains detailed indicators from each country's national statistical office and central bank. Includes extensive data on Austria's national accounts, Switzerland's busin [...]

[Dataset](#) | [Time series](#) | [Time unit](#) | [Aggregated](#)



[AMECO - European Commission Macro Forecasts](#)

[Haver Analytics](#)

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 [...]

[Dataset](#) | [Time series](#) | [Time unit](#) | [Aggregated](#)



[ANZ - Australia & New Zealand](#)

[Haver Analytics](#)

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 [...]

[Dataset](#) | [Time series](#) | [Time unit](#) | [Aggregated](#)



[ASM - Annual Survey of Manufactures](#)

[Haver Analytics](#)

U.S. statistics on 600+ industries covering 14 concepts such as spending, value added and cost of materials.

[Dataset](#) | [Time series](#) | [Time unit](#) | [Aggregated](#)



[ASMR - Annual Survey of Manufactures, States](#)

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	FameMart
Access Note	FAMEMART Access Required Statistics Canada website
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 Has Part: CANSIM PRIORITY dataset
License Agreement	Statistics Canada Open Licence Link
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 Demographic

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	Support	Access	Expand
Manage research data, and associated outputs, throughout its lifecycle	Support researchers in meeting to publishers' data submission requirements	Catalogue Reproducibility Packages	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

Reproducibility
package -
What is being
captured?

Read Me

Metadata for the Research
Publication

Research Data and Software

Metadata for the Research
Data and Software

Research Reproducibility Package

Description

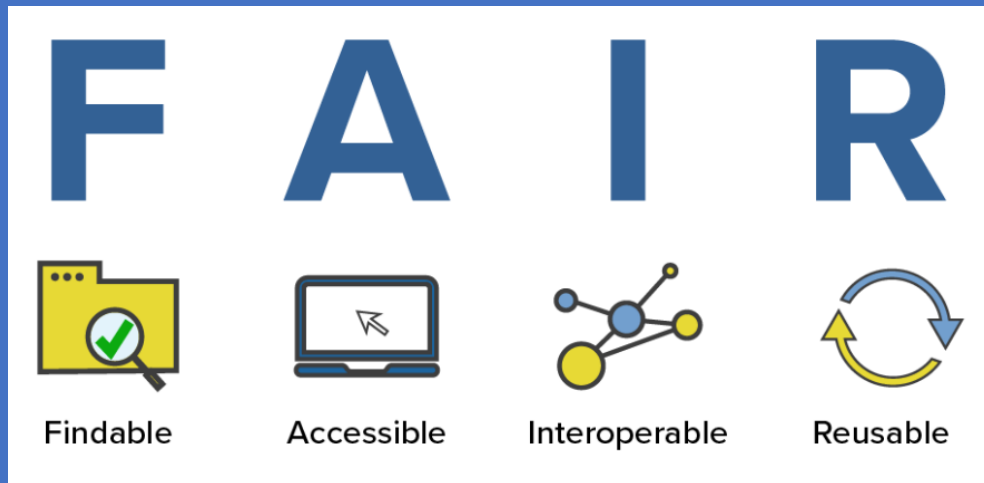
This data and software package contains a README file and two Excel data files on (1) the TVL of DeFi over time and (2) the decomposition in 2022 Q1.

Details

Resource Type	Reproducibility Package
Access Note	Contact K&IS; for access to these files
Access Support	Bank of Canada, Knowledge & Information Services RDM (MEDRSA@bank-banque-canada.ca)
Distribution Rights	The reproduction files provided for "Grasping De(centralized) Fi(nance) through the Lens of Economic Theory" can be shared publicly.
Publication Citation	Chiu, J., C.M. Kahn and T. Koepl. 2022. "Grasping De(centralized) Fi(nance) Through the Lens of Economic Theory." Bank of Canada Staff Working Paper No. 2022-43. https://doi.org/10.34989/swp-2022-43 .
Publication Year	2022
Language	English
JEL Codes	G, G2
Keyword	Digital currencies and fintech; Payment clearing and settlement systems
Creator	Chiu, Hing Lun (Jonathan) Kahn, Charles Koepl, Thorsten
Record Appears in	Reproducibility > Research Data (MEDR) All Data

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

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

Moving towards more Open Access and Digital Preservation (2023)

1

Migration of internal Research Reproducibility Package Records to external Bank website to improve the Bank's Open Access approach

2

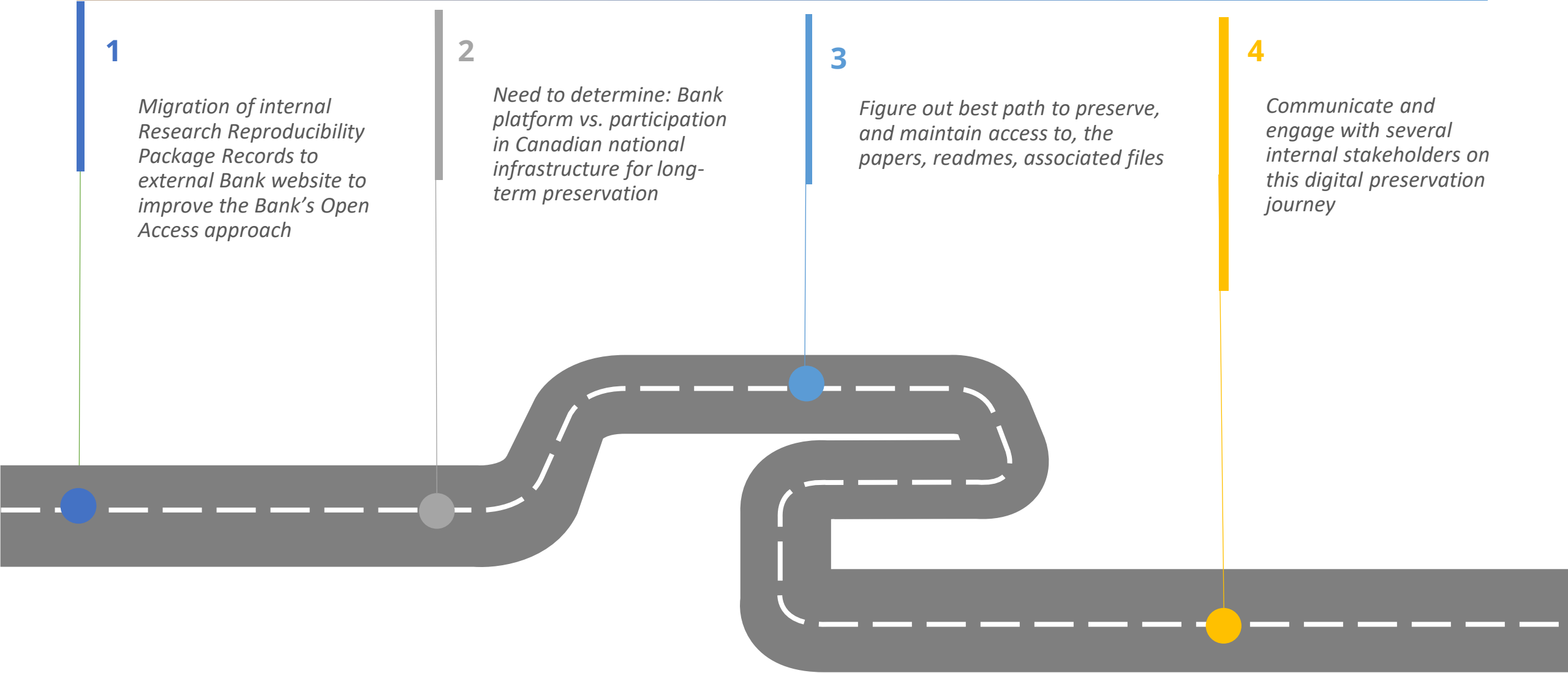
Need to determine: Bank platform vs. participation in Canadian national infrastructure for long-term preservation

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](mailto:msura@bank-banque-Canada.ca)

Jennifer Trower, Manager –
Digital Scholarship, Library
Systems and Metadata,
[jtrower@bank-banque-
Canada.ca](mailto: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