At your service: an interactive, flexible web-service for translating classification systems and taxonomies

Brandon Whitehead[1], Tawan Banchuen[2], Prof. Mark Gahegan[3], Will Smart[4], Sina Masoud-Ansari[5]
[1]Centre for eResearch, The University of Auckland, Auckland, New Zealand, b.whitehead@auckland.ac.nz
[2]School of Environment, The University of Auckland, Auckland, New Zealand, t.banchuen@auckland.ac.nz
[3]Centre for eResearch, The University of Auckland, Auckland, New Zealand, m.gahegan@auckland.ac.nz
[4]Centre for eResearch, The University of Auckland, Auckland, New Zealand, will.r.smart@gmail.com
[5]Centre for eResearch, The University of Auckland, Auckland, New Zealand, sina.masoud.ansari@gmail.com

INTRODUCTION

Geoscientists have developed and applied classification systems and taxonomies in an effort to enhance and share their understanding of complex geoscience information. However, the classification systems and taxonomies used throughout the geosciences for land cover and land use, soils, geology are neither static nor universal; the classes that we use to represent the Earth vary considerably with time and from place to place. As a general example, think of a geologic map. There are instances when mapped data contains desirable categories, but the granularity of those categories is not suitable. There are also instances where two or more maps (or data sets) lack the desired categories, or the categories across maps are semantically incompatible. This is to be expected, given that: (i) new science, social and economic agendas change what we may wish to differentiate when we look at a map and (ii) new technologies make differentiation of previously inseperable classes more reliable, thus viable.

As another example, consider the nuances of classifying geologic time. Numerous databases of the Earth's geologic history have been constructed based on differing representations of time; geoscientists need to be able to translate one timescale to another to construct comprehensive and meaningful understanding. Chronos.org [1] provides schemas and tools for geoscientists to align common timescales as well as tools to create unique timescales and mappings for specialised data.

This paper describes ongoing work to create a web-based semantic translation service that allows users to: (i) experiment with mappings between classification systems and taxonomies; (ii) visualize translation maps using a given mapping; and (iii) persist their translation maps, and share them with others. Semantic equivalence and similarity are supported via underlying ontologies, which also facilitate the merging and re-grouping of classes.

CLASSIFICATION SYSTEMS AND SCHEMAS

Our translation service is illustrated herein by experimenting with and interoperating between some of the various standard land cover and land use classifications in New Zealand, namely the Land Cover Database version 1.1 (LCDB1), Land Cover Database version 2 (LCDB2) [2], the EcoSat dataset [3] and the Land Use and Carbon Analysis System (LUCAS) [4]. LCDB1 and LCDB2 are nation wide data sets derived from satellite imagery and field reconnaissance with the primary function of tracking change in land cover and land use from 1996 to 2002. EcoSat is a family of projects focused on mapping ecosystem attributes from satellite imagery on a regional scale. The Land Use and Carbon Analysis System has been developed to help New Zealand meet the reporting requirements specified in the Kyoto Protocol [5]. The classification system for each respective data store has been created without considering interoperability a goal.

Table 1: Example schema mappings. Columns 1 and 2 show the equivalent EcoSat class for an LCDB2 category. Columns 3 and 4 show the equivalent LCDB2 classes for an EcoSat category.

LCDB2 class	corresponding EcoSat class	EcoSat class	corresponding LCDB2 class
Indigenous Forest	1. Indigenous Forest (1 to 1 mapping)	Indigenous Forest	1. Major Shelterbelts 2. Afforestation (not imaged) 3. Afforestation (imaged, post LCDB 1) 4. Pine Forest-Open Canopy 5. Pine Forest-Closed Canopy 6. Other Exotic Forests (1 to N mapping)

A semantic mapping was generated between all categories in each respective data store. Table 1 shows an example of a mapping between two data stores. The first and second column in table 1 show which class in the EcoSat data set corresponds to the Indigenous Forest class in the LCDB2 data set; while the third and fourth column show which classes in the LCDB2 data set have a semantic connection to the Indigenous Forest class in the EcoSat dataset. As the table suggests, the mappings are not be the same in each direction. This encoding provides a data user with the potential to compare and translate maps of land cover and land use, at the schema level. For example, a user could view the New Zealand Land Cover Database with the categories developed in the EcoSat product and visa versa.

TRANSLATION SERVICE

Semantic translation services are a relatively new technology. Many previously built systems were quite limited in terms of flexibility and extensibility, often having hard-coded computer scripts defining which translations are supported by the system. These limited systems provide little support for users to experiment with new mapping schemes. The technologies we use are fully open and standards compliant: Sesame is used as an RDF-store and reasoner, RDF/OWL are used for the ontology encoding, SPARQL is used to query the ontology for the mapping between classes. OpenLayers is used for interactive mapping, in turn sourcing translated map data from WMS and WFS services. Maps are styled using SLD, and other OGC standards.

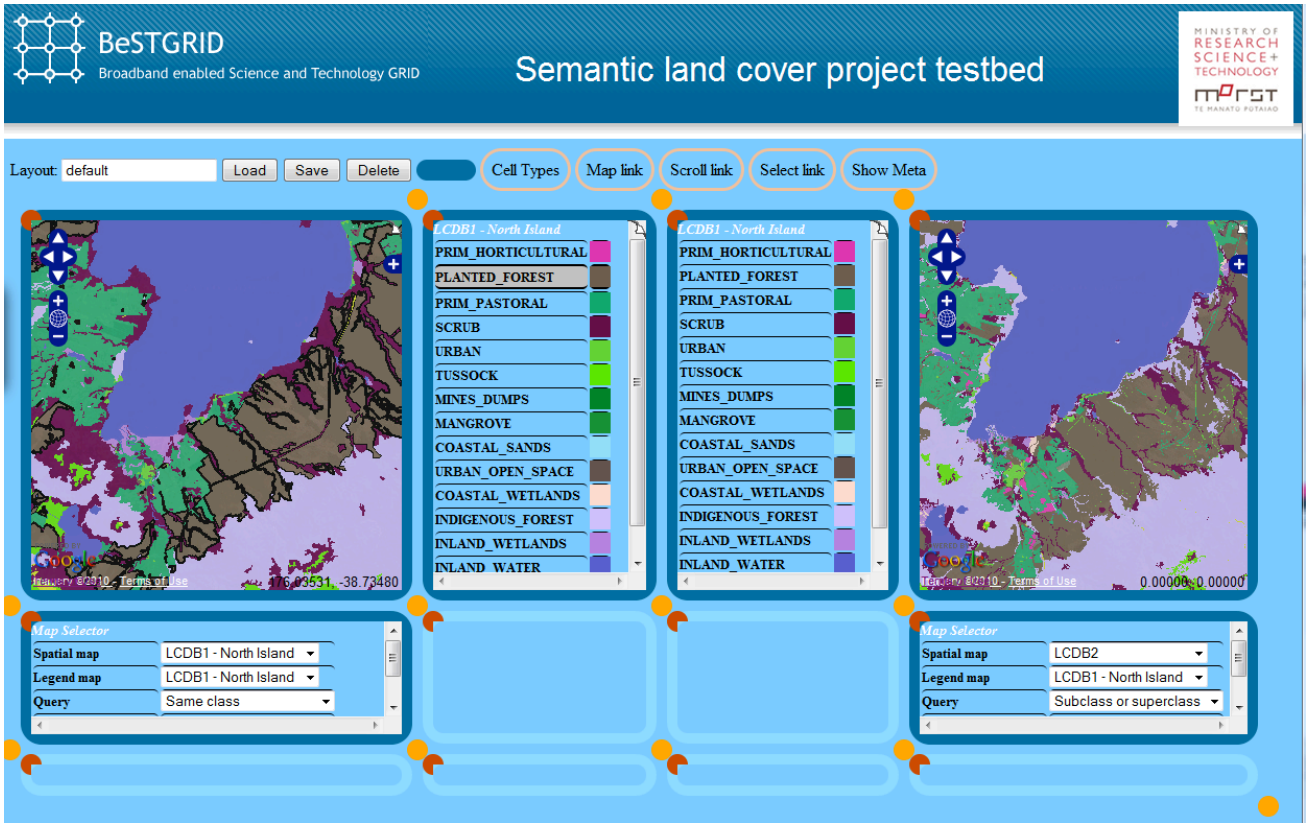

Figure 1: Coordinated view of the land cover change in LCDB1 and LCDB2 datasets.

On balance, our work makes two significant contributions: (i) our service has a highly interactive graphical interface allowing users to compare two classification systems or taxonomies (see figure 1), and to plan, test and refine new mapping schemes; and (ii) mapping schemes, once created, can be serialised into the repository, browsed through and applied in new situations by the same or different users. While the service currently focuses on land cover and land use in New Zealand, we expect the service to be fully extensible to cover other kinds of classified geospatial data sets, including biology, geology, soil, forestry and agricultural data.

REFERENCES

1. *Chronos.org*. Available from: http://www.chronos.org, accessed 20 Sep 2010.
2. *New Zealand Land Cover Database*. Available from: http://www.mfe.govt.nz/issues/land/land-cover-dbase/, accessed 20 Sep 2010.
3. *EcoSat*. Available from: http://www.landcareresearch.co.nz/services/informatics/ecosat/, accessed 20 Sep 2010.
4. *Land Use and Carbon Analysis System*. Available from: http://www.mfe.govt.nz/issues/climate/lucas/, accessed 20 Sep 2010.
5. *Kyoto Protocol*. Available from: http://unfccc.int/kyoto_protocol/items/2830.php, accessed 20 Sep 2010.