



Data architecture for statistical modernization: an integrated approach

Flavio Rizzolo

Statistics Canada



Delivering insight through data for a better Canada

A changing data space

Rapidly changing and increasingly complex economy and society



Proliferation of data and data providers



Data revolution, ingenuity and innovation



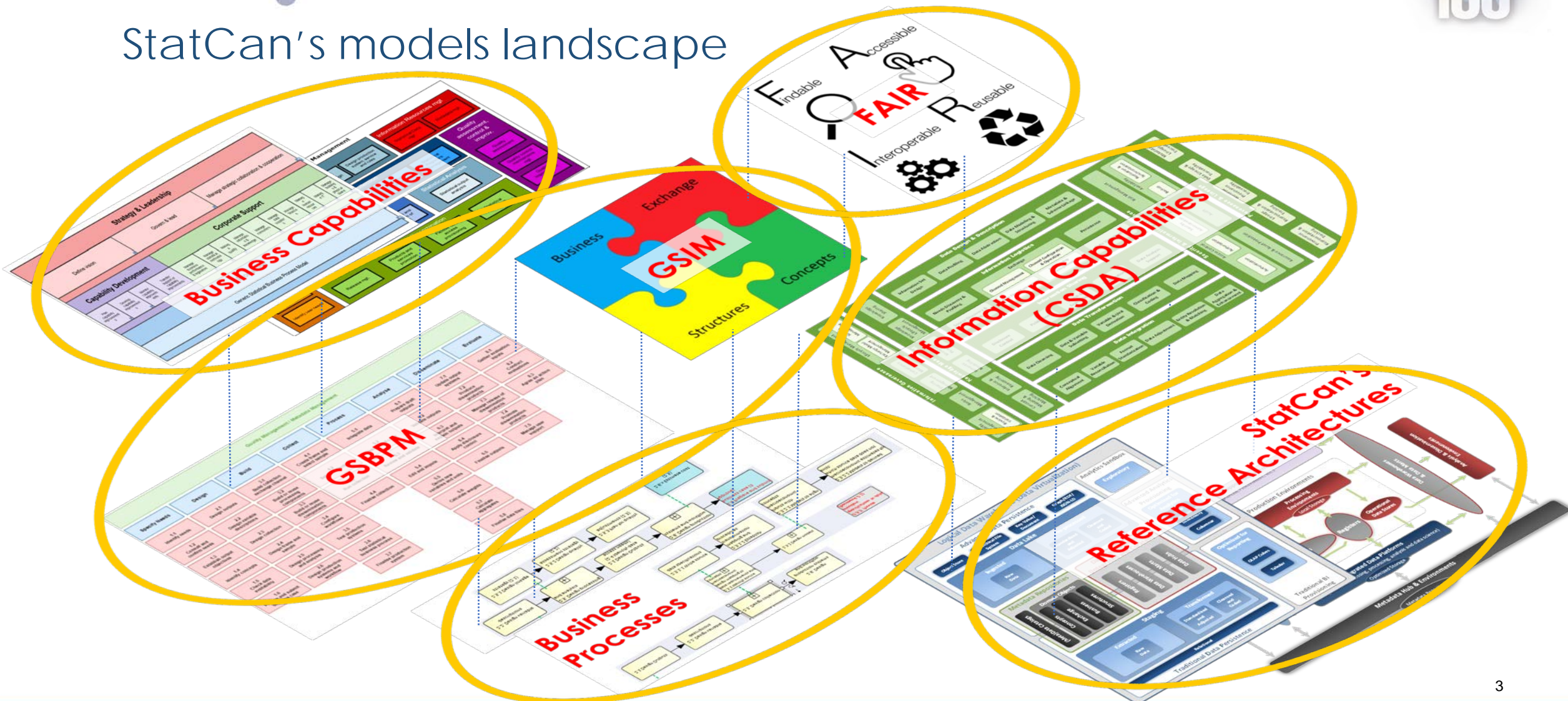
Increased expectations and demand for “real-time” and micro/detailed data



Key goals of modernization

- **Trusted** and **secure** information that is easy to **find** and **use**
 - Data provenance and lineage
 - Data management and digitalization
 - Trust framework and privacy protection
 - Semantic search and discovery
- **Timely** and **comprehensive** information that is **comparable** and **interoperable** across domains
 - Open data standards (DDI, SDMX, SKOS/XKOS, NIEM, etc.)
 - Interoperability layers (technical, semantic and legal)
 - Linked open data ontologies and vocabularies (schema.org, StatDCAT, PROV-O, etc.)
- To navigate this complexity we need to start with **standardized models**

StatCan's models landscape



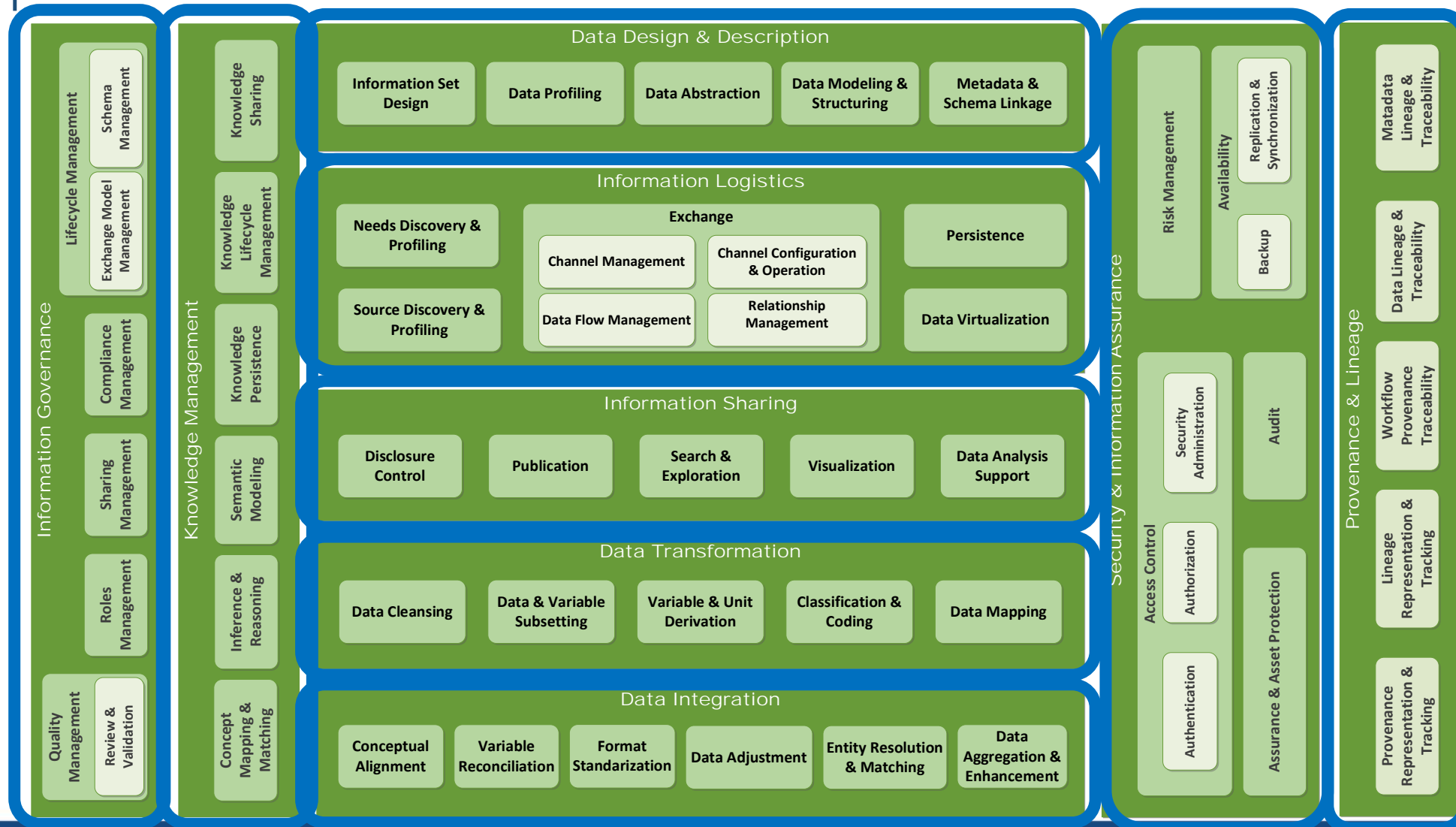
GSBPM and CSDA

- The **Generic Statistical Business Process Model (GSBPM)** provides a catalog of reusable statistical production process **building blocks**, or *sub-processes*, organized into eight conceptual classes, or *phases*
 - GSBPM sub-processes describe how we produce statistics
- The **Common Statistical Data Architecture (CSDA)** provides a catalog of **information functions** encapsulated within **capabilities**
 - CSDA capabilities describe what we do with data and metadata
- **CSDA** and **GSBPM** can both be used to develop **machine-actionable components**
 - Differences in scope, granularity and reusability degree
- Compatible blueprints for different **viewpoints**:
 - GSBPM → Business Architecture
 - CSDA → Information/Data Architecture

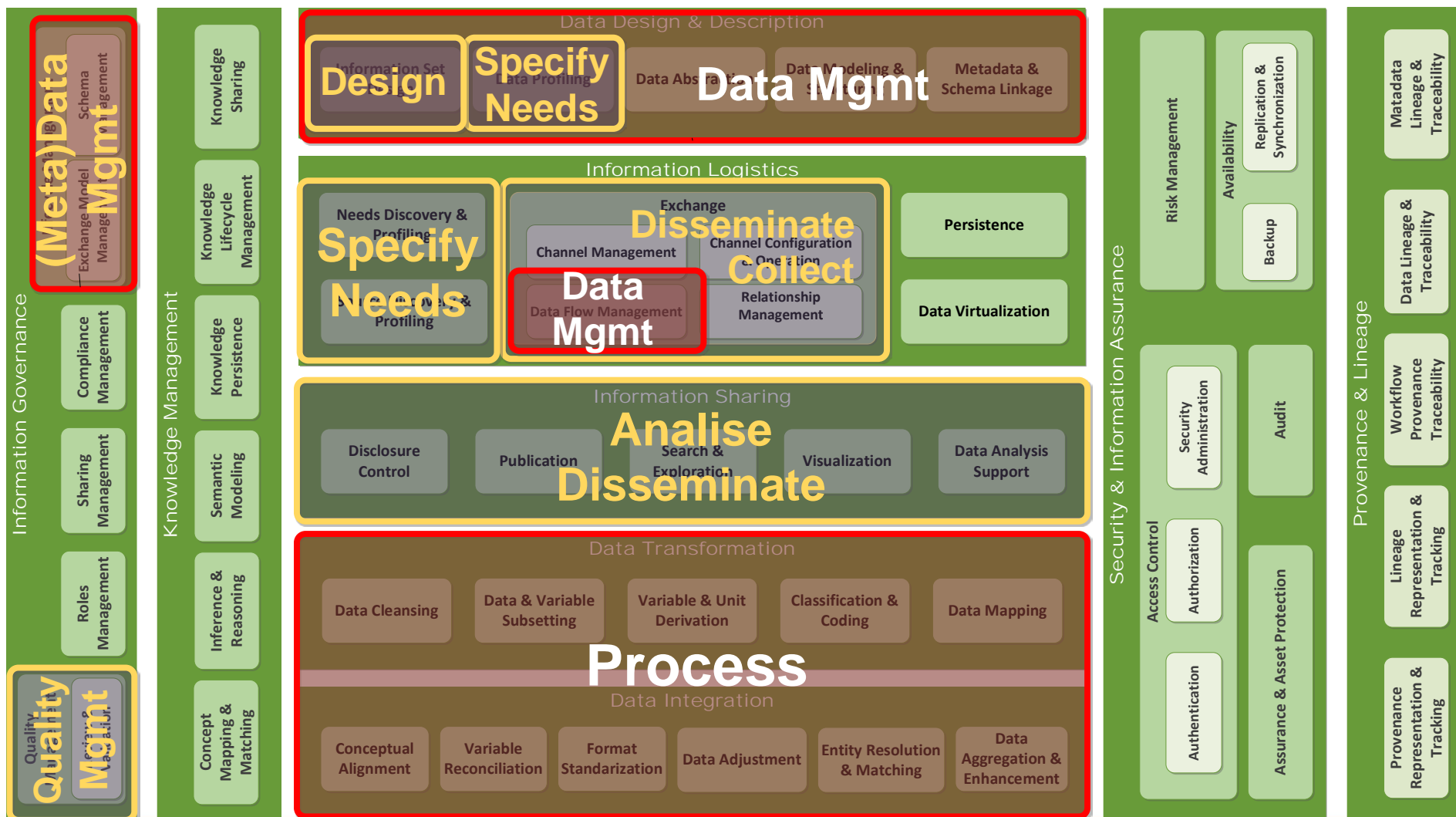
CSDA capabilities

- "Capabilities provide the statistical organisation with the ability to undertake a specific activity. A capability is only achieved through the integration of all relevant capability elements, e.g. methods, processes, standards and frameworks, IT systems and people skills" (CSDA)
- Capabilities are abstractions of "what" an organization does.
 - They are conceptual
 - They are completely separated from "how" the organization chooses to implement them
 - They are defined in terms of "what" and "why", rather than "how", "who" and "where"
 - They represent stable, self-contained business functions

Capabilities framework



Capabilities and GSBPM (preliminary)

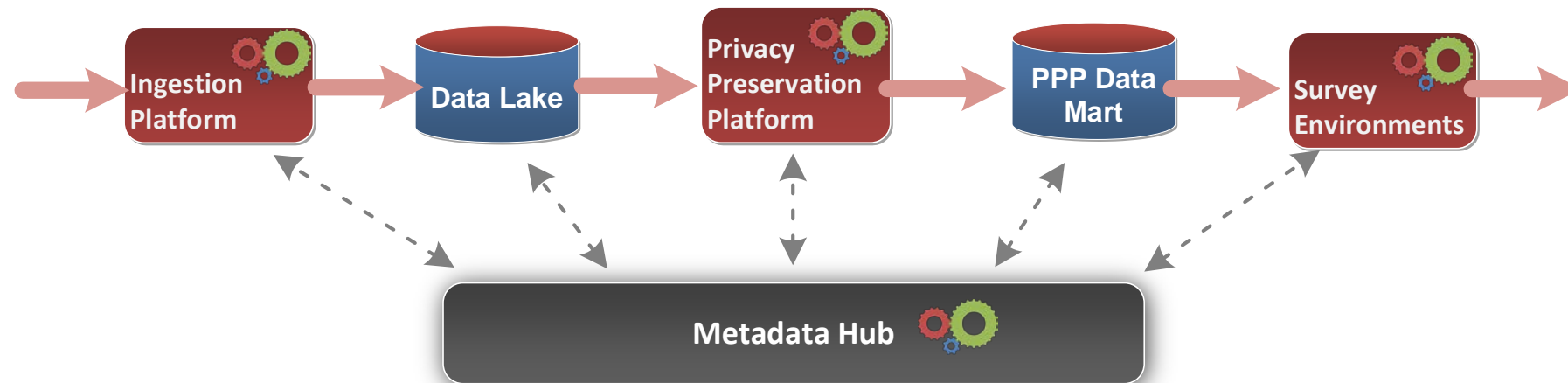


Capabilities use cases at Statistics Canada

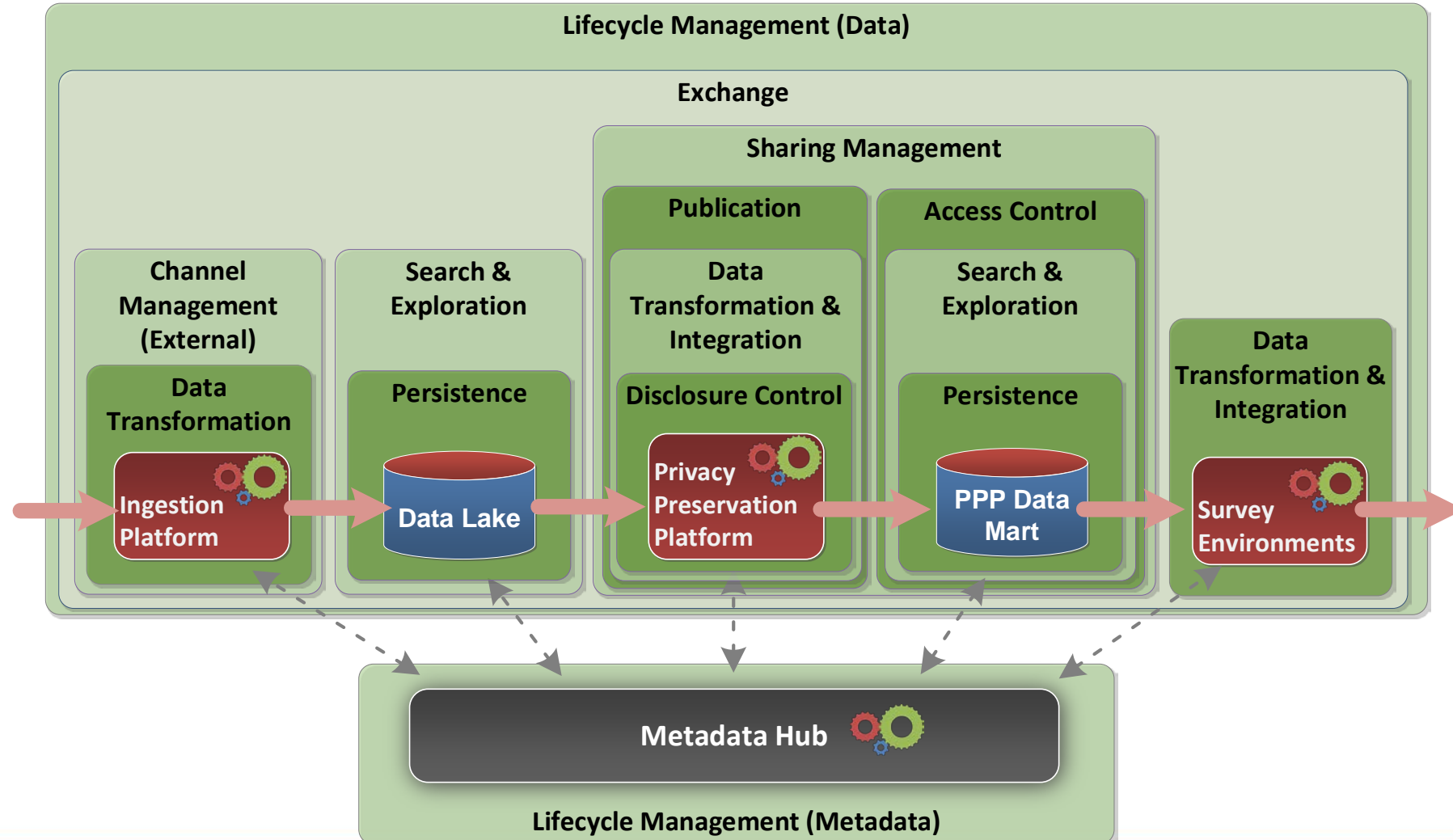
- Provide “connective tissue” between SMEs and developers
 - Serve to channel conversations with a common vocabulary between data, metadata, policy specialists and technical architects and specialists by establishing a common “canvas” or framework
- Ensure alignment with other enterprise reference architecture and guidelines to support the realization of our vision and strategies at an enterprise level
 - Map to processes, reusable software components, platforms, etc.
- Define a frame by which new projects and initiatives will be tested to ensure alignment
 - Facilitate evaluation by providing well-defined, technology-agnostic functionality descriptions

Using capabilities to characterize solutions

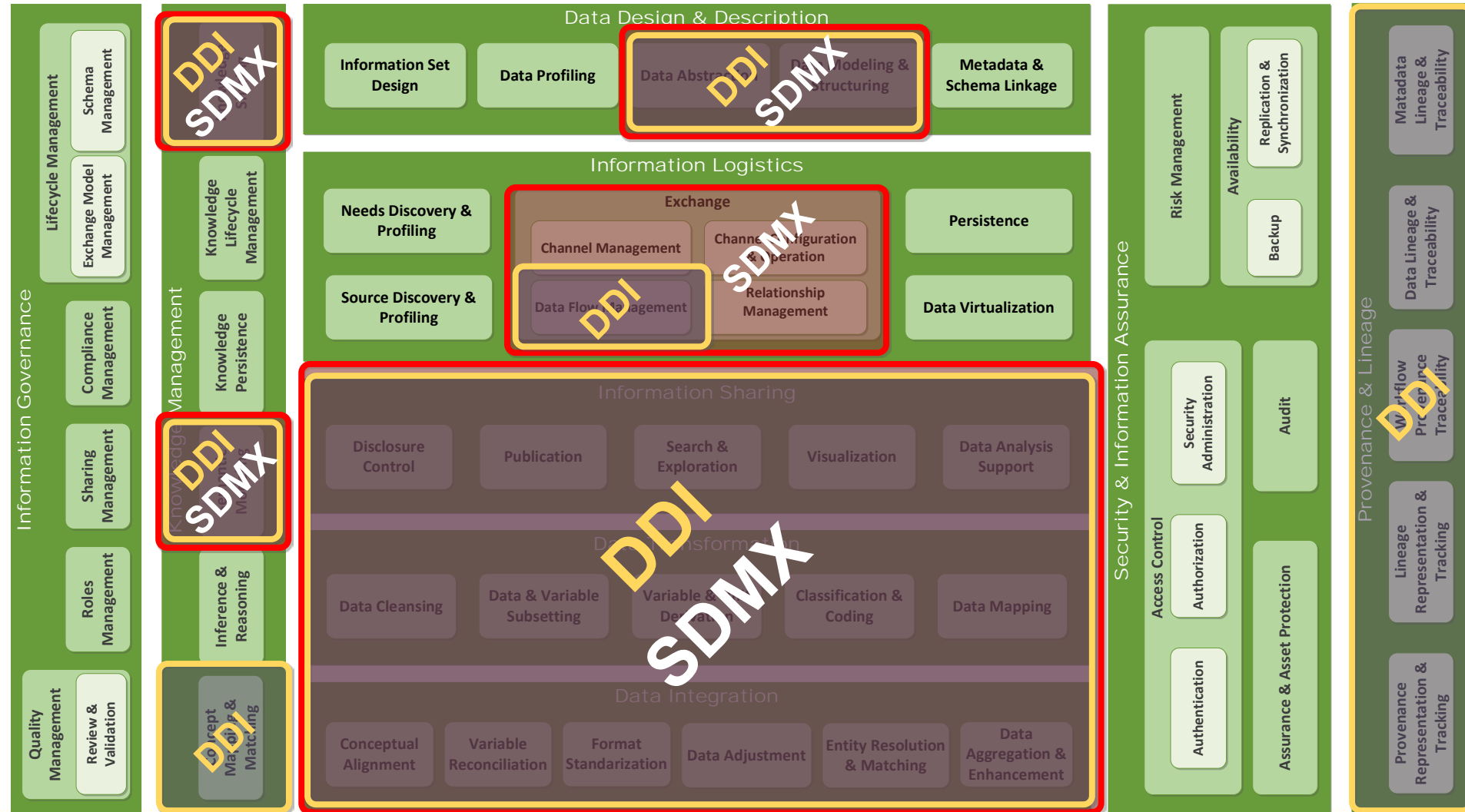
- Data coming into the organization needs to be pre-processed before being shared downstream
- Some of the data needs to be de-identified and keys maintained in a more secure environment
- Processing **metadata** is **standardized** and **managed** throughout



Using capabilities to characterize solutions



Future work: mapping capabilities to standards



Conclusions

- Models needed to satisfy multiple requirements
 - Modernization involves multiple stakeholders and viewpoints – A model for each occasion
 - Less is more, but ... “make things simple, but not simpler”
- Models need to evolve
 - Model and mapping management is a challenge
- Models need to be aligned
 - Diversity of interpretations, and implementations, get in the way of interoperability
 - Some work done – Lots still to do



Annex

FAIR Digital/Data Objects

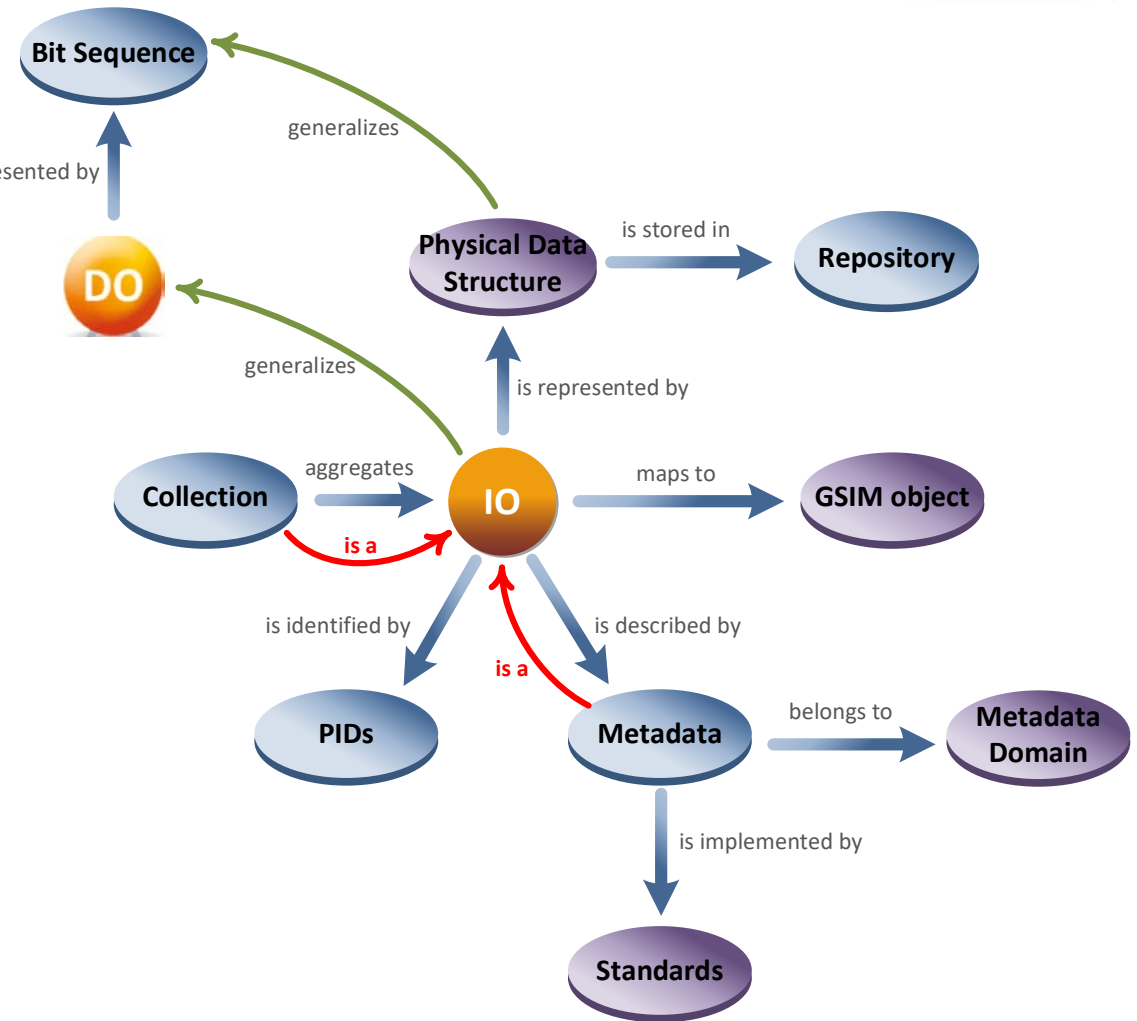
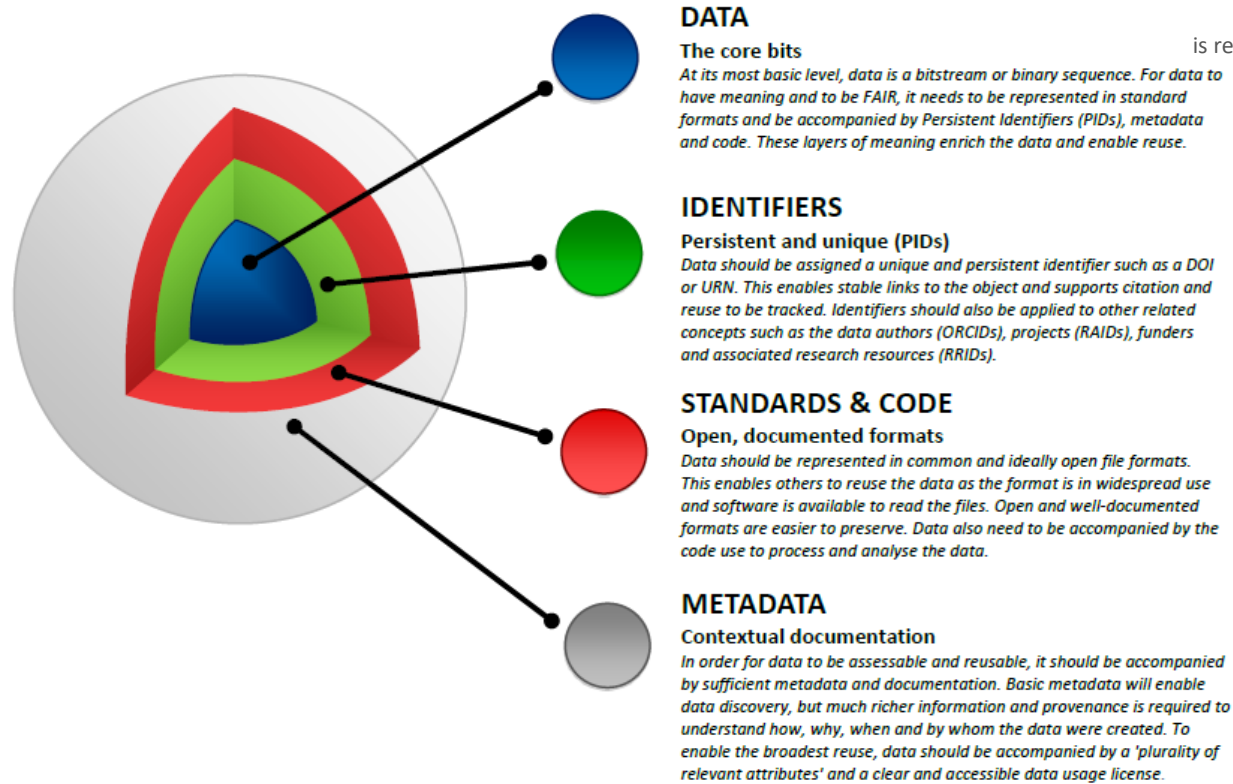


Figure 1: A model for FAIR Data Objects, noting the elements that need to be in place for data to be Findable, Accessible, Interoperable and Reusable.

Metadata standardization

- ✓ Supports quality management, information management and security.
- ✓ Streamlines statistical production, providing machine-actionable data descriptions; processes become repeatable, automatable and auditable.
- ✓ Key enabler for modernization and major project delivery.
- ✓ Facilitates data integration and usage by improving coherence and comparability across domains.
- ✓ Makes trusted and secure information easy to find.
- ✓ Improves timeliness and accessibility.
- ✓ Foundational for data discovery, harmonization, interoperability and transformation.

Ongoing work: mapping metadata domains to standards

