



# **Ontology Management and Selection In Re-Use Scenarios**

**Kim Finney (B.Sc, M.Sc)**

**Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy**

**University of Tasmania**

**School of Computing and Information Systems**

**October 2012**

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

The research associated with this thesis abides by the international and Australian codes on human and animal experimentation, the guidelines by the Australian Government's Office of the Gene Technology Regulator and the rulings of the Safety, Ethics and Institutional Biosafety Committees of the University.

This thesis may be made available for loan and limited copying and communication in accordance with the Copyright Act 1968.

## **Acknowledgements**

Several people require special mention who have supported me to deliver this thesis.

First, is my partner, Leanne Wilkes who has lived through this episode vicariously and who probably deserves the award more than I do. Without her critical eye, robust comments and stamina beyond belief in reading and re-reading this dissertation I would never have finished.

Second are my parents. My father, Bob Finney a man of prodigious talent whose enquiring mind inspired me to undertake this research in the first place, offered much useful criticism and advice. My mother, Brenda Finney as always practical and supportive, helped me push on when the end wasn't clearly in sight.

Last, but not least are my supervisors, Simon Milton, Chris Keen and Peter Marshall whose reviews, useful insights, experience and comments helped me to make a thesis.

I would also like to dedicate this dissertation to Elizabeth Bell (1916 – 2009), who didn't understand the Web, had never heard of an ontology and who disliked computers, but encouraged me none-the-less, just as she had done for everything else I've ever attempted in my life.

## Abstract

One of the main impediments to realising the Semantic Web vision is that most scientific data, even those data deployed on the web, are not generally expressed or encoded in an unambiguously defined, machine-interpretable manner. This is particularly the case for Antarctic-themed data. Ontologies that are linked to datasets via semantic annotation are required to achieve semantic-enablement of scientific data infrastructure. In scientific communities that adhere to the Open Geospatial Consortium Service-Oriented-Architecture (Web services) paradigm, Feature Catalogues are the repositories intended to manage and publish descriptions of dataset concepts. This thesis explores how Feature Catalogues can be ontologically-grounded to facilitate semantic annotation and in doing so addresses the lack of guidance in current standards about how to configure an ontologically grounded Feature Catalogue and how best to access the resources it contains for the semantic annotation of Web services. Also investigated is how ontology selection and evaluation is currently taking place in practise because ontology evaluation methodologies mentioned in the literature are resource intensive to apply, often requiring a high level of ontological expertise. Both contributions seek to lower barriers for ontology uptake and reuse within scientific communities.

To address these issues, two scientific communities of practise (i.e., AODN and SCAR) were used as case studies within a Design Science research method to ground-truth the design and to prototype an ontologically grounded, service-enabled Feature Catalogue. To address research questions pertaining to ontology selection and evaluation practise, fourteen experts (from outside of the AODN and SCAR communities) with experience in building semantically-enabled scientific infrastructure, were surveyed and interviewed to ascertain what ontology evaluation methods and criteria are being used in practise. A hierarchical evaluation model was established from analysed expert data using Template Analysis (Crabtree and Miller, 1992; King, 2004). The Analytical Hierarchical Processing (AHP) technique (Saaty, 1980), was then harnessed to establish the relative importance given by experts to each of the model elements.

The contributions arising consisted of an enhanced ISO 19110 Feature Catalogue model which accommodated additional concepts necessary to describe the observation-centric dataset paradigms of the two case study communities. The extended conceptual model was semantically grounded using the DOLCE (upper ontology) and expressed in both OWL and SKOS. Demonstration REST-based service interfaces (and REST query patterns) were created for serving Catalogue content to requesting Web clients. To the author's knowledge, no other Feature Catalogue implementation, founded on the ISO 19110 conceptual model, has attempted to model the Catalogue as an ontology, or permits access to Catalogue content via REST-based service interfaces. This thesis also delivers a

“practical” framework for evaluating and then selecting reusable ontological content which encompasses weighted model elements (indicating relative levels of importance), coupled with expert-derived evaluation metrics. Although the evaluation criteria listed in the framework are not novel in themselves, identifying which criteria are of most utility to experts who are operating in real-world scenarios, is an important contribution to practise.

# Table of Contents

	Page #
<b>Abstract</b>	<b>IV</b>
<b>List of Figures</b>	<b>XI</b>
<b>List of Tables</b>	<b>XV</b>
<b>Chapter 1. Introduction</b>	<b>1</b>
1.1 Research Motivation and Questions .....	2
1.1.1 Challenges in Creating Semantic Repositories For Feature-Centric Services .....	3
1.1.2 Challenges in Selecting and Evaluating Ontologies .....	9
1.2 Methods and Contributions Overview .....	12
1.2.1 Design Science Research .....	12
1.2.2 Qualitative and Quantitative Research Methods .....	16
1.3 Reader's Guide .....	18
<b>Chapter 2 Context Setting &amp; Related Work .....</b>	<b>23</b>
2.1 OGC/ISO and IT Standards (Services) Stack .....	26
2.1.1 Feature-Centricity Of OGC Services .....	28
2.1.2 SOA-Related Metadata .....	29
2.1.3 OGC Service Type Standards .....	32
2.1.4 Semantic Service Description and Orchestration .....	34
2.2 Ontologies .....	37
2.2.1 Characteristics Of An Ontology .....	37
2.2.2 OWL Ontologies .....	42
2.2.3 Ontology Types .....	45
2.3 Feature Catalogues .....	49
2.3.1 OGC-CSW (ebRIM) Embedded Feature Catalogue .....	51
2.3.2 Web Ontology Services .....	52
2.3.3 General Ontology Repository Characteristics .....	54
2.4 Ontology Selection and Evaluation .....	57
2.4.1 Evaluation Criteria .....	58

2.4.2 Evaluation Metrics and Methods .....	63
2.5 Summary .....	69
<b>Chapter 3. General Methods</b>	<b>71</b>
3.1 Research Philosophy and Validation .....	72
3.2 Feature Catalogue Design & Development Methods .....	87
3.2.1 Feature Catalogue Design .....	92
3.2.2 The Feature Catalogue Build Process .....	97
3.2.3 Feature Catalogue Evaluation .....	98
3.3 Practical Ontology Selection and Evaluation Methods .....	99
3.3.1 Expert Screening Survey .....	100
3.3.2 In-depth Interviews With Community Ontology Developers .....	103
3.3.3 Quantitative Survey – Relative Importance of Model Evaluation Criteria .....	105
3.3.4 Ontology Selection and Evaluation Framework .....	116
3.4 Summary .....	116
<b>Chapter 4. Feature Catalogue Design and Development Results</b>	<b>117</b>
4.1 ‘Datasets’, ‘Features’, ‘Observations’ and ‘Features Of Interest’ .....	119
4.1.1 Datasets .....	120
4.1.2 Features .....	121
4.1.3 Observations and Features-Of-Interest .....	124
4.2 Data, Data Models and Requirements Solicitation Via Use Cases .....	126
4.2.1 Review Of The Observation and Measurement Model .....	126
4.2.2 Evaluation Of CSML Feature Types For Modelling Biological Data .....	132
4.2.3 A Broader Review Of Biological Dataset Characteristics .....	136
4.2.4 Possible Biological Feature Type (Observation) Patterns .....	138
4.2.5 Defining The Feature Catalogue Use Cases .....	145
4.3 Review of Related ISO Standards For Defining Feature Types .....	150
4.3.1 Apparent Overlaps and Contradictions in ISO Standards 19110, 19109 and 19126	152
4.3.2 Referencing Between ISO 19110 and 19109 .....	157
4.3.3 ISO 19110 Implemented As A Register .....	157
4.3.4 ISO 19110-Omits ‘Collection’ Criteria .....	157

4.3.5	ISO 19110-Foreseeable Problems With Feature Type ‘Operations’	159
4.3.6	ISO 19110-Placeholder For UoM But Not For Datum	160
4.3.7	ISO 19110-Limitations On Temporal Referencing	160
4.3.8	ISO 19110-Limitations On Referencing Feature Type Symbology	160
4.4	Summary Of Requirements For An Enhanced Feature Catalogue Conceptual Model	161
4.4.1	Enhanced Feature Catalogue Model	161
4.4.2	Encapsulating Observation Features Using The Enhanced Model	166
4.5	Casting The Feature Catalogue Model As An Ontology	168
4.5.1	An Overview Of The Top-level Ontology – DOLCE	171
4.5.2	Useful Ontological Design Patterns	173
4.5.3	Initial Design For A DOLCE-based Feature Catalogue Content Model	176
4.5.4	Less Formal Approaches – The Simple Knowledge Organisation System	179
4.6	Feature Catalogue Ontological Repository Prototyping	180
4.6.1	Ontology Repository Store & Data Schema Creation Tools	181
4.6.2	Versions of DOLCE: DOLCE Lite-Plus vs DOLCE UltraLite	182
4.6.3	Problems Using SKOS For Annotation	184
4.6.4	Ontology De-bugging and Tooling	185
4.6.5	FCATOWL (The Ontologically-grounded Enhanced Feature Catalogue Model)	187
4.7	Prototype REST-Based Feature Catalogue Access Methods	210
4.7.1	The Enhanced Feature Catalogue Model In A Relational Database Form	212
4.7.2	Overview of REST-Based Catalogue Services and Service Descriptions	212
4.7.3	REST-Based Feature Catalogue Service Patterns (HTML and XML Output)	218
4.7.4	REST-Based Feature Catalogue Service Patterns (SKOS Output)	227
4.8	Feature Catalogue Design and Development Results Summary	236
<b>Chapter 5. Feature Catalogue Evaluation and Discussion</b>		<b>239</b>
5.1	‘Architectural Fit’ With Existing SCAR and AODN Infrastructure	241
5.1.1	Potential Use Of Feature Catalogues In SCAR and AODN Infrastructure For Semantic Annotation	245
5.1.2	FCATOWL – Alignment With Other Observation-Centric Ontologies	260
5.2	Current FCAT Limitations	275

5.2.1	<i>NonQualityAttribute</i> Typing .....	275
5.2.2	Attribute Enumeration .....	276
5.2.3	Properties and Access Service Descriptions .....	276
5.2.4	Resource Identifiers, Stability of URLs and Resolution Methods.....	277
5.2.5	Summary Of Limitations & Further Research .....	278
5.3	Community Readiness For Semantic Annotation Approaches .....	279
5.3.1	Evaluation Of Implemented Approaches Given Community Capabilities.....	279
5.3.2	Community Reaction and Feedback .....	287
5.4	Summary .....	288
<b>Chapter 6.</b>	<b>Practical Ontology Selection and Evaluation Data Analysis</b>	<b>291</b>
6.1	Expert Screening Survey .....	296
6.1.1	Screening Survey-Broad Characterisation of Experts and Communities .....	299
6.1.2	Expert Screening Survey-Potential Stratification Of Experts .....	304
6.1.3	Re-use, Selection and Evaluation Methods (As Derived From Screening Survey) ..	312
6.2	In-depth Interview Data Analyses .....	315
6.2.1	Codes and Template Development .....	320
6.2.2	Hierarchical Ontology Evaluation Criteria Model .....	321
6.2.3	Additional Information Relevant To Expert Stratification And Subsequent Data Interpretation .....	326
6.2.4	Summary .....	337
6.3	Quantitative Pair-wise Comparison Survey Data Analyses .....	338
6.3.1	Converting Raw Ratings To Weights .....	340
6.3.2	Improving Data Consistency .....	346
6.3.3	Analyses of Group Results .....	351
6.3.4	Weighted Criteria Model Data – Investigating Patterns In Expert Ratings .....	357
6.3.5	A Comparison Of Criteria Importance Between Interviews and Pair-wise Survey ..	367
6.4	Hierarchical Model Revision and Refinement .....	370
6.4.1	Expert Feedback & Hierarchical Model Revisions .....	370
6.4.2	Selection and Evaluation Framework .....	378
6.5	Practical Ontology Selection and Evaluation Summary .....	387

<b>Chapter 7.    Ontology Selection and Evaluation Discussion</b>	<b>391</b>
7.1 Inconsistencies in Expert Responses .....	392
7.1.1 Methods and Expert Judgement .....	392
7.1.2 Methods Trialled For Identifying and Improving Inconsistent Data .....	398
7.1.3 Obtaining A Group Result .....	399
7.2 Selection and Evaluation Methodologies .....	405
7.2.1 Precs of Expert Data .....	406
7.2.2 Reported Methods With Academic (Or External Community) Origins .....	408
7.2.3 Descriptions Of Community-based Methods Used .....	410
7.3 Governance Issues .....	421
7.3.1 Heavy vs Light Governance Perspectives .....	422
7.3.2 Institutional Backing and Community Mandate .....	425
7.3.3 Governance Framework .....	428
7.4 Summary .....	430
<b>Chapter 8    Conclusions</b>	<b>433</b>
8.1 Feature Type Catalogue Related Contributions .....	437
8.2 Ontology Selection and Evaluation Contributions .....	441
8.3 Methodological Limitations .....	444
8.4 Further Research .....	446
8.5 Summary .....	447
<b>Glossary</b>	<b>449</b>
<b>Bibliography</b>	<b>453</b>
<b>List of Appendices</b>	<b>491</b>
<b>Appendices</b>	<b>493</b>