

Interactive and digitally-enabled statistical standards and classifications

Side event

52nd Session of the United Nations Statistical Commission

February 2021

Purpose

- Advance the use of innovative technologies and approaches for cross-referencing and navigating between the various economic statistical standards, manuals and classifications
- Implement new methodologies for managing and describing data, and the categories to which they are classified through greater uptake of semantic web technology
- Allow digital integration with well-established library and other vocabularies, taxonomies and ontologies, to improve cross-disciplinary search capabilities of digitized documents

Concept Based Classification Model

- Concepts are the building blocks for everything
- Each concept has
 - a label and agreed definition
 - relationships to other concepts allowing for conceptual frameworks to be created and making for an easier way to merge and transfer data
 - categories within the concept that are linked to other concepts like in an electronic thesauri or neural network
- Categories are stored in a category set (similar to SDMX codelist)
- Content can be dynamically updated or added to by approved users

Metadata modelling

- Enables everyone to talk about the same concept, category and content in the same way
- Matrix style approach of relationships enables linking of multiple categories reducing parent-child structures
- Allows more customised and user-defined views
- Provides flexibility for presenting content without compromising consistency
- Makes search and discovery, retrieval and interoperability easier

Example 1: ISIC - Manufacturing

Manufacturing is defined as the “*physical or chemical transformation of materials into new products in plant or factories using power-driven machines and materials-handling equipment*”

Current Concept

Concept of manufacturing is based on:

- Build a factory
- Employ staff
- Invest in tools and infrastructure
- Obtain and source raw goods and transform good into a product

Outcome

- Classify activities into the types of processes used for manufacturing or by the products produced

New Concept

Concept of factoryless production is based on:

- Identifying services and intellectual property ownership

Outcome

- Classify by type of process or product produced as per ISIC
- Increasingly difficult to identify core business or primary activity and accurately classify

New methodologies

- Requires better usage of service-oriented architecture (SOA)
- Enables uptake of the Simple Knowledge Organisation System (SKOS)
- Allows integration with metadata standards - e.g. SDMX or ISO 11179
- Provides greater usage of taxonomies, thesauri, ontological engineering and concept management to mix structured and semi-structured data
- Encourages multiple output views, different labelling options, and multi-lingual content linked to an approved concept

Simple Knowledge Organisation System (SKOS)

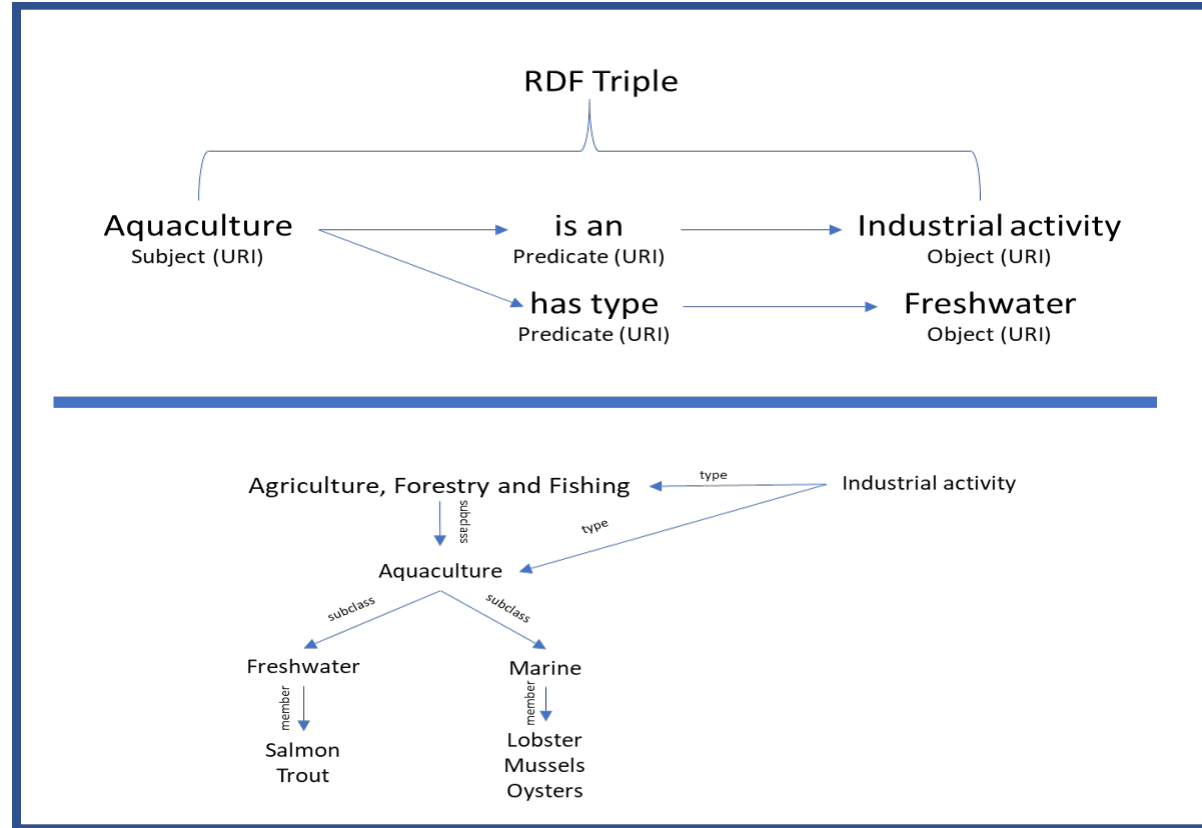
- Concepts can have multiple relationships like a neural network model
- Uses unique resource indicators (URIs) linked with lexical strings, assigning notations and links to other concepts
- Can organise content into informal hierarchies and networks using defined concept schemes
- URIs remove constraint of single descriptors or mutually exclusive labels
- Uses synonyms or aliases for categories
- Provides more granular metadata and easier integration and sharing of concepts and content

Resource Description Framework (RDF)

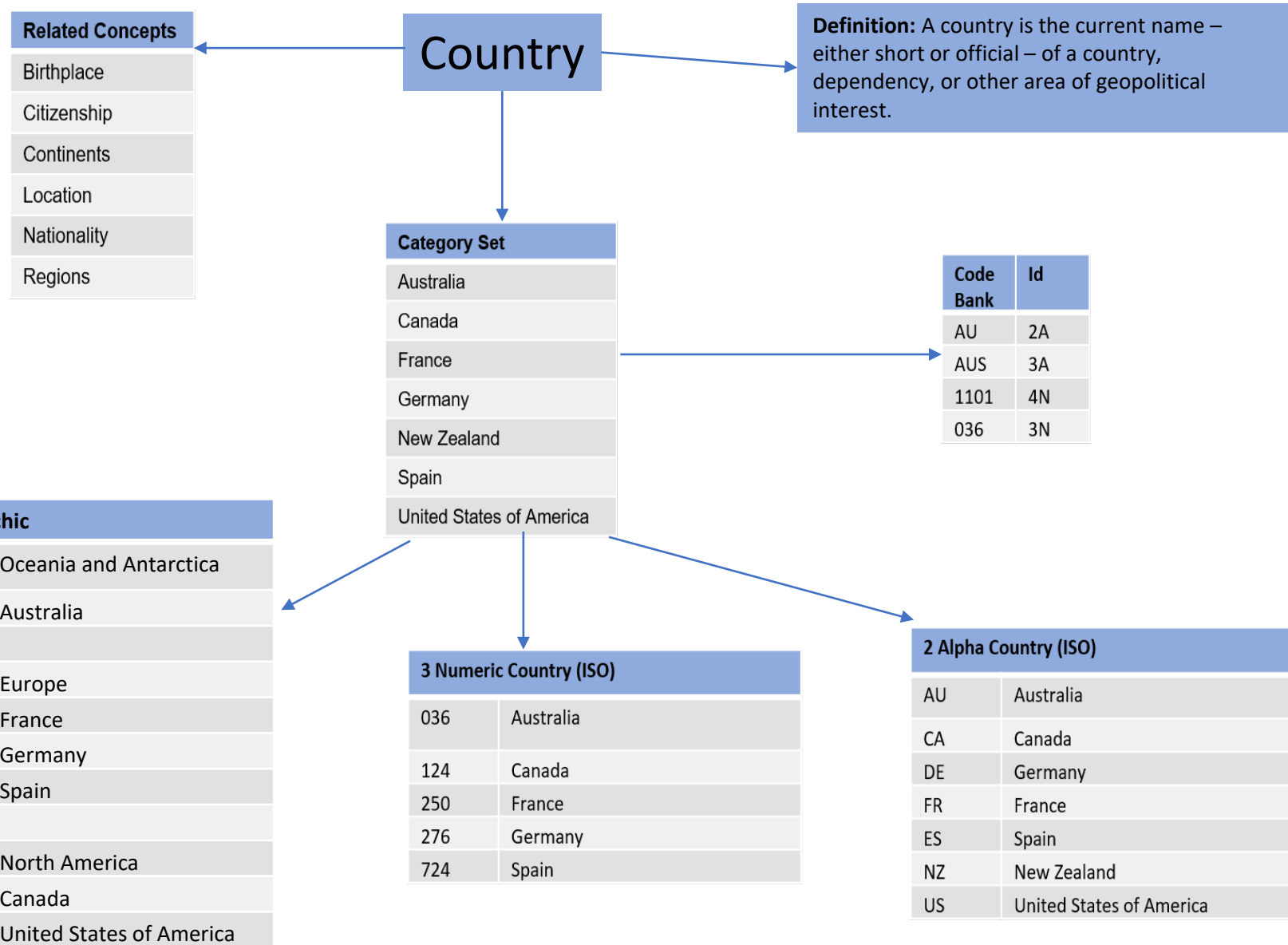
- Uses unique web identifiers for describing resources or entities
- Uses the RDF triple which comprises a subject (web resource), predicate (property) and an object (value)
- Allows classification content to be disassembled into component parts for easier integration and sharing
- Enables reconfiguration or repackaging into traditional content or user defined views
- Utilises graph networks or query language (such as SPARQL) to retrieve and manipulate data

Example 2: RDF

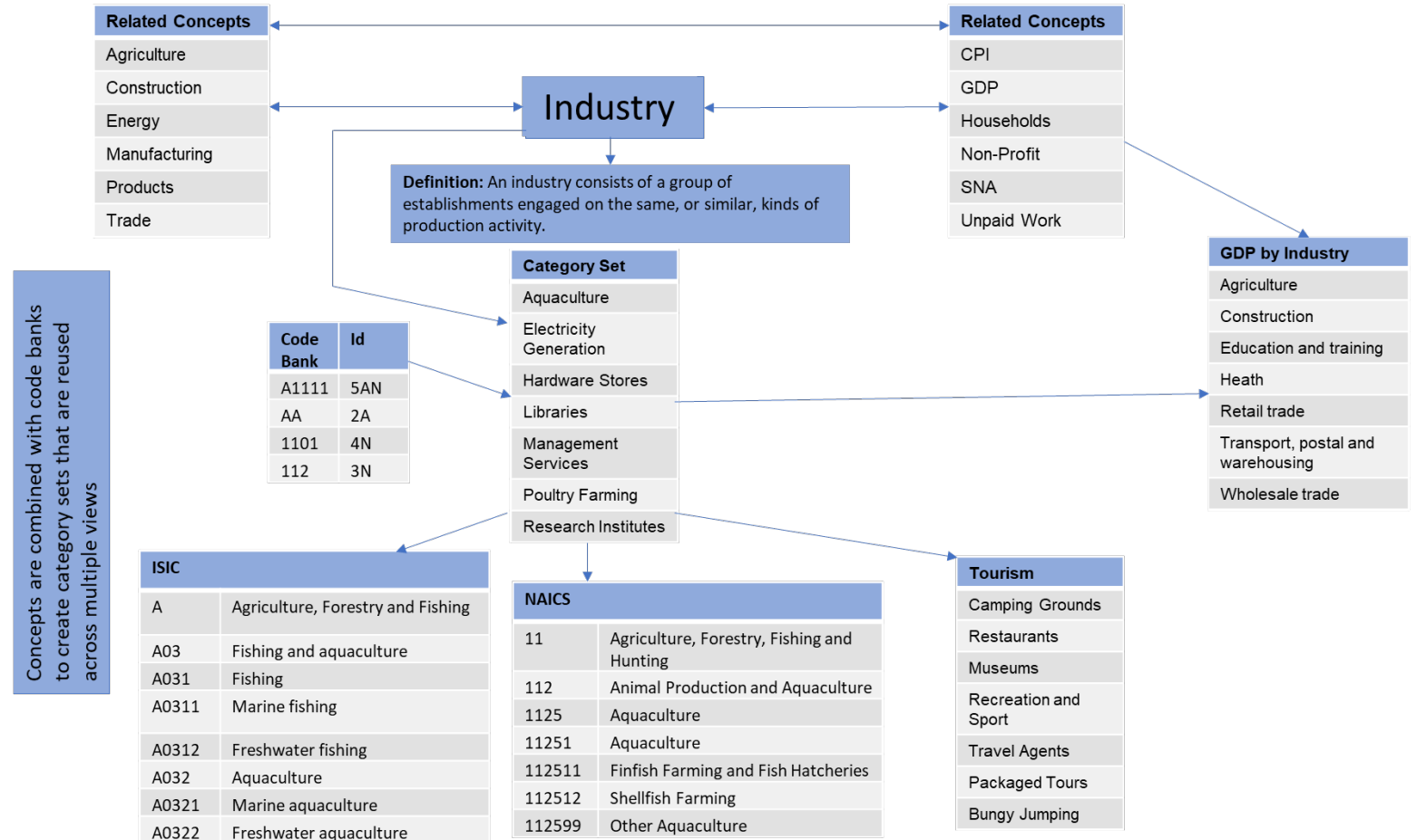
Example - Aquaculture



Example 3: Concept view - Country



Example 4: Concept view - Industry



Conclusion

- Moving to a metadata modelling approach for economic standards will take time
- The benefits are cost-reduction for national statistical offices, better resource utilisation, and greater responsiveness to user demand
- Ultimately it will make the economic standards more integrated, consistent, relevant and easier to use
- Use of semantic web and metadata modelling is the most practical way forward

Interactive and digitally-enabled statistical standards and classifications

Side event

52nd Session of the United Nations Statistical Commission