NorthWest Data Sharing Network Phase 2

2004 NWDSN Project Summary

Submitted To: NWDSN Steering Committee

Created By: Philip Kayal and Samuel Smith
Refractions Research Inc.
Suite 400 - 1207 Douglas Street
Victoria, BC, V8W-2E7
E-mail: pkayal@refractions.net
Phone: (250) 383-3022
Fax: (250) 383-2140

Table of Contents

1. Introduction

1.1. Background

This document reports the status of the operational datasets that were provided by Northwest Data-Sharing Network (NWDSN) partners for incorporation into the NWDSN data holdings by 31 March 2004.

Data received from the providers are catalogued against a list of all layers requested by the NWDSN. Noted are the layers that were received from each partner, the format(s) that the data were provided in, the appearance of the requested attributes, and finally the results of and any other comments relating to the forthcoming data treatments.

Upon receiving datasets from partners, files were backed-up on a machine external to the NWDSN, translated into working Shapefiles and pre-processed so as to facilitate translation into the relevant final standardized and geometrically valid format.

Each file standardized into its final format was subsequently loaded into the NWDSN warehouse's spatial database using a native translator, alongside a series of simple but effective QA/QC checks for valid geometries and consistent feature counts.

Finally, this document summarizes the current status of the NWDSN project and makes a series of recommendations for future stages of the NWDSN project.

1.2. Supporting Documents

If they have not already done so, readers are encouraged to familiarize themselves with the following documents:

- ***NWDSN Interim Data Model*** by P. Kayal and S. Smith. (NWDSN Interim Data Model.PDF)

- Northwest Data Sharing Network Phase 2 - **Task 2: Data Modeling for Data Exchange -** Data Integration Model by D. Evans, P. Kayal and S. Smith. (NWDSN_Data_Integration_Model_v1.PDF and NWDSN_Data_Integration_Model_v1_ADDENDUM.PDF)

- ***Data Sharing Network Phase 2 – Task 1: Local Data Administration*** by Rick Wheeler. Information regarding the spatial data and metadata collected from NWDSN stakeholders. (IFPA_DSN_Report.PDF)

- ***NW Data Sharing Network: Data Integration Report*** by P. Kayal and D. Evans. A complete assessment of the input datasets received, and a preliminary integrated data model. (NWDSN-Data_Integration_Report.PDF)

These documents are available for download as Adobe .PDF files from the NWDSN FTP server. From the login directory, navigate to the "docs" subfolder – file names follow the document title (in parentheses).

1.3. Requested Data Layers

Initial NWDSN data models requested a large number of operational datasets from participants. However, to ensure that data were received with enough time to build an online system for viewing and sharing data between providers, this list was pared down to fewer, more critical

layers. In a 16 February 2004, teleconference between the NWDSN Technical Committee and Refractions Research Inc., it was decided that the six NWDSN data providers would provide, from their individual data holdings, as many of the following seven data types as possible:

- Forest Block Features – Openings (Polygonal)
- Forest Block Features – Disturbances (Polygonal)
- Forest Block Features – Silviculture (Polygonal)
- Forest Block Features – Silvicultural Activity (Polygonal)
- Forest Road Features – Roads (Linear)
- Hydrographic Features – Lakes/Waterbodies (Polygonal)
- Hydrographic Features – Streams/Rivers (Linear)
- Hydrographic Features – Wells/Points of Diversion (Point)

Most of the participants were confident that they could provide some amount of Forest Block information, as well as information for Forest Roads. A few of the providers felt that they would be able to provide their internal hydrographic data layers (streams and lakes), which would be super-imposed against but not physically incorporated into TRIM Extended Base Map (EBM) and FC1 hydrographic data.

1.4. Requested Fields

Earlier NWDSN data models requested, for each data layer provided, the entire set of attributes solicited by the BC Ministry of Forests' Electronic Submission Framework data model. Where the February 2004 data-model requested as many as possible of fewer data types, it also requested as many as possible of a smaller set of critical attributes for each data-type. The interim list of requested attributes by data-types is as follows:

Forest Block Features – Common Fields

The following fields were commonly requested from all Forest Block Features (Openings, Disturbances, Silviculture, Silvicultural Activity):

Name	Description	Format	Example
LICENCEID	Forest Licence Identification Number	CHAR (12)	A12304
PERMITID	Forest Cut Permit Identification Number	CHAR (12)	CP406
BLOCKID	Forest Cut Block Indentification Number	CHAR (12)	BLKA
SOURCE	Data Source / Provider Code	CHAR (12)	CANFOR
UNIQUEID	Feature Identifier Unique to the Data Provider	CHAR (12)	2903837
CAPTUREID	Data Capture Method Code	CHAR (12)	GPS
STATUSID	Cut Block Activity Status Code	CHAR (12)	Logged

Forest Block – Openings

Forest Block Opening feature attributes were the same as the common list for all block features; there were no additional attributes.

Forest Block – Disturbances

In addition to the common Forest Block fields, the following three attributes were requested for Forest Block Disturbances:

Name	Description	Format	Example
DSTRBCODE	Code Describing the Disturbance on the Block	CHAR (12)	Logged
DSTRBSDATE	Disturbance Start Date	DATE (YYYYMMDD)	20040412
DSTRBEDATE	Disturbance End Date	DATE (YYYYMMDD)	20040412

Forest Block – Silviculture

In addition to the common Forest Block fields, the following two attributes were requested for Forest Block Silviculture Polygons:

Name	Description	Format	Example
STUNID	Unique Identification Code for the Standard Unit	CHAR (12)	8383099
STUNTYPE	Code Describing the Type of Standard Unit	CHAR (12)	NPR

Forest Block – Silvicultural Activity

In addition to the common Forest Block fields, the following six attributes were requested for Forest Block Silvicultural Activity Polygons:

Name	Description	Format	Example
STUNID	Unique Indentification Code for the Standard Unit	CHAR (12)	93809039
BASETECH	Code describing the basic silvicultural technique employed on the standard unit	CHAR (12)	PLANT
BASEMETH	Code describing the basic silvicultural method employed on the standard unit	CHAR (12)	CREWED
BASEOBJV	Code describing the basic objective of the silvicultural activity being performed on the standard unit	CHAR (12)	100/HA
ACTIVSDATE	Silvicultural Activity Start Date	DATE (YYYYMMDD)	20040412
ACTIVEDATE	Silvicultural Activity End Date	DATE (YYYYMMDD)	20040412

Forest Road Features

The following fields were requested from all Forest Road Features:

Name	Description	Format	Example
RPERMITID	Forest Road Permit Identificatin Number	CHAR (12)	R083993
RSECTIONID	Forest Road Section Identification Number	CHAR (12)	SEC203
SOURCE	Data Source / Provider Code	CHAR (12)	HFP
UNIQUEID	Feature Identifier Unique to the Data Provider	CHAR (12)	280932-4
CAPTUREID	Data Capture Method Code	CHAR (12)	TABLET
ROADNAME	The common name of the forest road feature	CHAR (55)	Dove Creak Mainline
ROADCLASS	The classification of the forest road feature	CHAR (55)	SPUR
ROADSTATUS	The active status of the forest road feature	CHAR (12)	Deactivated

Polygonal Hydrographic Features

The following fields were requested from all Polygonal Hydrographic Features:

Name	Description	Format	Example
SOURCE	Data Source / Provider Code	CHAR (12)	BABINE
UNIQUEID	Feature Identifier Unique to the Data Provider	CHAR (12)	809890-3
CAPTUREID	Data Capture Method Code	CHAR (12)	SKETCH
LAKECLSS	Waterbody Classification Code	CHAR (12)	L5
LAKEFISH	Have fish been confirmed in the lake?	CHAR (12)	YES/NO
FISHSPCO01	Code of the first most prevalent fish species	CHAR (55)	BCG
FISHSPCO02	Code of the second most prevalent fish species	CHAR (55)	SPT
FISHSPCO03	Code of the third most prevalent fish species	CHAR (55)	SAL

Linear Hydrographic Features

The following fields were requested from all Linear Hydrographic Features:

Name	Description	Format	Example
SOURCE	Data Source / Provider Code	CHAR (12)	FRASER
UNIQUEID	Feature Identifier Unique to the Data Provider	CHAR (12)	9304903
CAPTUREID	Data Capture Method Code	CHAR (12)	PHOTO
STREAMCLSS	Stream Classification Code	CHAR (12)	L5
STREAMFISH	Have fish been confirmed in the stream?	CHAR (12)	YES/NO
FISHSPCO01	Code of the first most prevalent fish species	CHAR (55)	BCG
FISHSPCO02	Code of the second most prevalent fish species	CHAR (55)	SPT
FISHSPCO03	Code of the third most prevalent fish species	CHAR (55)	SAL

Point Hydrographic Features

The following fields were requested from all Point Hydrographic Features:

Name	Description	Format	Example
SOURCE	Data Source / Provider Code	CHAR (12)	PIR
UNIQUEID	Feature Identifier Unique to the Data Provider	CHAR (12)	909043-4
CAPTUREID	Data Capture Method Code	CHAR (12)	SURVEY
POINTTYPE	Code describing the type of Hydrographic Point	CHAR (12)	WELL
HPVALUE	Magnitudinal value describing a measure associated with the Hydropoint	FLOAT (6,2)	12.78
HPDESCR	General Description of the Hydropoint	CHAR (55)	Depth (m)

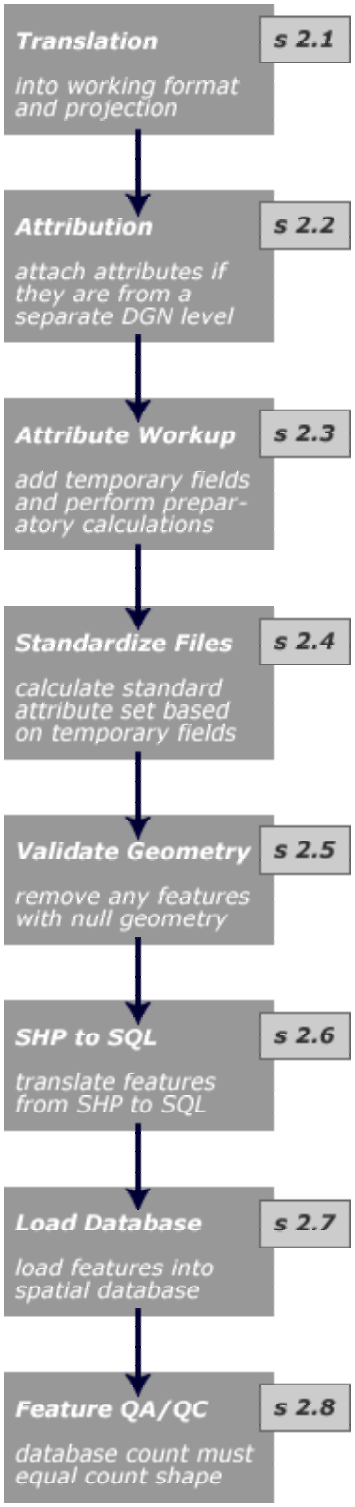
1.5. Acceptable Data Formats

The NWDSN data model, initially proposed early in 2004, requested that all data be delivered in a hybrid eXtensible Markup Language (XML):Geographic Markup Language (GML) format, describing both the attributes, and spatial extents of NWDSN features, respectively. While the XML:GML data format complied with an MOF mandate to move all forest company licencee data submissions to an open-source spatial data format, discussion in February 2004 revealed that not all of the NWDSN participants would be able to deliver their NWDSN provisions in XML:GML format by the proposed delivery date of 27 February 2004. As such, providers would submit their data to the NWDSN in one of the following formats (in order of preference):

XML:GML (.XML) > **ESRI Shapefile** (.SHP) > **Microstation Design File** (.DGN)

2. Data Preparation Process

NWDSN Data were worked up according to the following workflow diagram:



Numbers (in light grey tabs) relate the steps in this diagram to sections in this document.

2.1. Working Format and Projection Translations

Data were submitted to the NWDSN in either .SHP or .DGN (no XML:GML submissions were made) with a variety of attribution schemas. Different attribution schemas needed various amounts of recalculation to be compatible with the NWDSN data models previously described. Additionally, files provided in .DGN format required as a minimum Feature Manipulation Engine (FME) translation into .SHP format. Where the target data layer was polygonal, varying amounts of polygon building were effected using custom FME scripts.

The working projection of the NWDSN is the conformal conic BC Albers, commonly used to project this region of the world.

2.2. Attaching Point Based Attributes

Of the datasets provided in .DGN, none implicitly attached attributes to features. In some cases attributes were attached to points that were graphically grouped to the linework that comprised parent polygons. Wherever feasible, using a combination of custom FME and ArcView Avenue scripts, as many attributes as possible were tied back to the relevant .SHPs resultant from .DGNs.

2.3. Attribute Workup

In preparation for the conversion of working .SHPs into final standardized versions that would be consistently loaded into the NWDSN spatial database, a set of "working attributes" was added to each of the intermediate Shapefiles. Working attributes sets were identical to the minimum set of attributes for each NWDSN feature type described previously.

The addition of these working fields allowed mandatory field values to be calculated (e.g. *StringFieldC* = *StringFieldA* + *StringFieldB*), cast (e.g. 12 to "12") and/or assigned ("NoData") from existing fields without overwriting values in any of the original fields. Original fields were retained in each of the files.

2.4. Shapefile Standardization

A small set of ArcView Avenue scripts was developed to automatically create a consistently named and attributed Shapefile for each of the temporary Shapefiles that were worked up previously with the temporary attributes.

2.5. Selection of Legitimate Geometries

Initial attempts to load some Shapefiles into the spatial database were thwarted by null geometries in some of the subject Shapefiles, particularly those whose origins were a spatial database system (ArcSDE, GeoDatabase or other). These materialized in the Shapefiles as database records for which there was no corresponding spatial feature, and were easily removed by marquee selecting across the maximum extent of the entire theme in an ArcView View window and converting the selection to a new Shapefile. Only features with visible, thus legitimate geometries were selected and consequently exported.

2.6. Shape-To-SQL Translations

Standardized Shapefiles were re-uploaded to the NWDSN FTP server for spatial loading into the database.

Shapefiles were named as per the following table and stored in a subdirectory of the providers' folder named "standardized".

Naming schema: *<provider>_<type>_<subtype>.shp*		
Where <provider> is one of: • BABINE • BCTS • CANFOR • FRASER • HFP • PIR	Where <type> is one of: • Blocks • Roads • Hydro	Where <subtype> is one of: [for <type> = Blocks] • Openings • Disturbances • Silviculture • SilviculturalActivity [for <type> = Roads] • Roads [for <type> = Hydro] • Polys • Lines • Points

Once uploaded, Shapefiles were translated into file-based geometric SQL statements using the following syntax:

shp2pgsql <provider>_<type>_<subtype> <provider>_<type>_<subtype> nwdsn **&>** <provider>_<type>_<subtype>**.sql**

Note that the filename of the .SQL file is identical to that of the Shapefile as is the name of the table (both without an extension) in the NWDSN database.

2.7. Database Loading

.SQL files were run to load geometries into the NWDSN database using the following syntax:

psql –U postgres nwdsn –f <provider>_<type>_<subtype>**.sql &>** <provider>_<type>_<subtype>**.log**

This command uses the .SQL file created previously and outputs the results of the translation to an identically named .LOG file. LOG files were checked for the absolute absence of any load errors.

2.8. Post Load QA/QC

Once loaded, each operational table in the NWDSN database was checked by feature count. The number of records in the NWDSN.<provider>_<type>_<subtype> table object was expected to equal the number of records in the standardized and validated Shapefile previously created.

As shown in the data summary tables each layer immediately or eventually passed the QA/QC criteria.

3. Current Project State

3.1. NWDSN Partner Data Holdings

As NWDSN datasets were uploaded by FTP to the NWDSN server, Refractions Research catalogued the format and nature (eg. data type) of the upload. As well, feature counts were recorded at each stage of the work-up, standardization and loading processes. Balanced feature counts between the standardized Shapefiles (with non-null geometries) and the resultant tables served as a measure of QA/QC on the data preparation and loading process. A tabular summary of all datasets uploaded and worked up for use in the NWDSN follows:

		BLOCKS				ROADS	HYDROGRAPHIC		
		openings	disturbances	silviculture	silvactivity	roads	lines	points	polygons
BABINE	Format In	SHPn (*1)	SHPn	SHPn	nil	SHPn	nil	nil	nil
	Shape Recs	1382	991	2307	nil	3708	nil	nil	nil
	Valid Recs	1382	991	2307	nil	3672	nil	nil	nil
	DB Recs	1382	991	2307	nil	3672	nil	nil	nil
	Pass QA/QC	PASS	PASS	PASS	nil	PASS (*4)	nil	nil	nil
BCTS	Format In	GEODB (*2)	nil	nil	nil	SHPn	nil	nil	nil
	Shape Recs	1991	nil	nil	nil	4339	nil	nil	nil
	Valid Recs	1987	nil	nil	nil	4339	nil	nil	nil
	DB Recs	1987	nil	nil	nil	4339	nil	nil	nil
	Pass QA/QC	PASS	nil	nil	nil	PASS	nil	nil	nil
CANFOR	Format In	SHPn	SHPn	SHPn	nil	SHPn	SHPn	nil	SHPn
	Shape Recs	5017	4378	11194	nil	5870	70307	nil	16760
	Valid Recs	5017	4378	11194	nil	5870	70298	nil	16752
	DB Recs	5017	4378	11194	nil	5870	70298	nil	16752
	Pass QA/QC	PASS	PASS	PASS	nil	PASS	PASS	nil	PASS
FRASER	Format In	DGN (*3)	nil	nil	nil	DGN	nil	nil	nil
	Shape Recs	465	nil	nil	nil	1756	nil	nil	nil
	Valid Recs	465	nil	nil	nil	1756	nil	nil	nil
	DB Recs	465	nil	nil	nil	1756	nil	nil	nil
	Pass QA/QC	PASS	nil	nil	nil	PASS	nil	nil	nil
HFP	Format In	SHPn	nil	SHPn	nil	SHPn	nil	nil	nil
	Shape Recs	2751	nil	3906	nil	5606	nil	nil	nil
	Valid Recs	2751	nil	3906	nil	5606	nil	nil	nil
	DB Recs	2751	nil	3906	nil	5606	nil	nil	nil
	Pass QA/QC	PASS	nil	PASS	nil	PASS	nil	nil	nil
PIR	Format In	DGN	nil	nil	nil	nil	nil	nil	nil
	Shape Recs	471	nil	nil	nil	nil	nil	nil	nil
	Valid Recs	471	nil	nil	nil	nil	nil	nil	nil
	DB Recs	471	nil	nil	nil	nil	nil	nil	nil
	Pass QA/QC	PASS	nil	nil	nil	nil	nil	nil	nil

Table Notes:
(*1) “SHPn” denotes a non-standard Shapefile
(*2) “GEODB” denontes a GeoDatabase
(*3) “DGN” denotes a Microstation Design file
(*4) “nil” indicates that no data were uploaded

Table Colour Key:
___ = Data submission format
___ = Congruent record numbers
___ = Discrepant record numbers

3.2. Review of Project Deliverables

Refractions Research has completed the following steps to address and achieve the specific requested deliverables of the NWDSN project:

1. *Working with MSRM BISD, determine an existing or establish an interim data standard and schema that supports the integration and dissemination of the data without loss of strategic information. The standard and schema will also support the loading of the data into MSRM's SDE.*

   Two data standards were developed and documented during this project: future (XML:GML) data standards and interim (Shape) data standards. The contractor team worked closely with MSRM and NWDSN stakeholders to completely understand their needs and to define the optimal data schemas. The attribution associated with spatial objects is sufficiently rigorous to describe all relevant strategic information.

   Integrated data can be loaded into SDE. The future XML:GML data standards and schema conform with MSRM's Electronic Submission Framework.

2. *Provide data format solutions that support interim data sharing network objectives without compromising a longer term GML format based solution. Ensure that the proposed Logical Data Models support SFM planning requirements. The solution will be consistent with stakeholder operational and business requirement priorities stated in the Governance.*

   The interim data standards were specifically developed with a long-term XML:GML solution in mind. In fact, complete XML:GML data models were fully developed prior to the creation of interim data models.

   The structure and content of the integrated data models support Sustainable Forest Management planning requirements.

   The contractor team worked closely with NWDSN stakeholders throughout this project, to ensure that data models are consistent with stakeholder operational and business requirements.

3. *Assess the MSRM's GML standards as of April, 2003 and propose a schedule to develop and implement an operational integrated data solution by March 31, 2004.*

   The project commenced much later than April 2003. It was late in July 2003, when *Task 2: Data Modeling for Data Exchange*, was approved. The project was put on hold in October, before being restarted in mid-November. So the time-frame for this project shifted considerably. Most of the work was done between December 2003 and March 2004.

   The schedule that was originally proposed was revised in July, and then revised again in November. An interim integrated data solution has been achieved and fully implemented effective March 31, 2004.

4. *Clearly identify the differences in the MSRM/MOF and DSN Logical Data Models as of May, 2003 and report on Integration and Data Warehouse concerns by July, 2003.*

   As mentioned above, the time-frame for this project shifted considerably.

   There were a number of MSRM/MOF data exchange standards that were investigated by the contractor project team. All of the NWDSN integrated data models were specifically developed to ensure consistency with existing and proposed MSRM/MOF standards.

The main MSRM/MOF data exchange standards that were followed include:

- ESF – Electronic Submission Framework (http://www.for.gov.bc.ca/his/esf/)
- RESULTS – ESF Silviculture (http://www.for.gov.bc.ca/his/esf/index_ressub.htm)
- ABR – As-built Roads Electronic Submission

The current state of the LRDW should be investigated at the time when the NWDSN is ready to move to an XML:GML data schema.

5. *Provide an effective automated translator of stakeholder internal data to an agreed on common integrated schema, format and projection. The format must be portable and scaleable (into SDE) and consistent with the requirements of the LRDW.*

   Data processing scripts have been written to translate stakeholder data sets into the schema, format and structure of the integrated data model. Scripts are written to allow easy modification, if the stakeholder data delivery schemas change in the future. There are data processing scripts for:

   - Blocks:
     - Openings
     - Disturbances
     - Silviculture
     - Silvicultural Activity
   - Roads
   - Hydrology
     - Polygons
     - Lines
     - Points

   All of the automated data processing scripts are available on the NWDSN server.

   In addition, there are data processing scripts for many of the non-standard data formats encountered during this project, including IGDS.

   Data models and data are portable and can be imported into SDE within the LRDW.

6. *We anticipate the need for limited data administration April 1, 2003 to March 31st, 2004.*

   Data administration has been provided during this period; primarily by Jeff Lounsbury.

4. Summary and Future Directions

Looking towards future iterations of the NWDSN project, Refractions Research has identified several areas, which would add support to the project's requirements and enhance its functions.

Earlier NWDSN data integration models proposed using a single common format for all data submissions. In order to develop a working system in time for a 31 March 2004 delivery date, data were accepted in whichever of three common spatial data formats NWDSN partners were able to provide. The BC MOF has stipulated that within the foreseeable future, all forest inventory and activity data submissions will be made in XML:GML format – using a structure/standard published by MOF (Results, FTA and/or ABR). Refractions encourages the use of XML:GML as the common NWDSN submission format. Adopting this single data format structured to a defined standard, not only supports the MOF mandate, but would also facilitate automatic NWDSN partner data updates using a single set of server-side translation scripts.

Finally, in the interest of providing a complete set of information, all partners are encouraged to review their data holdings and where gaps exist in their NWDSN provision, do their utmost to supply any outstanding data.