# Yellowstone to Yukon Framework Data Demonstration Project
## Final Report

Abstract:

The Transboundary Framework Data Project was a collaborative demonstration GIS data compilation effort focused on creating seamless GIS data sets across local, state, and national boundaries in the U.S./Canada region of the Rocky Mountains. In June 2000, twenty-two partners entered into a contract with the United States Federal Geographic Data Committee and Canada's federal GeoConnections (Natural Resources Canada) in a cooperative effort to compile framework datasets for the Crown of the Continent area of the Rockies from the best available sources. These data were tested, documented with FGDC compliant metadata, made searchable through on-line clearinghouses, and made available (where licensing allowed) to the public. The project resulted in a set of nine vertically integrated framework data layers.

Project Summary:

Modern high-speed processors and the pushbutton analysis capabilities of modern software allow complex spatial analysis and modeling to be performed relatively effortlessly. For most projects, the logjam in the analysis process continues to be availability of high-quality spatial data at scales appropriate for the analysis. Additional complexities in compiling data occur when attempting to perform analyses across jurisdictional boundaries. Differences in data models, source scale, and availability make compiling even the framework data a challenge. Different branches of the same agency, different agencies, or even different countries may manage similar adjacent land parcels. Data from some sources may be free and easily downloaded from the Internet, while other sources for similar data may have a fee structure and carry significant licensing and use restrictions.

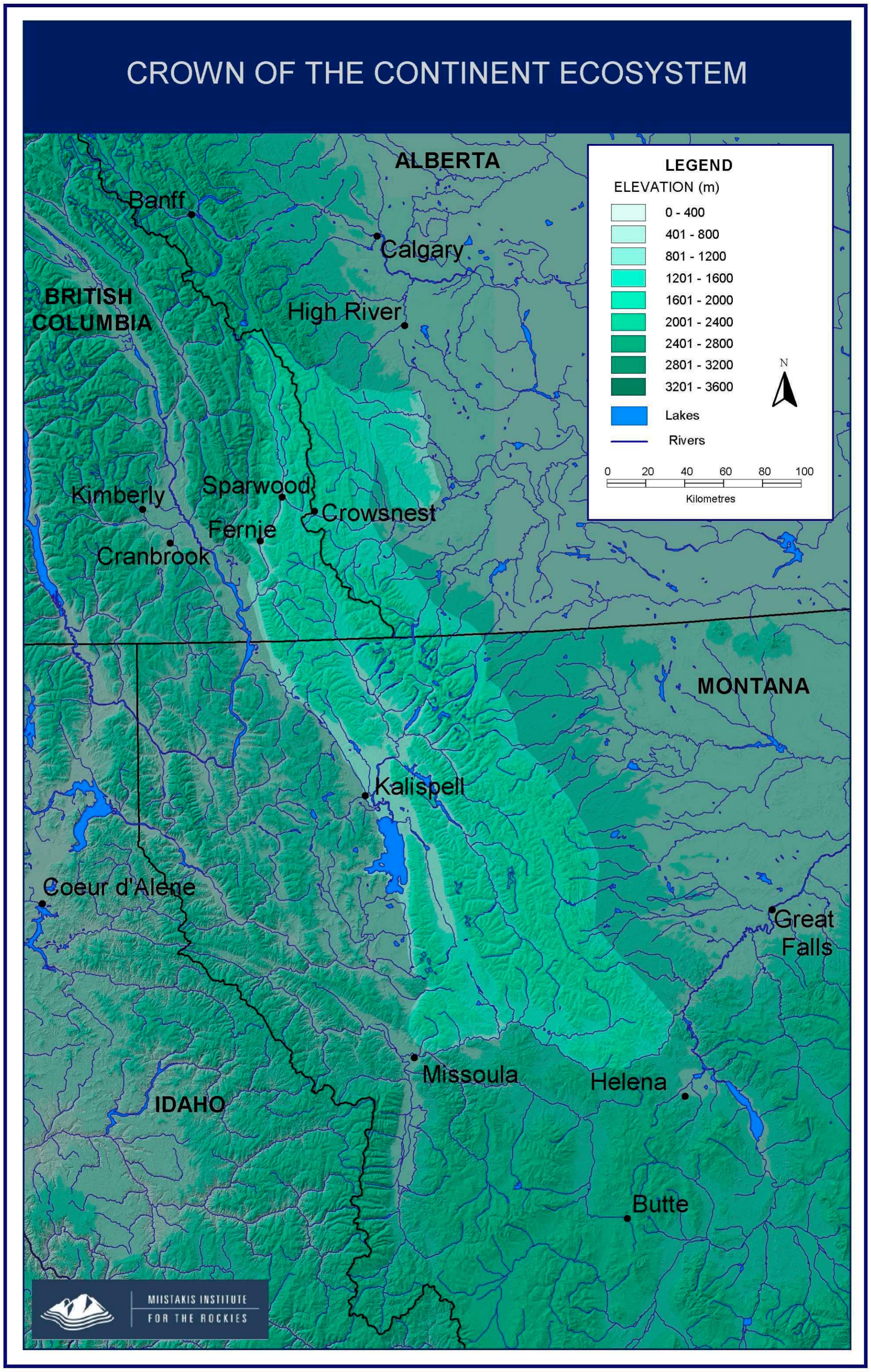
Wildlife and natural resource modeling and analysis projects are perhaps most affected by these and other technical challenges. Correctly identifying and prioritizing wildlife habitat is often dependent on merging cohesive data sets across jurisdictional boundaries. These challenges continue to grow as the spatial scale of the project expands across state, provincial, and national boundaries. Geographic Information System (GIS) analyses and mapping are heavily used to these ends. To perform analysis at these scales requires overcoming obstacles ranging from resolving different approaches to modeling the Earth's surface (datums, map projections), to converting between the different hardware and software platforms used, to sharing data due to varying approaches to copyrights.

In June 2000, twenty-two GIS-oriented partners entered into a contract with the United States Federal Geographic Data Committee (of the U.S. Geological Survey) and Canada's federal GeoConnections (Natural Resources Canada) to overcome these obstacles for an area straddling the U.S./Canada border. In exchange for funding, the partnership agreed to merge key datasets in an area known as the Crown of the Continent Ecosystem (Figure 1). Furthermore, with the assistance of the U.S. National Biological Information Infrastructure (NBII), the partners agreed to set up two NBII 'nodes' capable of distributing these datasets and associated metadata with other GIS users.

This effort arose out of a need identified by The Yellowstone to Yukon Conservation Initiative (Y2Y). Y2Y seeks to protect wildlife habitat and movement corridors through a variety of parks, zoning and easement mechanisms throughout the northern Rocky Mountains and adjacent ranges of North America. The Initiative, comprised of over 150 separate organizations, is the largest land-based conservation effort on this continent. As such, it faces unprecedented challenges in achieving its vision across multiple jurisdictions.

On November 19-21, 1999, Y2Y convened a diverse group of top scientists in Jasper, Alberta, with particular strengths in large carnivore and aquatic ecology. Led by the International Union for the Conservation of Nature's (IUCN) bear committee chair, Dr. Stephen Herrero, and the founder of conservation biology, Dr. Michael Soule, this group identified the key data layers needed to address transboundary ecosystem questions on such a broad scale. Specifically, in order to develop a transboundary strategy that addresses ecological connectivity and integrity questions, they recommended that Y2Y develop fifteen themes of information.

On April 13-14, 2000, GIS analysts and data managers from throughout the United States' and Canada's Intermountain West convened to assess the status of existing framework datasets and mechanisms needed to improve these layers of information across the Yellowstone to Yukon region. Several GIS datasets consistently reoccurred as they evaluated each of the scientist's themes – these were deemed 'framework' data.


CROWN OF THE CONTINENT ECOSYSTEM
ALBERTA
BRITISH COLUMBIA
MONTANA
IDAHO
Banff
Calgary
High River
Sparwood
Kimberly
Crowsnest
Fernie
Cranbrook
Kalispell
Coeur d'Alene
Great Falls
Missoula
Helena
Butte
LEGEND
ELEVATION (m)
0 - 400
401 - 800
801 - 1200
1201 - 1600
1601 - 2000
2001 - 2400
2401 - 2800
2801 - 3200
3201 - 3600
Lakes
Rivers
N
0 20 40 60 80 100
Kilometres
MIISTAKIS INSTITUTE FOR THE ROCKIES

The core layers that emerged as Y2Y's framework dataset were:

- Hypsography (landform or Digital Elevation Model) (Resolution: 25-30m)*
- Landsat 6/7 TM and regional IRS ortho imagery (Resolution: 30m/15m with 5mIRS samples)*
- Hydrology (Scale: 1:20000/1:100000)*
- Roads and trails (Scale: 1:20000/1:24000)*
- Resource extraction (well sites, cut blocks) (Scale 1:20000/1:24000)*
- AVHRR-derived land cover (Resolution: 1km)*
- Government units/Stewardship (Scale: 1:20000/1:24000)*
- Cadastral (parks, wilderness areas, national/state forests, private land)*
- Climate (varies)
- Census (to rural census division)
- Toponymy (place names) (Scale: 1:20000/1:24000)*
- Geodetic control to ensure consistent integration of layers*

Secondary layers included:

- Land claims/reserves/cultural elements (Scale: 1:20000/1:24000)*
- Land use zoning (Scale 1:20000/1:24000)
- MODIS imagery (assorted sensors)
- Watersheds (to HUC 6)
- Species richness (index of biodiversity)
- Special elements

*Themes compiled for this project

Project Activities:

Upon award of the grant in late April 2000, the project organizers assigned responsibility for compiling and documenting specific framework layers to two Canadian and three U.S. GIS organizations. These assignments were made based on each shop's interests, expertise, and experience with a particular data set. At this point, it was important to get a solid understanding of all available data sources for each of the framework layers. This involved a search and review of various federal and state/provincial data sources from both the U.S. and Canada. Each shop then assessed the utility of each data layer for creating a "seamless" transboundary data layer. Important considerations included having similar scale and attributing across boundaries. A recommendation was then made for the source data and attributes to be included in each framework layer.

A project workshop was then held in August 2000 in Calgary to review each data layer and agree on the set of standards each shop would use for the dataset. These included deciding on a target scale, projection parameters, and distribution format. From this meeting, a Draft Framework Architecture Report, including a data dictionary, was compiled and distributed to the project team. This report served as the guidelines for preparing each data layer.

From August 2000 through February 2001, each shop compiled and documented their assigned layers. FGDC compliant metadata was prepared for each data layer for distribution with the dataset and for incorporation into a National Spatial Data Infrastructure (NSDI) spatial data clearinghouse node. The Faculty of Geomatics Engineering at the University of Calgary then peer reviewed each layer for consistency, spatial accuracy, and vertical integration. The resulting report "Y2Y Framework Dataset Review" serves as a source for understanding the quality of resulting data and also as a guideline for improving the dataset in the future.

The task of making the data searchable through NSDI registered spatial data clearinghouses is being implemented on both the U.S. and Canadian sides. For the U.S., the project team entered into an informal partnership with the USGS Northern Rocky Mountain Science Center (NRMSC) in Bozeman, Montana. The NRMSC is in the process of setting up a National Biological Information Infrastructure (NBII) clearinghouse node to serve data pertinent to the Northern Rocky Mountain region. This node is in affiliation with the existing Greater Yellowstone Data Clearinghouse (GYADC) housed at Montana State University, Bozeman. For more information on this node, please visit the NRMSC web site at http://www.nrmsc.usgs.gov/. The Miistakis Institute for the Rockies hosts the Canadian Clearinghouse. The Canadian site focusses on running simultaneous spatial and biodiversity data nodes with the FGDC server and a NABIN-approved Species Analyst server running concurrently. Examples of dataset use will be available from Miistakis in summer 2001 through an online mapping application (ARCIMS or equivalent).

The Canadian spatial data server will be registered through the Canadian CEONET gateway and linked to Environment Canada's Ecological Monitoring and Assessment Network (EMAN). This should allow it to be searched from any FGDC portal on the web. A user will be able to view images of the data and metadata online before linking to downloads. Because certain Canadian data sets carry specific licensing agreements and are only available on request to certain groups or individuals, users requesting this data will first link to a page detailing the use restrictions and requiring them to become a member of the Conservation Data Consortium[1] before receiving the data. Copyright considerations are the only delays in registering the spatial node online. Once these are resolved, the project data will be obtainable through a NSDI registered spatial data Clearinghouse node.

Project Outcomes:

The project resulted in eleven GIS data layers for the study area (Table 1). These include vector data sets for features such as roads and boundaries, point data sets for geodetic control and oil/gas wells, grid data sets for elevation and some imagery, and raster images. In total, the data fits onto three CDs, excluding the imagery. The digital elevation models (DEMs) comprise the bulk of the data.

| Dataset | Scale | Coverage | Tiled? |
|---|---|---|---|
| DEM | 1:20 000 Canada<br>1:24000 U.S. | Crown of Continent plus | Yes – 100K |
| Hydrology | 1:20 000 Canada<br>1:100 000 U.S. | Crown of Continent | Yes – 100K |
| Transportation | 1:20 000 Canada<br>1:24000 U.S. | Crown of Continent | Yes – 100K |
| Imagery | Varies | Crown of Continent plus | No |
| Resource Extraction | 1:20 000 Canada<br>1:24 000 U.S | Crown of Continent | No |
| AVHRR Land Cover | 1.1 km | Yellowstone to Yukon | No |
| Gov't Units/Cadastral | 1:250 000 | Yellowstone to Yukon | No |
| Cultural | 1:250 000 | Crown of Continent | No |
| Toponymy | 1:20 000 | Yellowstone to Yukon | No |
| Geodetic Control | 1:20 000 | Yellowstone to Yukon | No |
| Watersheds | HUC 5 | Crown of Continent | No |

Table 1. Summary of Framework Datasets Completed

[1] The Conservation Data Consortium is a GIS data user group organized to help facilitate the sharing of geospatial data.

The geographic extent of each layer varies.  Each layer covers the Crown of the Continent study area as a minimum. Several layers cover the entire Yellowstone to Yukon region (Figure 2).  The actual extent of coverage was left to the partners who were compiling the data.  Generally the study coverage was extended where data was easily available and little time was required to compile the additional data.  Large datasets are tiled along a 1:100 000 National Topographic Series grid (Canada) and its U.S equivalent (1 degree latitude by 2 degrees longitude) for ease of distribution.


Figure 2: Crown of the Continent Study Area within Yellowstone to Yukon Region

Besides the datasets themselves, the project also resulted in several reports:

**Yellowstone to Yukon Framework Data Architecture** – an overview of framework datasets available in the Yellowstone to Yukon Region including all standards adhered to in assembly and a data dictionary for each theme.

**Yellowstone to Yukon Network Architecture Report** – an overview of the distribution system for framework datasets. The architecture report functioned as a progress report for the project.

**Yellowstone to Yukon Framework Dataset Review** – an external review of the quality of both metadata and data compiled through the project

**Cumulative Effects Analysis – an example application of Transboundary Framework Datasets** – results of a pilot study in the Crown of the Continent Region applying transboundary framework datasets.

Finally, the project resulted in the construction of metadata and the establishment of a clearinghouse server at the Miistakis Institute for the Rockies. The server will be mirrored by one at the Northern Rocky Mountain Science Centre as described in the previous section.

The project and the FGDC/GeoConnections contribution was described in presentations at both the Canadian Biodiversity Network Conference in Ottawa (March 2001) and the Intermountain GIS Conference in Boise (April 2001).

<u>Issues, commonalities, difficulties and challenges:</u>
The project team was able to accomplish a great deal on a very limited budget. This is attributable to the long-term relationships shared by the partner organizations. Many of the participants have been performing GIS modeling and analysis in the Yellowstone to Yukon region for many years. This gave the project team a significant head start on data collection and compilation. In some cases, the entire data set already existed at a partner facility and it was only necessary to clean-up, verify, and document the data.

The primary challenges resided in merging data of varying source scales and attributes. In most cases this matter was handled by selecting source data with similar scales. For example, 1:24,000 scale on the U.S. side and 1:20,000 scale on the Canadian side. The positional accuracy and geometric characteristics of each data set are similar and merge quite well across the border. In other situations the fit was not as good. For example, the hydrologic coverage on the U.S. side is comprised of 1:100,000 scale National Hydrologic Dataset (NHD) features. On the Canadian side, the hydrology consisted of TRIM data from BC and various data sources from Alberta. These data sets have radically different stream densities and spatial accuracy. These can sometimes be resolved by filtering on certain attributes such as stream order, but these convenient filtering attributes are not always available.

For many reasons, compiling consistent attributes across the U.S./Canada border was particularly challenging. Each country uses a different data model in order to attribute their features. For example, while the oil/gas attribute tables shared many common fields such as Total Depth and Owner, the hydrologic attribute tables shared no common fields. This makes it very difficult to compile a consistent attributing system for the unified data sets. Where possible, common fields were appended and not duplicated. Uncommon fields were retained in each data set where different jurisdictions had different attributes. Additionally, an effort was made to create “crosswalk” tables to help bridge the attribute differences between data sets.

Attribute challenges surfaced particularly in transportation files. Unlike hydrology or resource extraction layers, road files are compiled by a number of jurisdictions. Fine scale data is often available from a number of sources for the same areal coverage and each have strengths and weaknesses. A blend of TIGER, USGS DLG, National Forest, National Park and proprietary datasets are needed to produce the best results. Creating this blend, then matching it across the Canadian border is quite time-consuming. Our study area involved five national forests and each had widely varying quality and quantity of attribute data. As two forests had no database behind them at all, yet had the finest available topology, we were restricted in offering a seamless classification that worked across the whole study area.

Another challenge that the team faced with many of the data sets resides in each country’s approach to data licensing and availability. All of the U.S. data used for the framework data is freely available to anybody. This is a result of the Freedom of Information act that puts information collected with public dollars available to the public. In Canada, data collected with public dollars generally remains the property of the Crown or Provincial governments. The data is licensed to the purchaser, and cannot passed to another entity without violating the licensing agreement. The cost of the data is prohibitively expensive for most potential users. The result is a geomatics industry that is often severely limited by the availability of low-cost, high-quality data. The impact this has on the project is that most of the core framework data sets on the Canadian side are licensed and not freely distributable. This situation is seeing some movement toward lowering costs or eliminating them altogether, though the final outcome is still unknown.

Vertical integration of the framework data proved to be very difficult due to the varying scales, data sources, and intended uses of the input data. In general, while data shapes and positions are similar between data layers (i.e. government units vs. reservations), it was beyond the scope of this project to ensure that the boundaries in each data layer were exactly coincident. To ensure complete vertical integration would require extensive data processing and analysis and is not likely to happen any time in the near future.

<u>Recommendations for Framework Development:</u>
This demonstration project illustrated a number of challenges which should be kept in mind when future frameworks are developed across the binational border:

a) *Understand the copyright restrictions on source data from each jurisdiction included:* The differences in copyright over Canadian and American data is the single greatest obstacle to creating transboundary datasets. For instance, any dataset which includes provincial data from British Columbia, cannot be redistributed without the consent of provincial authorities. Even if datasets are altered or significantly enhanced during the integration process, B.C. licensing forbids redistribution. Unless the partners can assert copyright over the interated product, this limits the utility of the integrated dataset.

b) *Always start with source data:* Project partners found legacy errors in interpolated files, such as DEMs, that they then had to address. An error, such as an incorrect datum transformation, can endure through future versions of the file. Without access to source survey data, it is difficult to repair this error in an integrated dataset.

c) *Focus on a common classification first, before integrating data:* By building the data dictionary first, a GIS shop is forced to evaluate inconsistencies across classification systems in different jurisdictions. A theme is much easier to integrate if one starts with the database, then considers the topology, rather than the other way around. Include legacy classifications in the database so that future users can query out the original data if at all possible (this results in a larger database and numerous empty fields but produces a more flexible result).

d) *Do not assume source data is correct:* Our review found that the Digital Elevation Models did not correspond with Geodetic Survey information even though both came from government sources. Clear errors exist in the former which would affect subsequent analysis with the data. Although we couldn't repair these errors, we have carefully documented them in attached metadata.

e) *Keep intended uses in mind in making decisions about how data should be integrated:* Even though a framework dataset should support as many uses as possible, weighting between intended uses can help in decisions about where to allocate the most resources. Intended uses will affect choices surrounding themes, scale and attributes. For instance, a framework supporting hydrological modeling will need a high level of accuracy during interpolation of a DEM. A soil theme will be more desirable than a transportation theme. A framework supporting carnivore habitat analysis will have little use for a soil theme but place high value on transportation (human use) and vegetation data. Although an accurate DEM is desirable, resolution isn't as critical as with hydrological applications.

*f) Try to integrate data of consistent scales and currency:* Differing data densities create problems with both analyses and cartographic output. This is particularly true of both transportation and hydrological data. Usually densities occur as a factor of scale. Current Canadian provincial data is often of finer scale than current U.S. state/federal data (i.e. 1:20 000 vs 1:24 000). As a result, Canadian data can appear more feature rich than American counterparts resulting in difficulties with integration. Furthermore, finer scale U.S. data was often less current than coarser scale data. Older data is often less accurate (roads and trails may be added/reclaimed, streams may shift as a result of floods). Choices reflected tradeoffs between consistent scales and currency. Therefore, it is important to be cognizant of the tradeoffs (and implications of these) when balancing between scale and currency in source data.

*g) Consider a strategy for updating datasets as improved data becomes available:* Generally, we see a trend towards higher resolution and more accurate data as errors are addressed over time. Improvements are made through local projects as well as focused government data reviews. These improvements should be captured and made available to the framework community either through various versions offered through the clearinghouse and/or through a periodic update of the 'master' file. In either case, tracking changes through the lineage portion of affiliated metadata is essential.

h) *Consider standardized interface for searching for data spatially on web:* Present Z39.50 compatible servers, such as ISITE lack an intuitive reference tool for searching data spatially across national borders. Certain repositories, e.g. CEONET, GEODE, have developed their own interface for searching data on their repository. A common look and feel interface which is capable of searching framework data across repositories and across borders would particularly help in knitting together GeoConnections and FGDC pilot nodes.

<u>Assessment of Project Success</u>

The Y2Y Framework Data Demonstration project had both successes and failures:

a) The **assembly of a diverse project team** with diverse interests ensured that a variety of strengths were brought to bear on project tasks and that the results met a variety of needs. Conversely, a large project team meant that delays related to one or two contractors could affect the team as a whole. The Miistakis Institute and Mountain West GIS needed to spend more time than expected ensuring that datasets and metadata compiled by a variety of shops were compiled to the same standards.

b) The **layers were compiled, integrated and evaluated as described** in our original proposal. The framework datasets are the best transboundary datasets available in the region and present considerable capacity for transboundary wildlife research in the region. Conversely we detected errors during the course of the project in source data that we are using. On one hand this is positive for use of these datasets in high profile wildlife research would have continued oblivious to errors. On the other hand, we now have complex errors to fix, particularly in the Digital Elevation Model (see DEM evaluation report in the Project Results and Reports compendium). In short, the project improved the transboundary datasets through its execution but the data is still far from perfect and partners are more cognizant of that than ever before.

c) The **datasets are available online** for widespread partner use. The project developed not only the integrated datasets but the FGDC compatible metadata to support distributed use. In the process of compiling metadata, partners were fully trained by EMAN and NBII staff, thereby raising our collective capacity to catalogue other shared datasets. The datasets have been tiled and are available through a customized web interface to allow for download by authorized partners.

A FGDC compliant ISITE server has been established alongside NABIN Species Analyst at the University of Calgary's Miistakis Institute. As of June, 2001 Miistakis is still awaiting a legal opinion on the copyright implications of registering the ISITE server and therefore allowing widespread access to the datasets. **Current Canadian copyright law and the related policies of Canadian governments surrounding data distribution is potentially a major impediment to the ultimate success of this project** – the unmitigated availability of transboundary datasets to enable collaborative ecosystem-based research and management across the 49th parallel.

In summary, this project was a mitigated success. Partners met project objectives during an extended timeline. The datasets will be used for transboundary wildlife analyses by partner institutions.

<u>Plans for follow-on activities including outreach:</u>
At a 'next steps' meeting in Kalispell, Montana partners agreed to continue work on extending the datasets both spatially and in content. Although the FGDC/GeoConnections Framework project is completed, the pilot effort laid the groundwork for integrating datasets in adjacent areas. For instance, an integrated transportation/human use dataset is presently being compiled for the Central Canadian Rockies Ecosystem (Banff/Yoho/Kootenay region) directly to the north using our framework standards. Further work between the Crown of the Continent and the Greater Yellowstone Ecosystem is expected in the near future using these standards. Other partners throughout the Yellowstone to Yukon Region are expected to learn from our experiences and use these standards when integrating their own datasets. Negotiations are underway to add a fine scale vegetation layer to the ten layers presently compiled. Work continues to ensure the

datasets are given the widest possible distribution in association with complementary biodiversity data through the Miistakis Institute and the Northern Rocky Mountain Science Centre.

Partners remain in touch via email and we expect many of them will reconvene at future workshops regarding application of the data.

User-demand requirements for framework data in a national level spatial data infrastructure:

The Yellowstone to Yukon framework project immediately attracted a large number of partners because the need for these datasets was so obvious. In fact, on the Canadian side, the results of this project have helped catalyze a new research partnership through the University of Calgary, Parks Canada, EMAN and the Miistakis Institute. The availability of this data will enable transboundary research and decision support which was simply not possible before.

In our experience, properly catalogued framework data lays the groundwork for a host of different ecological analyses. Without it, ecological research and decision-making occurs using disparate datasets of varying quality. Redundant efforts, tracking down pieces of the same data elevate the costs of both academic research and commercial analyses in support of government decisionmaking. In fact, when such analyses are undertaken to meet regulatory requirements (such as environmental impact analyses incorporating wildlife habitat suitability modelling), delays in approval resulting from a lack of information can run into hundreds of thousands of dollars worth of lost revenue by a project proponent. Conversely, errors made in such assessments from approvals based upon shoddy data can result in even more expensive reclamation costs.

On behalf of our project participants, we believe that government investments in framework data *which is then made widely and freely available to both private industry and the public alike is a necessary and justifiable subsidy.* The economic, social and environmental costs of NOT providing this data, far outweigh lost recovery revenues. In the information age, this is a necessary and fundamental investment in the economic and environmental security of both Canadians and Americans. We expect that use of these framework datasets in the immediate future will bear this out.

Conclusions:

The resulting framework data layers are a great improvement on the datasets that previously existed.  The systematic compilation and review of each layer provided critical information on the quality of the data and what the user can reasonably use the data for. Additionally, the exercise identified significant shortcomings in the spatial accuracy, attribute quality, and physical extents of the data that were previously unknown or undervalued. This provides an important stepping-stone for developing strategies for improving the data down the line. In the future the project team would like to maintain and improve the quality of the existing data through a series of maintenance updates. These would include both replacing obsolete or incorrect data with more accurate information, as well as expanding the coverage to include the entire Yellowstone to Yukon geographic region. In addition, we expect the partnership developed by this team will continue to work towards applying and extending these datasets through a variety of projects in the immediate future. The Yellowstone to Yukon Framework Data Demonstration Project was a fruitful exercise with lasting results.

Acknowledgements:

The project participants would like to thank the many public and private organizations in both the U.S. and Canada who came together to make this project a success. The U.S. Federal Geographic Data Committee (FGDC) and Canada's federal GeoConnections (Natural Resources Canada) provided the primary funding for the project and advice and guidance through Milo Robinson, Eric Goods and Marc Lemaire. The countless members of the Yellowstone to Yukon Conservation Initiative provided invaluable inspiration to take on the project, as well as important financial support. Additional contributions of time, expertise, and equipment by the participating organizations were critical to the successful completion of this project.