BRITISH COLUMBIA

Corporate Services for the Natural Resource Sector

Information Management Branch

Metadata Publishing Consolidated Reference for the Natural Resource Sector

Last Updated:	May 8, 2012
Version:	1.0.0
Document:	NRS_Metadata_Publishing_1.0.0

Document History

Version	Description of Change, Review or Approval	Author	Date
1.0	A first guide to create a consolidated reference for publishing metadata within the Natural Resource Sector.	David Skea	Dec 2011

Contents

1 Introduction

1.1 NRS Corporate Services Centralization

In the fall of 2010 the British Columbia Government carried out a major restructuring of the natural resource ministries. The resulting set of ministries became:

- Ministry of Aboriginal Relations & Reconciliation (MARR)
- Ministry of Agriculture (AGRI)
- Ministry of Energy and Mines (MEM)
- Ministry of Environment (ENV)
- Ministry of Forests, Lands and Natural Resource Operations (FLNR)

As part of this restructuring a decision was made to centralize all corporate services into a single entity supporting the five ministries. This resulted in the creation of CSNR (Corporate Services for the Natural Resource Sector) which included the Information Management Branch (IMB). The role of the IMB is to support the design, development, deployment and running of hardware and software systems to support the work of the sector.

1.2 Centralization Goals

A goal (one of many) of this centralization is to simplify operations by reducing the number of standards and procedures used to develop new application and possibly reduce the number of applications by identifying common components and eliminating duplicate functionality.

A second goal is to try and achieve a more integrated world view between the five resources ministries.

This document will address the question of publishing metadata or relational database schema and associated data dictionary entries. Metadata associated with geo-spatial information can be found at: 'XML Schema Profile for DataBC Geographic Data Discovery Service Application Database'

2 Metadata Publication Approaches

Within the sector there are two major approaches to publishing metadata.

2.1 Agriculture/Environment/ILMB Approach

The approach taken by the Ministry of Environment and Ministry of Agriculture focused on the use of a tool that interacts directly with the Oracle Designer Repository to publish to the Web. This tool is known as the 'Repository Object Brower' or ROB. Additionally, there is a tool that lists all applications known as the 'Infrastructure Reporting System or IRS. This lists all applications, the servers they run on, contact people and databases they connect to. There is no obvious link between this tool and the ROB.

2.2 Forests Approach

The approach taken by the Ministry of Forests and Range (Now Ministry of Forest, Lands and Natural Resource Operations) is to maintain a central web page listing all applications with links to their data model ERD's, entity and table definitions and BCGW publication. Associated application metadata entries are available through a central data dictionary.

3 Agriculture/Environment/ILMB Metadata Publication

3.1 Repository Object Browser (ROB)

ROB is a tool to that directly accesses an Oracle Designer repository. Making a data model accessible to ROB is simply a matter of setting access permissions within the repository. The ROB does not directly support the display of diagrams (ERD's) but does allow PDF's and other accompanying documentation to be stored in a 'file' folder that can be drilled into. It is up to the model developer to generate the PDF and upload it to this directory. Synchronization issues between the model and the diagram can occur.

3.2 Using ROB to Browse the Ministry CASE Repository[1]

3.2.1 Purpose

This document is a quick launch pad to using the **ROB (Repository Object Browser)** tool to browse application systems in the Ministry production repository rdrprod1. This is not an in depth explanation of the use of ROB, and is not a replacement to the existing ROB on line help.

Using this document as a starting point, the reader should be able to be online browsing the appropriate areas of the repository, including downloading of diagrams.

3.2.2 Audience

This document is intended for Ministry staff or business partners who have access to the CSD intranet and an interest in browsing the Ministry CASE repository. It is assumed that the reader will have some knowledge of CASE (things such as ERD diagrams, entities, etc.), basic intuition of navigation web pages, and some familiarity with the Ministry application(s) of interest.

This document (and the ROB tool) is not intended for updating CASE applications, or development of new applications. Staff requiring this level of access to the repository should contact IMB Data Architecture.

3.2.3 Getting Started

The ROB has a default set up which automatically connects you into the repository without requiring you to have a repository account. The URL is

http://a100.gov.bc.ca/int/robread

---

[1] Taken directly from the old ILMB site http://csdgww.bcgov/imb/da/services/rob.html.

The ROB home page will have the Ministry banner.

To Brows the Repository use the Navigate button or click on Repository Navigator..

This is probably the best way to locate an application of interest and browse the application contents. Upon selecting this option, you are given the options to drill down into workareas, configurations, or all containers.

Workareas are logical groupings of Ministry applications, and are used to classify particular applications or versions of applications according to its current status in the Ministry systems development life cycle. To enter a workarea click on the plus sign on the left side of the name.
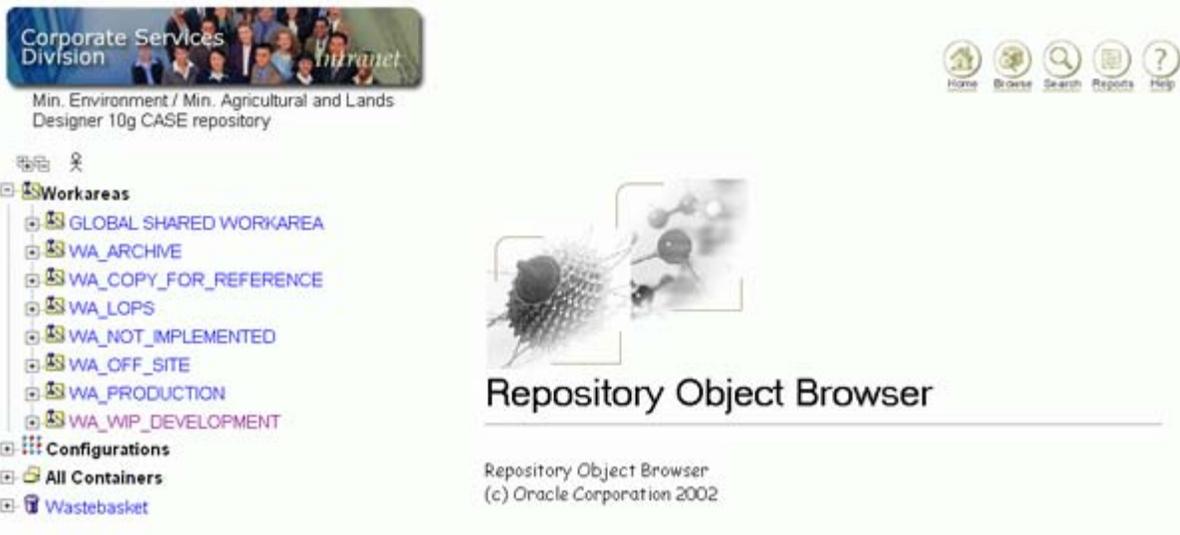
3.2.4 Definitions of these workareas are:

- GLOBAL SHARED WORKAREA – a default workarea not currently being used.
- WA_ARCHIVE – corresponds to applications/versions that have been superseded by a newer version and are in a staging area to be removed from the repository, or older versions of an existing application that are needed online for reference
- WA_OFF_SITE – corresponds to applications/versions that are currently being developed in an external repository. This workarea will reflect the application state at the time the application was handed over to the developer and will not be line with the latest changes.
- WA_COPY_FOR_REFERENCE – corresponds to applications/versions that belong to an external data custodian. An example of this is where Ministry of Forest and Range (MOFR) data modesl are needed by the discovery service for definition.
- WA_NOT_IMPLEMENTED – corresponds to applications/versions in a state of halted development.
- WA_LOPS – corresponds to applications/versions for the Local Publication SDE warehouse data models
- WA_PRODUCTION – applications considered in a full working state.
- WA_WIP_DEVELOPMENT – applications currently undergoing significant development inside the Ministry repository. Note that minor updates to bring CASE apps up to date with changes to production systems may be made directly to WA_PRODUCTION (such as a change to a single column in a database table).

Using workareas is probably the best way to begin browsing the repository as it provides a context to help find the particular application/version of interest. Most Ministry staff will be interested in WA_PRODUCTION and WA_WIP_DEVELOPMENT applications. Note: If a application is in both of these workareas, this will indicate its two states a production release (WA_PRODUCTION) and a development state (WA_WIP_DEVELOPMENT). NOTE: applications

containers are identified by the Application Acronym, this acronym in just about all cases will translate to the table definition prefix, in the database where it resides.

Some applications have configurations, which is a specific collection of application versions through time. Additionally there is an All Containers option within Repository Navigator which may be a convenient way to locate an application and all available versions irrespective of workarea. These may be used to find an alphabetical listing or to review version history of a application. Most browsing should be done in WA_PRODUCTION though.

3.2.5   Tips on Browsing Diagrams

Though it will list all existing diagrams and associated meta-data, the ROB product does not directly display diagrams. As a workaround, staff at Data and Database Services will endeavour to create PDF diagrams for requested production applications, and upload these PDF files directly to the application system. These diagrams can be found, downloaded and viewed by drilling into the 'Files' component of an application system, then clicking on the name to download.
When prompted, click 'Open' to view the diagram in a PDF viewer. Note that some applications may have other files uploaded (such as code files) and are not to be confused with diagram PDF files.

If you require diagrams that are not yet added, contact Data Architecture.

The left pane of the navigator allows drill down by clicking on the plus sign the right pane is for information. Clicking on icons in the information panel will provide further information about objects.

For Searching the Repository use the button or click on Repository SEARCH. Searching accepts % as wild cards and can be used to find, for example, multiple instances of a table definition. So for example a search name of "CORP%CODE" with an Element Type of "Table Definition", leaving Workarea as "Workarea" would list all tables with a prefix of CORP and a suffix of CODE.

3.2.6   Other Tips on ROB

Application containers names are based on the application acronym. This is 3 to 5 letters, where any suffixing is just to further instantiate the meaning of the data model. So if there is an application folder called ACAT and one called ACAT_WHSE, both are data models of the ACAT dataset. ACAT_WHSE is a warehouse data model.

Advanced users of the repository may take advantage of the Reports, Object Browser features. These tools may be useful for identifying potential for reusable objects, or other purposes. Use the On Line Help System built into ROB to find out more about using these features.

If you have any trouble finding the application you need, require diagrams that are not yet available, or have any general problems or comments on ROB, please contact Data Services.

3.3   Class Models

In addition to the entity relational data models developed in Designer, MAL/MoE also has data models developed in UML. This 'Object Oriented' approach tends to deal more with applications than strict data centric developments. Data models developed in UML can include: class models, object models, state diagrams, sequence diagrams, activity diagrams and other UML diagrams. These models are stored in XMI format (http://www.omg.org/spec/XMI/) in a Subversion repository. The repository can be found at:

http://axe.env.gov.bc.ca:8080/

4 Forest, Lands and Natural Resource Operations Metadata Publication

The approach to publishing metadata followed by the Ministry of Forest, Lands and Natural Resource Operations is based on the manual construction of a web page containing the models and the use of two applications for managing the data dictionary.

4.1 Data Models

 After the data model review has been signed off, the vendor or data administration staff generate three artefacts describing the model. These are listed by application on the following Web site:

http://www.for.gov.bc.ca/his/datadmin/models/models.htm#models

The artefacts are:

1. A PDF file, named XXX_ERD.pdf, of the ERD data model. Guidelines for what the ERD should look like are given in a section called "Model Display Standards" in "Guide S7 - Modelling Standards for Relational Applications" which can be found at:

   http://www.for.gov.bc.ca/his/datadmin/s7.pdf

2. A PDF report, named XXX_ENT_ATT.pdf, containing the Entity and Attribute Definitions Report.
3. A PDF report, named XXX_TABLE_COL.pdf, containing the Table and Column Definitions Report.

These artefacts are stored at http://www.for.gov.bc.ca/ftp/HIS/external/!publish/Web/data-models/ and the above Web page points to them. Changes to these artefacts are done manually.

4.2 Integrated Data Dictionary

In addition to the artefacts described above information about all data models is made available through a tool called the "Integrated Data Dictionary". This tool allows users to search for entities or attributes, corporate codes or organizational (org) units. The tool is similar to the ROB but simpler and easier to use. The integrated data dictionary spans the entire Ministry of Forests business area, i.e., an enterprise model but will be in conflict with other Ministries in the Sector. The integrated data dictionary can be found at:

http://www.for.gov.bc.ca/pscripts/isb/idd/iddmain.asp#

4.3   Integrated Spatial Data Dictionary

For mostly historical reasons there is a separate data dictionary for spatial data features (spatial data entities). This tool can be found at:

http://www.for.gov.bc.ca/pScripts/isb/idd/isddmain.asp

5 NRS Metadata Inventory

An initial inventory of metadata and metadata content standards includes:

5.1   NRS Data Models

http://www.for.gov.bc.ca/his/datadmin/models/models.htm#models

https://a100.gov.bc.ca/int/robread/cdwp.startup

5.2   NRS Data Dictionaries

IDD (Integrated Data Dictionary) - http://www.for.gov.bc.ca/pscripts/isb/idd/iddmain.asp

 ISDD (Spatial Component) - http://www.for.gov.bc.ca/pScripts/isb/idd/isddmain.asp

GeoBC Metadata content - http://www.data.gov.bc.ca/dbc/geo/geometadata/profile.page

GeoBC Discovery Service - https://apps.gov.bc.ca/pub/geometadata/home.do

5.3   NRS Applications

https://extranet.for.gov.bc.ca/AppDev/Applications/WebApps/webapps.htm

https://extranet.for.gov.bc.ca/AppDev/Applications/stdapps/stdapps.htm

https://apps.gov.bc.ca/int/irs/init.do

6 Future Metadata Publishing Direction

6.1   Future: An alternate view (Semantic Metadata)

A common limitation of the metadata publishing standards we currently deploy is that they are intended to be consumed only by human operators. Computer based search is limited, if possible at all. The future for metadata publication should be centred on development of the Semantic Web.

Searching metadata is difficult when we are talking about a hundred or so high quality, high effort to produce, spatial data sets. It becomes much worse when you include objects such as spread sheets, code blocks, ERD data models and completely unmanageable when you think about millions of Web pages. A different approach is required.
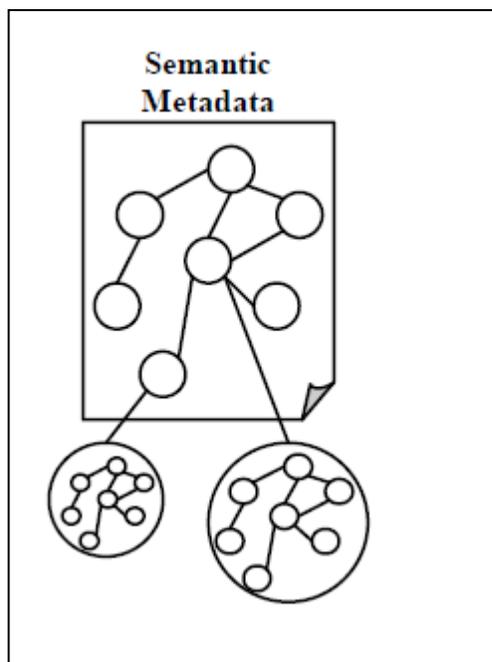
6.2   A New Approach Needed

"Let us assume we have committed to a strategy that relies on some kind of common model. Inevitably, we will have to integrate with another partner who has not committed to the same strategy. How do we align our metadata then? How do we explain what the terms mean? How do we connect our terms and relationships to their terms and relationships"?[2]

"Finally, once we decide to move beyond the realm of metadata how does actual data fit in? Do we have to start from scratch with a new common data model? How do we connect our metadata to our data? By now it should be clear that many of the technology choices of the past twenty years or so have simply entered the fray at the wrong level. We attempt to model domains and processes and people and data separately using insufficient and rigid abstractions. We ignore realities about how information is produced and consumed. Information does not have a format and rarely has boundaries. There is no explicit distinction between data and metadata. Perhaps most fundamentally, we do not and usually will not agree to a consistent world view. Any top down IT initiative that ham-handedly tries to get around this reality is doomed to fail as we have seen time and time again".

"We need another strategy that combines the efficiencies of top down efforts with the reality of organic, bottom up perspectives. We need a data model that frees us from the constraints of

---

[2] From http://www.infoq.com/articles/roa-resource-metadata

a particular schema, language, object model, product or world view. We need to encourage people to agree where they do, but allow them to disagree as well. We need to let them care about different things and be fully supported by any modeling activities. Where we cannot support them centrally, we need to allow their needs to be met on the edge. This is not to say we are heading toward anarchy. Where we need to validate and restrict, we certainly can".[3]

6.3   New Approach Direction

An new approach is emerging that uses a generalized model to allow structured and semi-structured data to be mixed, exposed, and shared across different applications. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data based on the Resource Descriptor Framework (RDF).

RDF has features that facilitate data merging even if the underlying schemas differ, and it specifically supports the evolution of schemas over time without requiring all the data consumers to be changed. It extends the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link (this is usually referred to as a "triple"). Using this simple model, it allows structured and semi-structured data to be mixed, exposed, and shared across different applications.

6.3.1   New Approach Examples

Let's consider a single item of interest, in this case a Web service for computing Slope, Aspect and Elevation for a supplied polygon (sometimes referred to a SEA Service). Here are some facts about this service offering:

- Vivid Solutions created the service http://someserver.com/SEA/slope?
- Vivid Solutions is a consulting company.
- The service http://someserver.com/SEA/slope? was published on 2011-11-05.
- The service http://someserver.com/SEA/slope? is a terrain reporting service.

Notice that although all these facts are different they all have the same pattern: a resource, a property of that resource and a value. Explicitly:

Resource	Property	Value
Vivid Solutions	Created	http://someserver.com/SEA/slope?
Vivid Solutions	Is_a	Consulting company
http://someserver.com/SEA/slope?	Published on	2011-11-05
http://someserver.com/SEA/slope?	Is_a	Terrain reporting service

---

[3] From http://www.infoq.com/articles/roa-resource-metadata

It is important to note that the value, as in the case of the first fact, is itself a resource. Hence the collection defines a graph.
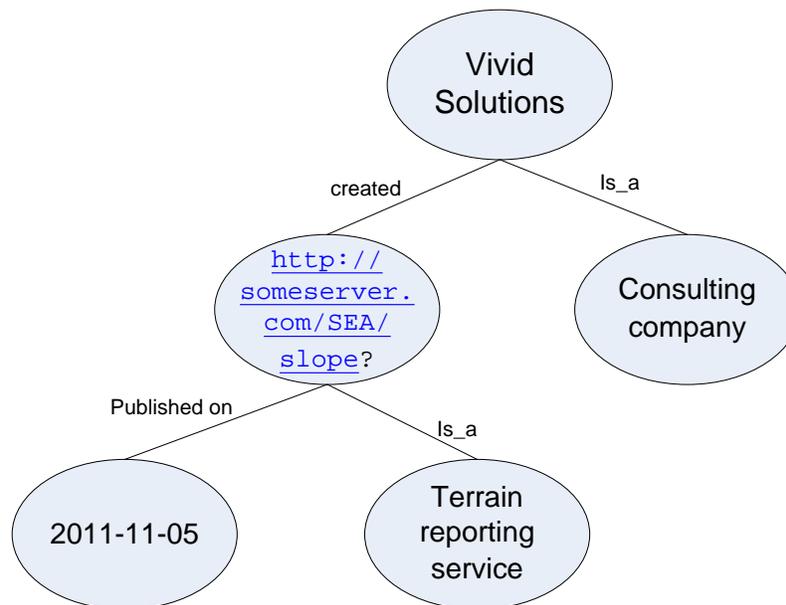
This particular graph has no loops but unlike the hierarchical structure of a taxonomy there is no restriction on loops.

In the above graph each node-link-node represents a resource-property-value triplet. A collection of such triplets (one or more) define a graph. To model these graphs in a computer accessible way the W3C has developed a standard called the Resource Description Framework and within this standard a triplet is often called a RDF triplet. Future work on discovery services and publishing metadata to support such services will almost certainly follow this standard.