

A Metadata Management System for Web Based SDIs

by

Andrew Heath Phillips

A Thesis

Submitted in Fulfillment of the Requirements for the Degree

of

Master of Geomatic Science

Department of Geomatics

The University of Melbourne

Parkville, Victoria 3052

December 1998

Declarations

This is to certify that the thesis comprises only my original work except where due acknowledgement is made in the text to all other material.

This thesis is approximately 30,000 words in length.

Andrew H. Phillips

December 1998

Abstract

The process of decision making is best undertaken with the consideration of as much information as possible. One way to maximise the amount of information that is being used in the process is to use metadata engines. Metadata engines can be used to create virtual databases which are a collection of individual datasets located over a network. Virtual databases allow decisions to be made using data from many different databases at many different locations on a network. They shield the user from this fact. From the users point of view they are only using data from the one location.

This thesis investigates some of the concepts behind metadata engines for Internet based Spatial Data Infrastructures. The thesis has a particular emphasis on how metadata engines can be used to create virtual databases that could be of use in the planning and decision making processes. The thesis also investigates some current spatial data technologies such as SDIs, data warehouses, data marts and clearinghouses, their interoperability and their relationship to metadata engines. It also explores some of the more recent spatial data applications that have been developed in the context of metadata engines and Spatial Data Infrastructures.

A major aspect of the thesis was the development of a prototype metadata engine to test the concepts investigated in this thesis. The prototype was developed by modifying the public domain software "Isite" to act as a metadata engine rather than just a metadata search system which is its intended function. The many different approaches that could be used to develop a metadata engine are discussed along with the actual approach that was used in this thesis. The thesis finishes by discussing the many limitations that the prototype has along with some of the improvements that could be made to the prototype to eliminate these limitations.

Acknowledgements

The Department of Infrastructure (Victoria) for their generous sponsorship of this thesis.

Greg Wapling, Jo Moylan, John Hanna, and Bruce Thompson for their insight and help at the Department of Infrastructure.

Ed Young, Ross Honeyman, and Ian Ferguson for their insight on metadata and local government at Land Victoria.

Tony Boston for allowing me to gain access to the Prototype Distributed Data Directory source code from ERIN.

Ben Hatton his assistance in understanding the Isite source code and subsequent help in removing some of the errors that were present in the source code.

The Department of Natural Resources (Queensland) for their demonstration and information on the BLIN system.

Tai On Chan and Wolfgang Effenberg for their insight into the research topic and general guidance on undertaking a research project.

Iestyn Polley for the invaluable assistance he supplied when linking his spatial data viewer application with the metadata engine prototype that was developed in this thesis.

Finally special thanks has to go to my co-supervisors Professor Ian Williamson and Dr. Chukwudozie Ezigbalike. Their assistance in helping me to focus on the task at hand was appreciated greatly.

Table of Contents

<i>Declarations.....</i>	<i>i</i>
<i>Abstract.....</i>	<i>ii</i>
<i>Acknowledgements.....</i>	<i>iii</i>
<i>Table of Contents</i>	<i>iv</i>
<i>Table of Figures</i>	<i>viii</i>
<i>List of Acronyms.....</i>	<i>ix</i>
1 INTRODUCTION	ERROR! BOOKMARK NOT DEFINED.
1.1 Problem Definition.....	Error! Bookmark not defined.
1.2 What is a Metadata Engine?	Error! Bookmark not defined.
1.3 Research Objectives	Error! Bookmark not defined.
1.4 Research Process	Error! Bookmark not defined.
1.5 Justification.....	Error! Bookmark not defined.
1.6 Structure of the Thesis.....	Error! Bookmark not defined.
2 RELEVANT CONCEPTS	ERROR! BOOKMARK NOT DEFINED.
2.1 Introduction.....	Error! Bookmark not defined.
2.2 Spatial Data Infrastructures	Error! Bookmark not defined.
2.2.1 Data	Error! Bookmark not defined.
2.2.2 Communication	Error! Bookmark not defined.
2.2.3 Common Standards and Procedures.....	Error! Bookmark not defined.
2.2.4 Partnerships	Error! Bookmark not defined.
2.2.5 Technology	Error! Bookmark not defined.
2.3 Interoperability	Error! Bookmark not defined.
2.3.1 The Spatial Data Transfer Standard	Error! Bookmark not defined.
2.3.2 The Open Geodata Interoperability Specification	Error! Bookmark not defined.
	defined.

2.3.3 Difference between SDTS and OGIS.....	Error! Bookmark not defined.
2.4 Data Warehousing.....	Error! Bookmark not defined.
2.4.1 The Operational and Data Warehousing Environments	Error! Bookmark not defined.
2.4.2 Properties of a Data Warehouse	Error! Bookmark not defined.
2.5 Data Marts	Error! Bookmark not defined.
2.6 Data Mining	Error! Bookmark not defined.
2.7 Clearinghouses.....	Error! Bookmark not defined.
2.8 Chapter Summary.....	Error! Bookmark not defined.
3 THE CURRENT SITUATION.....	ERROR! BOOKMARK NOT DEFINED.
3.1 Introduction.....	Error! Bookmark not defined.
3.2 Victorian Developments	Error! Bookmark not defined.
3.2.1 Victoria's Geospatial Information Strategy	Error! Bookmark not defined.
3.2.2 DOI GIS Strategy	Error! Bookmark not defined.
3.2.3 Land Channel	Error! Bookmark not defined.
3.2.4 GI Connections.....	Error! Bookmark not defined.
3.3 National Developments	Error! Bookmark not defined.
3.3.1 ANZLIC Metadata Guidelines	Error! Bookmark not defined.
3.3.2 Australian Spatial Data Directory	Error! Bookmark not defined.
3.3.3 The Basic Land Information Network – BLIN	Error! Bookmark not defined.
3.4 International Developments	Error! Bookmark not defined.
3.4.1 New Brunswick Real Property Internet Information Service	Error! Bookmark not defined.
3.5 Chapter Summary.....	Error! Bookmark not defined.
4 THE METADATA ENGINE - CONCEPTS.....	ERROR! BOOKMARK NOT DEFINED.
4.1 Introduction.....	Error! Bookmark not defined.
4.2 The Metadata Engine Concept.....	Error! Bookmark not defined.
4.3 The Five Data Types	Error! Bookmark not defined.
4.4 The Data Models	Error! Bookmark not defined.

4.4.1 Option 1: Metadata and Metadata Engine on local machine	Error!
Bookmark not defined.	
4.4.2 Option 2: Metadata Engine on local machine, Metadata stored on custodians server	Error! Bookmark not defined.
4.4.3 Option 3: Metadata engine and metadata located on one central server	
Error! Bookmark not defined.	
4.4.4 Option 4: Metadata engine located on an independent central server, metadata located on custodians servers.....	Error! Bookmark not defined.
4.4.5 Option 5: Combination of options 3 and 4.	Error! Bookmark not defined.
4.5 Live Connection to the servers	Error! Bookmark not defined.
4.6 Gaining Access to the Remote Databases	Error! Bookmark not defined.
4.6.1 Pages Approach	Error! Bookmark not defined.
4.6.2 Separate File Approach	Error! Bookmark not defined.
4.7 Chapter Summary.....	Error! Bookmark not defined.
5 THE METADATA ENGINE - DEVELOPED	Error! Bookmark not defined.
5.1 Introduction.....	Error! Bookmark not defined.
5.2 The Adopted Data Model	Error! Bookmark not defined.
5.3 Adopted access metadata data storage method	Error! Bookmark not defined.
5.4 Isite	Error! Bookmark not defined.
5.5 The access metadata file	Error! Bookmark not defined.
5.6 How the Prototype works	Error! Bookmark not defined.
5.7 Modifications made to Isite to develop the Prototype	Error! Bookmark not defined.
defined.	
5.7.1 Adapting Isite to run as originally intended by CNIDR on two different servers.....	Error! Bookmark not defined.
5.7.2 Modifying the Isite source code to return access metadata	Error!
Bookmark not defined.	
5.7.3 Linking the developed metadata engine to a spatial data viewer	Error!
Bookmark not defined.	
5.8 Chapter Summary.....	Error! Bookmark not defined.

6 PROBLEMS, LIMITATIONS AND IMPROVEMENTS. ERROR! BOOKMARK NOT DEFINED.

6.1 Introduction.....	<i>Error! Bookmark not defined.</i>
6.2 Limited Querying	<i>Error! Bookmark not defined.</i>
6.3 Limited Access.....	<i>Error! Bookmark not defined.</i>
6.4 Cryptic Search Criteria.....	<i>Error! Bookmark not defined.</i>
6.5 Hard Coding of Source File Locations	<i>Error! Bookmark not defined.</i>
6.6 Points and Lines Display Only.....	<i>Error! Bookmark not defined.</i>
6.7 Accuracy and Integration.....	<i>Error! Bookmark not defined.</i>
6.8 Charging	<i>Error! Bookmark not defined.</i>
6.9 Privacy	<i>Error! Bookmark not defined.</i>
6.10 Speed	<i>Error! Bookmark not defined.</i>
6.11 Metadata Divergence	<i>Error! Bookmark not defined.</i>
6.12 Firewalls	<i>Error! Bookmark not defined.</i>
6.13 Data Currency.....	<i>Error! Bookmark not defined.</i>
6.14 Chapter Summary.....	<i>Error! Bookmark not defined.</i>

7 CONCLUSIONS AND RECOMMENDATIONS ERROR! BOOKMARK NOT DEFINED.

7.1 Introduction.....	<i>Error! Bookmark not defined.</i>
7.2 Summary and Conclusions of Thesis.....	<i>Error! Bookmark not defined.</i>
7.3 Recommendations.....	<i>Error! Bookmark not defined.</i>

References.....	4
------------------------	----------

Table of Figures

- Figure 1-1:** Multi-level integrated research framework for the management of geospatial information technology.**Error! Bookmark not defined.**
- Figure 2-1:** Example of a SDTS connection between an Oracle database and an Ingres database.**Error! Bookmark not defined.**
- Figure 2-2:** Example of what a network built using OGIS compliant software will look like.....**Error! Bookmark not defined.**
- Figure 3-1:** The Pages Concepts.....**Error! Bookmark not defined.**
- Figure 3-2:** Structure of the Prototype Distributed Data Directory**Error! Bookmark not defined.**
- Figure 4-1:** Virtual database architecture.**Error! Bookmark not defined.**
- Figure 4-2:** Example of Metadata and Metadata Engine located on the users own machine.**Error! Bookmark not defined.**
- Figure 4-3:** Example of metadata engine on local machine whilst the metadata is stored on custodians server.**Error! Bookmark not defined.**
- Figure 4-4:** Example of metadata and metadata engine located on a single central server.**Error! Bookmark not defined.**
- Figure 4-5:** Example of a central metadata engine, and distributed metadata approach.**Error! Bookmark not defined.**
- Figure 4-6:** Example of a model that combines both options 3 and 4.**Error! Bookmark not defined.**
- Figure 5-1:** Architecture of the metadata engine prototype.**Error! Bookmark not defined.**
- Figure 5-2:** Isite Information System Architecture.....**Error! Bookmark not defined.**
- Figure 5-3:** Http to z39.50 stateful gateway architecture.**Error! Bookmark not defined.**
- Figure 5-4:** Example of planning access metadata that appears in the "directory.txt" file.**Error! Bookmark not defined.**
- Figure 5-5:** Screen dump of the gateway.html web page.**Error! Bookmark not defined.**
-

Figure 5-6: The spatial data viewer.....**Error! Bookmark not defined.**

Figure 5-7: The data manager menu.....**Error! Bookmark not defined.**

List of Acronyms

3-D	Three Dimensional
AAPT	Australian Associated Press Telecommunications
ABS	Australian Bureau of Statistics
AEC	Australian Electoral Commission
ANSI	American National Standards Institute
ANZLIC	Australia New Zealand Land Information Council
API	Application Programming Interface
ASDD	Australian Spatial Data Directory
ASDI	Australian Spatial Data Infrastructure
ASGC	Australian Standard Geographical Classification
ATS	Automated Titles System
AusPost	Australia Post
BLIN	Basic Land Information Network
CAD	Computer Aided Design
CD	Compact Disc
CGI	Common Gateway Interface
CISP	Computer Inventory of Survey Plans
CNIDR	Center for Networked Information Discovery Retrieval
CRCS	Content, Reform and Compliance Services
DBMS	Database Management System
DCDB	Digital Cadastral Database
DNR	Department of Natural Resources (Queensland)
DNRE	Department of Natural Resources and Environment (Victoria)
DOI	Department of Infrastructure (Victoria)
DSS	Decision Support System
DXF	Digital Exchange Format

ERIN	Environmental Resources Information Network
ESD	Electronic Service Delivery
FGDC	Federal Geographic Data Committee
FIPS	Federal Information Processing Standard
GI	Geospatial Information
GIS	Geographic Information System
GPS	Global Positioning System
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
ID	Identification
IT	Information Technology
IVAS	Integrated Valuations and Sales database
IVR	Interactive Voice Response
KDD	Knowledge Discovery in Databases
LGA	Local Government Authority
LGPMIS	Local Government, Planning, Marketing Information Services
NBGIC	New Brunswick Geographic Information Corporation
NISO	National Information Standards Organization
ODCS	Organisational Development and Corporate Finance
OGC	OpenGIS Consortium
OGIS	Open Geodata Interoperability Specification
PATS	Property Assessment and Taxation System
PDDD	Prototype Distributed Data Directory
PTC	Public Transport Corporation
SAPI	Search Application Programming Interface
SCDB	Survey Control Database
SDI	Spatial Data Infrastructure
SDMB	State Digital Map Base
SDTP	Spatial Data Transfer Processor
SDTS	Spatial Data Transfer Standard
SPES	Strategic Planning and Economic Services

SQL	Structured Query Language
TCP	Transfer Control Protocol
TQM	Total Quality Management
TVP	Topological Vector Profile
USA	United States of America
USGS	United States Geological Survey
USL	Universal Systems Limited
VSDD	Victorian Spatial Data Directory
WA	Western Australia
WALIS	Western Australian Land Information System
WWW	World Wide Web

<C:\WINNT\Profiles\sam\Desktop\ian's thesis\andrew\Chapter 1.doc>

<C:\\WINNT\\Profiles\\sam\\Desktop\\ian's thesis\\andrew\\Chapter 2.doc>

A Clearinghouse is an advanced form of search engine designed for the dissemination of spatial data. They are part of the overall Spatial Data Infrastructure and allow a person with access to a network to search that network to find what spatial data exist, what the characteristics of the dataset are, and how they can get hold of the data if a direct link to the data does not exist. The research that is being undertaken in this thesis has many similarities to the concepts that are involved in clearinghouses. The thesis essentially extends the concepts of the clearinghouse. It allows a clearinghouse to not just to allow the users to find what data sets exist and are available, but also allow the user to query and display several of them at once whilst they remain in the custodian's care.

<C:\\WINNT\\Profiles\\sam\\Desktop\\ian's thesis\\andrew\\Chapter 3.doc>

<C:\\WINNT\\Profiles\\sam\\Desktop\\ian's thesis\\andrew\\Chapter 4.doc>

<C:\\WINNT\\Profiles\\sam\\Desktop\\ian's thesis\\andrew\\Chapter 5.doc>

<C:\\WINNT\\Profiles\\sam\\Desktop\\ian's thesis\\andrew\\Chapter 6.doc>

<C:\\WINNT\\Profiles\\sam\\Desktop\\ian's thesis\\andrew\\Chapter 7.doc>

References

- AAPT, 1998, *VicOne: A network for the 21st century*. AAPT. 4/12/98, <<http://www.net.aapt.com.au/vicone/>>.
- Althiede, P., 1992a, Design of a Spatial Data Transfer Processor. USGS. 1/12/98, <<ftp://sdts.er.usgs.gov/pub/sdts/articles/ascii/design.doc>>, (Reproduced from *Cartographic and Geographic Information Systems*, **19**(5)).
- Althiede, P., 1992b, An Implementation Strategy for SDTS Endcoding. USGS. 1/12/98, <<ftp://sdts.er.usgs.gov/pub/sdts/articles/ascii/encode.doc>>, (Reproduced from *Cartographic and Geographic Information Systems*, **19**(5)).
- ANSI, 1992, *Application Service Definition and Protocol Specification for Open Systems Interconnection*. American National Standards Institute. 20/7/1998, <<http://www.cni.org/pub/NISO/docs/Z39.50-1992/>>.
- ANZLIC, 1996, *ANZLIC Guidelines: Core Metadata Elements Version 1*, Report: ANZLIC Working Group on Metadata.
- ANZLIC, 1997, *Core Metadata Elements for Land and Geographic Directories in Australia and New Zealand*. Australia New Zealand Land Information Council. 5/6/1997, <<http://www.auslig.gov.au/pipc/anzlic/metaelem.html>>.
- Arseneau, B., Coleman, D., and MacLeod, G., 1997a, The Development of the Land Gazette. In *Proceedings of URISA*, (Toronto, Canada: URISA), 9 pages (CD only).
- Arseneau, B., Coleman, D., and Quek, S. H., 1997b, Internet Access to Real Property Information. In *Proceedings of GIS'97*, (Vancouver. British Columbia, Canada: Geoplace), 13 pages (CD only).
- Brobst, 1996, *Critical Success Factors in Building a Data Warehouse*, Powerpoint Presentation, Melbourne, The 1996 Data Warehousing Solutions Symposium.

- Buehler, K. and McKee, L., 1996, *The OpenGISTM Guide: Introduction to Interoperable Geoprocessing* (Wayland, Massachusetts: OGC Technical Committee, Open GIS Consortium, Inc.).
- Codd, E. F., 1990, *The Relational Model for Database Management* (Massachusetts: Addison - Wesley).
- Dilly, R., 1995, *Data Mining*. The Queen's University Belfast. 17 July 1997, <http://www-pcc.qub.ac.uk/tec/courses/datamining/stu_notes/dm_book_1.html>.
- DNR, 1997, *Basic Land Information Network (BLIN)* .Department of Natural Resources, Queensland. 5/5/98, <http://www.dnr.qld.gov.au/infomercials/blin_fax.html>.
- DNRE, 1997a, *Land Channel Implementation Strategy*, Department of Natural Resources and the Environment, Land Channel Working Group, Implementation Strategy.
- DNRE, 1997b, *Selling Geospatial Information: Best Practice Principles: A Summary*, Land Victoria, Department of Natural Resources and Environment, Summary.
- DNRE, 1997c, *Victoria's Geospatial Information: Building the Foundations 1997-2000*, The Department of Natural Resources and Environment, Land Victoria, Strategic Plan.
- DNRE, 1998, *GI Connections - About This Site*. Department of Natural Resources and Environment, Victoria. 10/6/98, <<http://www.giconnections.vic.gov.au/content/about/about.htm>>.
- Ezigbalike, I. F. C., 1988, *Land Information Systems Development: Software and Management Considerations*. PhD Thesis, The University of New Brunswick.
- FGDC, 1996a, *The National Spatial Data Infrastructure*. Federal Geographic Data Committee (US). 3/3/97, <<http://www.fgdc.gov/NSDI/Nsdi.html>>.

- FGDC, 1996b, *Questions and Answers*. Federal Geographic Data Committee (US). 19/6/97, <<http://www.fgdc.gov/Clearinghouse/background.html>>.
- Gamiel, K. and Warnock, A., 1994, *The Isite Information System, Version 1.05* (North Carolina: The Clearinghouse for Networked Information Discovery and Retrieval).
- Glover, J., 1997, OPEN GIS™ - Removing the barriers to a spatial data infrastructure. In *Proceedings of AGI 96*, (Birmingham, London: Association for Geographic Information), 5 pages (CD only).
- Gore, A., 1998, The Digital Earth. *The Australian Surveyor*, **43**(2), 89-91.
- Hatton, B., 1997a, *Prototype Distributed Data Directory*. ERIN. 4/11/97, <<http://www.erin.gov.au/database/metadata/isite/index.html>>.
- Hatton, B., 1997b, *Systems Architecture*. ERIN. 22/10/97, <<http://www.erin.gov.au/database/metadata/isite/arch.gif>>.
- Holsheimer, M. and Siebes, A., 1994, *DATA MINING: The Search for Knowledge in Databases*, CWI, No. CS-R9406.
- Inmon, W., 1995a, *What is a Data Mart?* D2K: Data to Knowledge. 18 July 1997, <<http://www.d2k.com/d2k/library2.html>>.
- Inmon, W., 1995b, *What is a Data Warehouse?* Prism Solutions. 16 July 1997, <http://www.cait.wustl.edu/cait/papers/prism/vol1_no1/>.
- Kildow, M., 1996, *The value of Metadata (A NSDI report)*. US Fisheries and Wildlife Services. 4/9/97, <<http://www.r1.fws.gov/metadata/meta.html>>.
- Korth, H. F. and Silberschatz, A., 1991, *Database Systems Concepts* (Singapore: McGraw - Hill).

- Lazar, R., 1992, The SDTS Topological Vector Profile. USGS. 1/12/98,
<<ftp://sdts.er.usgs.gov/pub/sdts/articles/ascii/tvppaper.doc>>, (Reproduced from
Cartographic and Geographic Information Systems, **19**(5)).
- Lennon, P. and Berenyi, M., 1997, BLIN: Basic Land Information Network:
Captilising on our Information resource. In *Proceedings of AURISA*,
(Christchurch, New Zealand: AURISA), 5 pages (CD only).
- Masser, I., 1998, *Governments and Geographic Information* (London: Taylor and
Francis).
- McKee, L., 1996, *Building the GSDI*. The Open GIS Consortium. 28/8/97,
<<http://www.opengis.org/articles/gsdi.html>>.
- Mooney, D. J. and Grant, D., 1997, The Australian National Spatial Data
Infrastructure. In *Framework for the World*, edited by D. Rhind. (New York:
John Wiley & Sons Inc), 1st, ed., pp. 187-201.
- Nebert, D., 1996, *Information Architecture of a Clearinghouse*. Federal Geographic
Data Committee (US). 4/8/97,
<<http://www.fgdc.gov/clearinghouse/www5/clearinghouse.html>>.
- Orr, K., 1996, *Data Warehousing Technology*. The Ken Orr Institute. 9/9/98,
<<http://www.kenorrinst.com/DWPaper.htm>>.
- Phillips, A., Ezigbalike, C., and Williamson, I., 1998a, Spatial Data Infrastructure
Concepts. *The Australian Surveyor*, Forthcoming.
- Phillips, A., Williamson, I., and Ezigbalike, C., 1998b, The Importance of Metadata
Engines in Spatial Data Infrastructures. In *Proceedings of AURISA 98*, (Perth:
AURISA), 10 pages (CD only).
- Polley, I., 1998, *Facilitating the use of cadastral data through the World Wide Web*.
Master of Geomatic Science Thesis, The University of Melbourne.

- Rush, G., 1997, Land Information Management in Queensland. In *Proceedings of iKUSASA CONSAS 97*, (Durban, South Africa: South African Council for Professional and Technical Surveyors), 12 pages (CD only).
- Shelley, E. P. and Johnson, B. D., 1995, Metadata: Concepts and Models. In *Proceedings of Third National Conference on Management of Geoscience Information and Data*, (Adelaide, Australia: Australian Mineral Foundation), pp. 4.1-5.
- USGS, 1996, *What is SDTS*. Mid Continent Mapping Centre. 11/3/98, <<http://mcmcweb.er.usgs.gov/sdts/whatsdts.html>>.
- Wilson, E., 1998, Manuals go click. *The Age*, Melbourne (Australia), 30/6/98, 1.