Classifications and Indicators: Modernising the Statistical Business Registers

Andrew Hancock[1]
Statistics New Zealand, Christchurch, New Zealand

Abstract: Economic data and statistics are becoming more accessible, more complicated, and open to greater and wider interpretation and application. Whilst the System of National Accounts (SNA) or the Balance of Payments manual (BPM) may heavily underpin the theory, concepts, definitions, and outputs used to produce economic data, there is still a need for business outputs to identify their target populations and respondents to utilise these massive frameworks. Statistical Business Registers (SBRs) provide the underlying statistical infrastructure including the organisational and business statistical units and their characteristics, facilitating the integration of administrative data and conducting of business surveys, in the production of economic statistics.

The classifications and indicators that underpin the SBRs do not always keep up with the contemporary reality of the economy thus affecting how businesses are classified within the SBR and therefore any data that may be produced. In the New Zealand context, classifying industrial activities is complicated by the use of a joint Australian and New Zealand classification (ANZSIC), a business industry classification (BIC) utilized by the Inland Revenue and Accident Compensation Commission, and the Stats NZ industry output classification (NZSIOC). Keeping these current is challenging enough but adding new indicators such as recognition of Māori businesses to the SBR

[1] Author: Andrew Hancock, Statistics New Zealand, Private Bag 4741, Christchurch, New Zealand
Email: andrew.hancock@stats.govt.nz

highlights the challenges of consistency, time-series, and relevance in a fast-moving information dependent society.

Metadata modelling along with the use of semantic software tools enables significant advances to be explored in the way that traditional statistical classifications are developed, maintained, updated, and implemented for and in Statistical Business Registers.

New methodologies for managing and describing data, and the categories to which they are classified can benefit from a greater uptake of semantic web technology, such as Simple Knowledge Organisation Systems (SKOS), and Resource Description Frameworks (RDF).

This paper explores new approaches to statistical classifications and their role in the future of SBRs through the use of metadata, conceptual and entity modelling rather than the traditional methodology of hierarchically structured, sequentially code based statistical classifications.

1. Introduction

Statistical Business Registers (SBRs) provide the fundamental infrastructure for the production and output of business and economic statistics as they contain a comprehensive list of businesses and other undertakings engaged in the production of goods and services.

The information they hold about a business may include the name, address, ownership links, number of persons employed, and codes or classification categories indicating industry, location, institutional sector, business type and the nature of overseas transactions. In addition, the SBR provides a framework for selecting survey respondents for business surveys and enables consistent classification of data collected in those surveys.

The classification of the data they hold however, is very dependent on the alignment with the key economic frameworks such as the System of National Accounts (SNA), the Balance of Payments manual (BPM) or System of Environmental-Economic Accounting (SEEA). These frameworks, as a starting point, provide the taxonomical and conceptual basis for information management, data description and the production of official economic statistics. It is also essential that the SBR utilises statistical classifications such as the International Standard Industrial Classification (ISIC), or the Central Product Classification (CPC) to enable consistency in the presentation of data and information.

But SBRs, along with other users of statistical classifications, are faced with the challenges that those classifications bring, that is, they are difficult to use, integrate, maintain and update across statistical systems or information management systems because of their traditional hardcopy nature. SBRs, and economic statistics, need to take account of the rapidly moving world that now exists. Instantaneous information availability, the use of social media tools such as Instagram™ or Twitter™ reinforces the fact that human communication and interaction is changing and staying with the current methodologies and infrastructure for SBRs does not keep pace with a changing economy. Whilst traditional approaches and frameworks provide consistent time-series they don't enable contemporary data to be created that will influence policy and decision-making. Further they do not enable systems and data to be easily and more dynamically updated to take account of real-world changes in industries, business types or institutional sector.

As national statistical offices grapple with expenditure on the maintenance and upgrading of legacy IT systems and move towards more cloud-computing, as a service technology and greater use of application programming interfaces (apis), they also need to investigate ways to integrate and identify new data for the global indicator framework and 2030 Agenda for Sustainable Development. The traditional approaches, as well as the new approaches being identified by national statistical offices are facing greater and greater scrutiny and evaluation. This coupled with the increased need for innovative solutions, smarter SBRs, real-time response to information collection and interpretation, highlights that new and more dynamic methodologies for integrating classifications and indicators into SBRs are necessary to ensure that economic statistics are timely, fit-for-purpose, relevant and cost-effective.

In the New Zealand context, the enactment of the 2022 Data and Statistics Act also changes the way in which data and information from the SBR can be disclosed to data users outside of the national statistical office. Section 39, subsection 2 (c) states that the Statistician may publish or otherwise disclose “*data about a business entity that the Statistician is satisfied is non-sensitive, including (but not limited to) the name, address, and contact details of the business entity, its location, its business type or structure, the industry and sector type it belongs to, and indicators of the size of the business, such as the number of employees*”. This places greater emphasis on ensuring that the classifications and indicators embedded in the SBR are current and reflective of the real-world economy.

1.1 Background

There is great expense in time and resource to update a SBR every time a new version of a statistical classification is created, or a new statistical classification is developed, and this, in both instances, negates and limits the ability to reflect the contemporary reality of the economic data. The SBR is dependent on grouping and organising information and data using common criteria, concepts, categories, and definitions using statistical classifications and frameworks. Changes to frameworks such as the System of National Accounts, and its functional classifications, has impact both in cost of reconfiguration of an IT system, but also in terms of managing time-series potentially through dual-coding against two industry classifications (old and new). This coupled

with the need to ascertain viable implementation timelines and the integration issues with other agency IT systems, for example the national tax system, poses numerous challenges which traditional classification processes clearly hamper.
Associated with these traditional challenges is the added demand for better and more detailed supporting metadata, which is not a strong point of statistical classifications historically, and as a result difficult to provide with SBR data

Metadata is usually described as information about information and often presented or characterised as either structural, reference, descriptive or administrative metadata. Metadata helps provide a common understanding of the meaning of data or the semantics of it. There are many metadata standards in use internationally and within and across SBRs, such as the Statistical Data and Metadata Exchange (SDMX), ISO 11179 Information Technology – Metadata Technologies, or the Data Documentation Initiative (DDI). These standards are based upon conceptual or entity models that chunk content into component parts for easier understanding, usage, and consumption, which is no different from a SBR or statistical classification. The challenges lie in ensuring the entities are described and presented in a consistent way and some of that will be addressed in the future through greater digitisation of the SNA or BPM, and the greater system integration that the use of cloud-based apis brings for SBRs and classification management systems alike.

For much of the metadata that supports economic statistics, it has been created for human consumption and understanding, and not so much for system-to-system interaction. When information is extracted from a statistical business register it may not be as extensive as it could be nor does it necessarily provide descriptive text that may clarify a classification category, concept, or particular treatment of a classification issue. Like statistical classifications themselves, the sharing of information across concepts and data within a statistical business register is limited by the traditional methodologies used for development, maintenance and storage of that information.

But whilst statistical business registers, statistical classifications and conceptual or reference metadata models are not that dissimilar, they have unfortunately suffered from being viewed as stand-alone entities rather than as an integrated package. They have also suffered from the mindset that everything is for hardcopy publication and a need for hierarchic structures with parent-child category relationships. They all hold concepts, definitions, codelists, entities, categories, and other similar attributes. They can all be broken down into component parts but unlike the statistical classifications and metadata models, the SBR are not that flexible due to their primary structural reliance on an economic statistical units model of enterprise, kind of activity unit and location.

The lack of dynamic system integration between the classifications and the SBR highlights and reinforces the wider realisation that traditional approaches to classifying information have to change. Human intervention is gradually being replaced with machine learning and automation, and the way in which SBRs consume statistical classification information needs to change. The process is still very much about a classification like industry or institutional sector being revised by the classification experts, and then either through a bulk load function or a person entering the information, the SBR is updated.

To better share ideas, issues and economic data requires standardised approaches but the current suite of economic classifications and frameworks have a perceived rigidity, dated content and inflexibility combined with an overlap in content, concepts and purposes which hampers their usability within a statistical business register, and hampers the ability to modernise.

2. Why do we need to modernise classifications for economic statistics?

The complex nature of many economic frameworks and the extensive time it takes to make change and redefine classification content means that users are increasingly struggling to appropriately use those same statistical classifications and understand what the data means. It also inhibits the ability to easily maintain and advance content within a statistical business register as the classification revision cycles are often not in alignment with SBR redevelopments nor appropriately consider the implementation issues for SBRs. The simple issue of introducing new classification codes, indicators, and/or category definition is not straightforward, but also dependent on the ability to recognise new businesses that may be classified to such new codes.

As an example, the revision of International Standard Industrial Classification (ISIC) will see a significant number of new and emerging industries (or economic activities) added to the classification, particularly when there have been significant advances in the financial sector and new activities in the environmental sector. Concepts such as manufacturing have changed with the development of factory-less production and this means that businesses may to be reallocated to new classification codes on the SBR. If a business is widely diversified in its nature, it becomes increasingly difficult to identify their core business or primary activity particularly if the focus changes to identifying services and intellectual property ownership from traditional production processes. The static and monolithic nature of statistical classifications updated on a cyclical basis are already out of date when they are published. What does this mean for a SBR and does it actually matter? A national statistical office may be happy to code new businesses to old industry or institutional sector codes but it may also be contingent on whether any new businesses are sourced from tax data and therefore the codes allocated within the tax agency may not be able to be changed – the inherent problem of human consumption rather than integrated machine-to-machine consumption and potential legacy and incompatible IT systems. Currently in the New Zealand context the industry classification which underpins the SBR, the Australian and New Zealand Standard Industrial Classification (ANZSIC), has not been revised since its release in 2006 thus affecting the accuracy of the kind-of-activity units that may be allocated to a business. That then impacts the allocation of businesses to specific surveys if they are not appropriately classified, and ultimately the data that may be released about a business from the SBR.

3. The vision for change

The vision for change is very much about a transitional approach to a new future that enables dynamic and real-time updating of classification content and its integration and assimilation into a statistical business register. National statistical offices and data users need time to understand and accept new ways of classifying economic statistics and how new processes will enhance the delivery of information.

As stated earlier, statistical business registers and statistical classifications are both concept based. An SBR may contain multiple concepts whereas as statistical classifications are usually developed around one main concept for example business type or institutional sector. What is important about the use of concepts is that it allows for greater relationships to be created between relevant attributes. Further it enables a more efficient and automated authorisation and dissemination process, and greater search and discovery of classification related information. Treating the storage of classification separately, as per an SBR system, will enable greater reuse and reduce duplication because of the modernisation that metadata modelling brings.

Data is now able to be sourced from tools, systems and processes that didn't exist 10-20 years ago. ATMs, Global Positioning Systems (GPS), mobile phones, supermarket scanners, internet activity and social media highlight the variety and volume of new activities and ways of describing entities and categories for which current classification revision models cannot keep pace with.

It is now an appropriate time to consider the use of relational databases, advances more widely in ontological engineering, semantic web technologies, computer created matrix software and innovative classifications managements systems such as the Stats New Zealand Ariā system as methodologies for improving the way statistical classifications are developed and maintained for, and integrated into, statistical business registers. Greater use of external facing apis will enable systems to consume and share information, particularly at a more granular entity level than previously possibly with stand-alone classifications embedded into a single repository such as an SBR.

Concept-based classification management approaches such as that being developed and trialled by Statistics New Zealand can add value to data by increasing the content and metadata that can be created and associated within a classification category. An advantage to this approach is that it allows for greater integration of administrative and statistical concepts and it becomes no longer necessary to focus on standardising everything to a single economic classification or standard. Semantic consistency across classification entities can be introduced and replace the current way of developing and maintaining economic statistical classifications, and how they are integrated and used within a statistical business register.

A concept-based classification management approach is the vision for change in New Zealand as the cyclical review process for producing statistical classifications has become more labour intensive and costly in time and resource, coupled with changing strategic drivers and national data needs. In the New Zealand context, the primary classification supporting the statistical business register is the Australian and New Zealand Standard Industrial Classification (ANZSIC). Having two countries, namely Australia and New Zealand, collaborate over a multi-year project to produce a classification which takes years to implement across the business surveys and national accounts, and which is out of date on release defeats the purpose of producing a regional classification. In addition, the determination of businesses and the allocation of their industry code is done on a country specific basis, and both countries implement changes to their statistical business registers independent of each other. At present the two countries are also divergent in their future thinking around classification management meaning that New Zealand will take an independent and different pathway from Australia and seek to pursue its visionary approach to classification management. This will have significant impact and benefit for the statistical business register.

For New Zealand the vision for change is about understanding the global processes for statistical classification development and identifying simpler and easier ways for a wider community of users to contribute to content and usage. The outcomes are designed for the New Zealand national data system, and not exclusively for Statistics New Zealand, and will take advantage of the changes being introduced with the enactment of the 2022 Data and Statistics Act. It is also about recognising that there are different classifications being used across the data system which may be related but not intuitively or dynamically linked, thus affecting data quality, consistency, and application to the statistical business register.

3.1 Metadata Modelling

Metadata modelling provides a new way of thinking which begins with a clearly defined concept, which may then have relationships to any number of other concepts or sub-concepts. Each concept is unique and forms a scope for all the entities or words that may then be categorised by that concept. This is effectively how a statistical business register works by linking together the different classifications and concepts it needs to describe any given business, and its activities.

This leads to the use of entity-relationship models and relational database thinking as a way forward for how economic classifications can be better developed and then integrated into a statistical business register.

An industry classification will generally define manufacturing as the physical or chemical transformation of materials into new products in plants or factories using power-driven machines and materials-handling equipment. The concept also includes the transformation of materials into new products. Consequently, the industry classifications need codes and categories to differentiate between the various types of manufacturing activities that exist, and then these are assigned to a business on the register. But usually, it is one industry code for each business based on its primary activity and this devalues the potential data if the business is engaged in multiple activities (and at multiple locations).

For example, if manufacturing includes the transformation of materials into new products, does this mean that all transformation of materials constitutes manufacturing? No – Logging, whereby a tree is cut down and turned into logs is not manufacturing. However, when the resulting logs are transformed into building frames or furniture it is considered as manufacturing. If the same business is undertaking both activities what is the industry code to be allocated for both business data and on the SBR? There is a conceptual relationship between the two activities as the second activity cannot be done without the first occurring, but the nature of statistical classifications, and especially the ideal of mutual exclusivity, prevents that relationship being shown without significant cross-referencing or use of inclusion/exclusion text. This leads to more categorisations to explain the relationships that exist in the real world thus making the classifications more unwieldy and difficult to use and requiring more structure to take account of code patterns, levels, and categories (including residuals).

A better approach is to treat entities or activities as stand-alone concepts that can be brought together and from which different thematic views of the economy can be created, but which still can be aggregated to give an overarching view of 'industry' or 'the economy'. It is about turning classifications on their side and removing parent-child relationships and replacing with familial relationships. However, the challenge of how to integrate this approach with an IT system such as a statistical business register is where the greater use of apis and semantic web technology can assist.

Whilst the traditional approach to developing statistical classifications may seem intuitive, the heavy text focus and the limitations of code patterns and sequences makes their maintenance and updating difficult, costly, and time-consuming, coupled with the implementation into statistical business registers. A classification has to be revised as a whole, or incremental adjustments made which take time and result in another version or edition of the classification. This is not a suitable approach for statistical business registers and introduces challenges for versioning and hard coding of classification related content. It is also hampered by economic classifications that are stored and used in a SBR being revised in isolation of each other. It makes sense to review an industry and products classification in parallel with each other, but business type or institutional sector are not directly related to other classifications so there is always a tension of when to introduce new classifications into the SBR in terms of IT system maintenance versus time-series maintenance for the data it holds.

Concepts become increasingly important as regardless of whether a user is talking about information off a SBR or data coded to an economic classification, everyone needs to be talking about the same concept, categories in the same way. Often the data may just be presented as a code but what does that code represent? Even if there is a label with the code does it cover the same thing?

In the New Zealand context, the category of Dairy Cattle Farming can be presented three ways: ANZSIC 016000 Dairy Cattle Farming, NZSIOC AA131 Dairy Cattle Farming, BIC A016010 Dairy Cattle Farming. They look the same, they all come from industry classifications but can the user be confident that the scope, coverage, and context is the same. This is why more emphasis needs to go into the definitional text and labelling rather than the codes, which are solely there as placeholders for data, and defined by a classification code sequence and structure. Having different codes for essentially the same entity creates risks and overheads for SBRs. Should the SBR hold three separate codes and labels or just hold the label with three separate codes linked to it?

Moving to a matrix style approach of relationships formed by linking multiple categories together and away from the traditional parent-child classification structures will enable more fit-for-purpose views of concepts and provide users with greater flexibility around their data, without comprising consistency. However, replicating this in a SBR conceptual or physical model needs consideration to ensure information and data is consistent.

4. The use of new methodologies

Changing from a traditional approach to the development and maintenance of economic statistical classifications will enable the use of new methodologies such as service-oriented architecture (SOA) to allow the integration with other system components or platforms, especially when utilising a cloud environment.

To evolve statistical classifications requires the consideration and uptake of the Simple Knowledge Organisation System (SKOS) and/or integration of other systems such as the Statistical Data and Metadata Exchange (SDMX), ISO/IEC 11179 or the Generic Statistical Information Model (GSIM). This will enable the classification developers to apply more flexibility through the use of classification components, structures and views than was previously possible. These international metadata standards provide information models which specify concepts, relationships, rules, and other elements that are not dissimilar to the content of statistical classifications and SBRs.

Introducing these different methodologies enables better usage of taxonomies, thesauri, ontological engineering, and concept management ideas to give new insights for how economic statistical classifications and standards can be developed and maintained. This has benefit to the relevance of data and content stored and classified in the SBR.

SBRs have a wealth of content and attributes about a business which are often difficult to extrapolate or use because of the stand-alone classifications in use, and the lack of integration between those classifications. Further statistical classifications currently do not easily enable the production or provision of multiple output views or labels because of the hardcopy, printed page psyche. Occasionally an aggregated thematic output view may be created for a narrow or specific purpose, for example, ICT industries but this is not readily integrated across and in systems. The New Zealand context highlights this in that the industry classification (ANZSIC) acts as an input only classification and all outputs and aggregations are then presented on the related output classification structure of NZSIOC. Ideally these should be one integrated classification and as such the ability to produce data on working or published industries is separated from the main classification. The output classification also uses different code patterns and because of the nature of the New Zealand economy, specific privacy and confidentiality rules are applied to ensure major business entities are not obvious in the data.

Moving to a SKOS or RDF approach for building, maintaining, and storing classification content, ideally in the cloud environment, from a central repository gives the flexibility and options needed to modernise. Use of apis then allows the SBR to dynamically interact with the concept/classification repository to consume the relevant information without the SBR have to stored or regularly download.

4.1 Simple Knowledge Organization System (SKOS)

In SKOS, concepts can have multiple relationships and this enables the utilisation of the thinking and ideals of electronic thesauri or neural network modelling. Hierarchic classifications which are the primary type used within an SBR nearly always require a parent-child relationship between each level because the structures work in a barrower to broader aggregation approach or vice versa depending on the viewpoint taken. This, like everything else with a classification, only enables a standalone output and introduces issues for coding survey responses, that is, can responses be coded across classification levels?

The use of unique resource indicators (URIs) for concepts within SKOS widens what can be described within a classification than previously. Concepts can be labelled with lexical strings that can utilise multiple languages, notations can be more easily assigned, links to other concepts can be created and all of this can be organised into informal hierarchies and networks using defined concept schemes.

SKOS makes for more granular metadata, easier integration, and a greater ability to share concepts and content across different classifications and/or views. It reduces the overhead of creating mappings between stand-alone classifications and the uris and apis provide greater options for utilising the data and information held within an SBR.

4.2 Resource Description Framework (RDF)

As with SKOS, RDF is a powerful option for statistical classifications by providing a methodology for structuring and storing classification content using unique web identifiers. It allows for the reuse of existing content across multiple classifications whether that be concepts, classification categories or associated indicators.

Currently with industry or product classifications, or large frameworks like the SNA or BPM, and even within SBRs everything is stored in its specific container. This leads to potential differences in the definitions provided for a category as it may not be the same person working on the SNA or the BPM, different editors come into the equation and then when concepts need updating in one, they cannot be flowed through to the other.

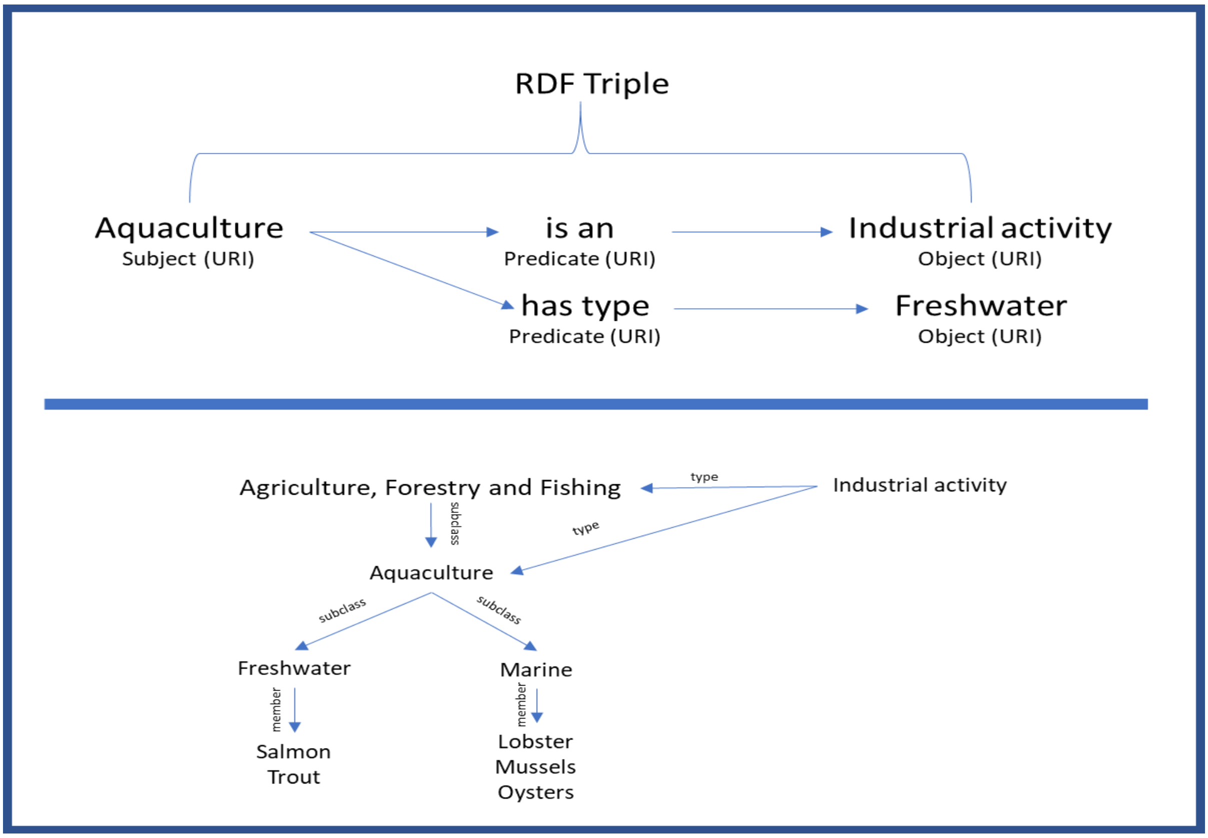
As an example, the concept of aquaculture could be stored once and reused or shared across an industry, product, trade, or sector classification, rather than having single entities in single classifications, each with potentially differing definitions. Would this work within a SBR as well as in the classification space – probably.

What makes RDF so powerful is the RDF triple. The triple is made up of a subject (which describes the web resource for the information), predicate or relationship (which defines a property that the information is sought about) and an object (which contains the value for that predicate) enabling classification content to be disassembled into component parts to enable easier integration and sharing with other systems or frameworks.

The use of the uris allows multidirectional consumption of content between systems and classifications that was never possible before. There is no longer the need to store multiple classifications within an SBR, or the continued employment of download processes of classification content from other systems into the SBR. It also removes the challenges for an SBR about recognising classification abbreviations or version ids to ensure that the latest content is always available.

A further benefit is that when something is changed in the classification there is no need to create a whole new version, and the resulting identity issues that creates because the entities are time-stamped and the use of the apis, along with the uris provides so much more flexibility.

An illustrative example relating to aquaculture is shown below:



Effectively, this process breaks everything down to enable a reconfiguration or repackaging into traditional classification frameworks or alternatively into user defined views which are linked together by their relationship to an overarching concept. By connecting the triples, a graph network of relationships is defined within a set of controlled vocabulary terms using graph knowledge software. An example of the software is SPARQL which is a query language used to retrieve and manipulate data stored in an RDF format.

5. Concept-Based Classification Model

The move to a concept-based classification model is a natural evolution of the process of developing statistical classifications. The general approach to developing a statistical classification is to identify all the things that need classifying and then determine whether these can comprise a flat list or require a hierarchic approach. With the latter it becomes a top-down, bottom-up process determining what parent-child relationships are needed, how many levels the classification will need, what the top or broad groupings need to be and then putting the whole structural development into a sequential code pattern.

A classification such as the Australian and New Zealand Standard Industrial Classification (ANZSIC) is often referred to as the industry classification even though it is actually a measure of economic activities. The classification has no definitive definition of the concept of industry to support it whilst the scope of the classification covers the productive activities within the production boundary of the System of National Accounts (SNA). SNA describes the production boundary as “the physical process, carried out under the responsibility, control and management of an institutional unit, in which labour and assets are used to transform inputs of goods and services into outputs of other goods and services” (SNA 2008, 1.40).

Further the SNA states that “A necessary condition for an activity to be treated as productive is that it must be carried out under the instigation, control and responsibility of some institutional unit that exercises ownership rights over whatever is produced” (SNA 2008, 1.43). ANZSIC applies these principles by grouping the processes and units into broader divisions which form the top level of the classification, e.g., agriculture, manufacturing, wholesale, or retail trade. Then the relevant categories are identified and included within that structure and their inclusion is either based on economic or statistical significance, user requirement or other factors. The process also requires that everything is sequentially numbered accordingly with residual categories included to try and be exhaustive in coverage.

The numbering pattern over time becomes a constraint as there are only so many categories available at each level, based off the number of levels in play. Eventually a point is reached where either the code pattern is changed, for example, from numeric to alpha-numeric or new levels are introduced. This not only impacts the time-series and the classification content but its usability for an SBR. It also poses issues if the underlying model or specifications for the SBR needs to be revised to accommodate the classification changes. This may lead to delays in implementation of those new classifications and a potential recoding of business with the flow-on effects for business surveys and sample selection.

The traditional approach fails to enable a full analysis of the variable being classified and results in a one size fits all attempt at measuring the national economy in a comparable and consistent way. Users then often think they are talking on the same page under the guise of the industry classification when in fact they may not actually know the concept or agree on the concept they are talking about.

For the new way of doing things the concept is the crucial element or entity around which everything is built. Each concept is given a label, and a user agreed definition which may be contextual, such that the definition forms the scope of what the concept measures.

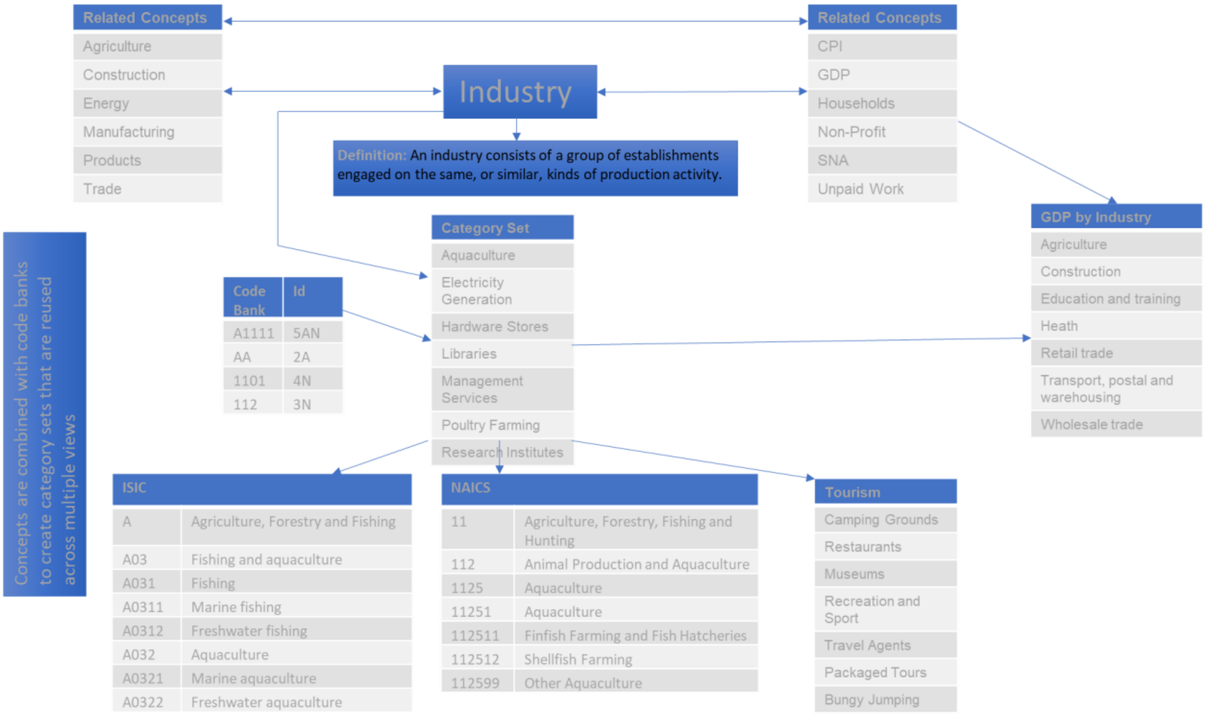
Each concept, as with the use of subjects, objects, and predicates within RDF, will have a relationship to other concepts. This enables a conceptual framework to be created and an easier way for merging and transferring data. This provides strong benefits for data sourced from an SBR. As concepts are related, the categories they contain are linked to other concepts like a neural network or electronic thesaurus.

The scope of the concept is not only framed by the definition but also by the creation of what is called a category set. This set contains all the words that are relevant to the concept and which are in scope of the definition. This is very similar to something like an SDMX codelist (but without the codes). For an SBR, this increases the options around classifying an activity or business allowing more ancillary activities to be identified or made visible and allows greater semantic linkages between classifications such as industry, business type or institutional sector. The data becomes more descriptive.

One of the other great challenges which is somewhat resolved through the new methodology is the management of various classification code patterns and their impact both on classification development and IT system implementation The new approach is through the creation of a code bank to support the concept and category set. The code bank can hold all the approved codes available to be used in building views and classifications off the concept, and these can be in any format, for example alpha, numeric and/or alpha-numeric. Through the use of the universal resource indicator users can determine which codes best suit their data and system needs. This gets rid of the problem of users not actually implementing a statistical classification because it only has one code pattern.

The category set is a dynamic list of words associated with the concept that can be added to at any time. All instances are time-stamped and approved (either manually or automated) and users are then notified of changes to the master list. Users can choose to adopt change immediately or business rules can be applied to the concept to release updated content at regular points of time, for example quarterly, six-monthly, or annually. The category set approach allows for user-driven or standardised views to be created by cutting and dicing the master set of categories, thus providing more flexibility to describe a business or enterprise.

An example can be shown using the concept of industry below. The primary concept of industry has a relationship to many other concepts which allows for the creation of related views. The concept has a definition, a category set, and a code bank attached to it. The category set can be used to produce standardised views to represent ISIC, or the North American Industrial Classification (NAICS), or an output view for the Tourism sector for example. Additionally, aggregated views such as GDP by industry can be created because of the relationships between the concepts which allows the category set to have wider application or linkages.



6. Conclusion

To move the development and maintenance of economic statistical classifications into the ideal of metadata modelling and conceptual classification management will take some time. But the benefits for national statistical offices in terms of cost-reduction, better resource utilisation and greater responsiveness to user demand, outweigh the continuation of the traditional time-consuming process of developing and maintaining statistical classifications that are out of date upon publication. The approach will also enable a greater usage of SBR data and content than previously available, and also reduce overheads in SBR system maintenance.

Users demand real-world reflection of the data and the current process does not enable that. Applying the thinking of metadata modelling and the greater use of conceptual relationships that are fully described, and which utilise the best features of the semantic web is the most practical way forward. Such an approach contains a wealth of information about the concept used in classification and provides rich and flexible information about the relationships and properties within economic statistical classifications . However, adaptability to change is the real key to future success and staying with the status quo approaches does not reflect contemporary reality.