Australian Bureau of Statistics

Implementation of Semantic Web Technology

Presenter: Michael Mecham,
Australian Bureau of Statistics (ABS)
Concepts, Sources & Methods (CSM) Ontology Project

• An ABS R&D project conducted in 2010/11
• Collaboration between
  – Macroeconomics and Integration Group (MIG)
    • statisticians / business customers
  – Technology Services Division
• Developers
  – Michael Mecham
  – Arupa Sarkar
• Business Sponsor
  – Peter Meadows
Overview

• Business need
• Our solution
• Proposed future directions
Why document Concepts, Sources & Methods (CSM)?

- **Principle 2.** To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the methods and procedures for the collection, processing, storage and presentation of statistical data.

- **Principle 3.** To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.

What we were modelling

Very high level definitions:

• Concepts underpinning statistics

• Sources of data for surveys

• Methods = how to
"Concepts" and CSM

A statistical framework can be considered a **conceptual** model of a **domain** within the real world (e.g., "the economy") for which statistical data can be collected and compiled on a well-defined and consistent basis.

The System of National Accounts (SNA) is an internationally agreed statistical framework.

Not all concepts associated with Australian National Accounts data compiled by the ABS will align precisely with the theoretical ideal set out by SNA. The description of Concepts within CSM allows sets of concepts to be understood in relation to each other.
Business needs

- Key needs:
  - Accessibility
  - Usability
  - Maintainability
  - Extendibility
What we did

• Creation of prototype to model Concepts, Sources and Methods of macroeconomic statistics:
  
  • First candidate survey: Australian System of National Accounts
  
  • Made available using a semantic file format (RDF)
  
  • Use of a familiar open source wiki-style web interface
  
  • We created separate instances of the CSM ontology:
    » Pragmatic reasons;
      • Environment 1 - Authoring (dynamic)
      • Environment 2 - Viewing (static)
High level implementation design

Overview of MIG Concepts, Sources and Methods
Ontology Tool Implementation
High level Macro Economic ontology

Example of Ontology-driven Network of MIG Concepts, Sources and Methods

- Production Account
- ASNA
- Prices
- Government Finance Statistics
- Balance of Payments and International Investment Position
- Secondary Income Account
- Macroeconomics
Our Semantic Solution

• Needed a structure to network the perceivable ideals underpinning a survey:
  • It is conceptual - “Concepts”
  • It’s inputs may be varied - “Sources”
  • It has a methodology – “Methods”
Our Semantic Solution

- **Perceivable** was key, therefore CSM could be modelled:
  - Developed model using the open source Ontology editor, Protégé → OWL/RDF
  - Our CSM Ontology modelling akin to Object Oriented modelling
  - Our goal was to produce a fully traversable network of CSM nodes for a web-based Wiki interface
Our Semantic Solution

• Brief technical view:
  • High level design showing CSM linkages to support the goal of a traversable network of nodes
Our Semantic Solution

• Brief technical view:
  • Source hierarchy
Our Semantic Solution

• Wiki interface:

  • Semantic MediaWiki plugin consuming RDF
  
  • Benefit: high degree of end-user familiarity

• Deployment issues
  – ABS intranet
  – Subset available on ABS website
Our Semantic Solution

• Wiki interface example views:
  • Let’s traverse some of the ASNA structure
Our Semantic Solution

- Wiki interface example views:
  - Looks familiar? Great!
Our Semantic Solution

• Wiki interface example views:
  • Scroll to the bottom, there’s something special...
Our Semantic Solution

• Wiki interface example views:
  • Here is where the real power of a CSM Ontology is realised

Category:Concept

(previous 200) (next 200)

Queries for categories

Ask for all instances of "Concept" and for all instances of its subcategories

Pages in category "Concept"

The following 200 pages are in this category, out of 691 total.

<

  • Construction in the reporting economy

A

  • ASNA Banks
  • ASNA Captive Financial Institutions and
    Money Lenders
  • ASNA Central Borrowing Authority
  • ASNA Computer software
  • ASNA Currency
  • ASNA Currency and Deposits
  • ASNA Depository Corporations
  • ASNA Deposits
  • ASNA Financial Auxiliaries
  • ASNA Financial Corporations Sector
  • ASNA Financial Intermediaries
  • ASNA Financial Investment Funds
  • ASNA General Government Sector

A cont.

  • Assets
  • Audiovisual and related services
  • Australian Bureau of Agriculture and
    Resource Economics
  • Australian Institute of Health and Welfare
  • Australian Prudential Regulation Authority
  • Australian System of National Accounts
  • Australian Taxation Office
  • Auxiliary insurance services
  • Average Retail Prices
  • Average Weekly Earnings

B

  • BOP CSM EXAMPLE 2010
  • Balance Sheet
  • Balance at Payments and International
    Investment

C cont.

  • Cultivated biological resources
  • Currency
  • Currency and Deposits
  • Currency and deposits
  • Current International Cooperation
  • Current International Cooperation model
  • Current Taxes on Income and Wealth
  • Current international cooperation
  • Current price values
  • Current tax
  • Current taxes on income, wealth etc.
  • Current taxes on income, wealth etc. Model
  • Current taxes on income, wealth etc. NA
  • Current transfers from the
    Commonwealth government to State and
    local government
Our Semantic Solution

- Wiki interface example views:
  - Aggregation and linking of ASNA CSM, exciting!
Ideas for future direction

• Integration of functionality to highlight contribution of individual components of higher level Major Economic Indicators
  • Hierarchical graph (e.g. tree structure)

• Integration of functionality enabling contribution to ABS CSM from the public

• Supporting ABS research: automation of updating of relevant CSM content (e.g. newsfeeds, external agency datasets....)
Proposal for future capability

Authoring space for external website. Should be seamless environment for ABS staff.

Concepts, Sources & Methods
ABS Website

ABS Intranet

Concepts, Sources & Methods
Document
Economic Intelligence

Secure Online Discussion Communities

Editing, Provider Contact
Research
New Feeds
Fostering International Comparability

• Statistical Frameworks such as SNA are designed to be consistent and shared and applied internationally

• National Accounts data is intended to be comparable around the world

• Shouldn’t CSMs be expressed in a readily comparable form?

• International collaboration, harnessing ontologies and other semantic web technologies, could help make this a reality?
Conclusion

• Significant organisational lessons in terms of potentially improved transparency of survey methodologies

• Shared lessons from this foray represent significant potential future value to statistical agencies

• As one of ABS’ first attempts in the ontology realm, a positive and successful outcome