FISH PASSAGE CENTER

2501 SW First Avenue, Suite 230, Portland, OR 97201-4752

Phone: (503) 230-4099 Fax: (503) 230-7559

http://www.fpc.org/

e-mail us at fpcstaff@fpc.org

MEMORANDUM

TO: Brian Allee
Anadromous Fish Managers

FROM: Michele DeHart
Henry Franzoni

DATE: October 13, 1999

RE: Regional Database Technology – the future

Recent Independent Scientific Review Panel (ISRP) comments on the Fish Passage Center 2000 proposal raised several issues regarding regional databases, including comments about StreamNet, DART and PITAGIS. The Fish Passage Advisory Committee (FPAC), and PSMFC both responded to the ISRP comments. The PSMFC response[1] to the ISRP comments discussed The Fish Passage Center and it's present and future relationship to StreamNet, the Coded Wire Tag database (CWT), and the Passive Integrated Transponder Tag System (PITAGIS). Subsequently, precipitated by the ISRP comments and responses, Brian Allee requested that we review the present status of regional data management activities with particular attention to present technology, efficiency and a concept for future regional databases. Following is our response for CBFWA consideration.

Our response was developed through a series of steps. First we reviewed the available documents and work statements for StreamNet to develop an understanding of the present status of StreamNet. Second, we reviewed the most current technology and national templates for large regional databases. Third, we developed a future concept for a regional database that meets the management needs of the agencies and tribes in a cost efficient way while avoiding duplication and redundancy.

---

[1] "*Response to the ISRP Review*" a memorandum from Randy Fisher, Executive Director of PSMFC, to the Northwest Power Planning Council, the Independent Scientific Review Panel, and the Columbia Basin Fish and Wildlife Authority, dated August 13th, 1999.

Brief History of the development of Streamnet – What does Streamnet do now?

The Historical StreamNet

StreamNet's predecessor project was the Coordinated Information System (CIS), established pursuant to a directive in the 1987 FWP (NWPPC, 1987, 206.d.2.C). In 1996, CIS was merged with the BPA-sponsored Northwest Environmental Database project and given the name "StreamNet." This action was taken in part as a cost savings measure, but it also involved recognition of the region's evolving data needs. A steering committee was formed and charged with project oversight, and a long-term data plan was prepared. In FY 1997, StreamNet enhanced the 1:100,000 river reach system, incorporating new datasets on water quality, Council Protected Areas, the Pacific Northwest Rivers Study final ratings, and a 1:100,000-scale fish distribution data layer. StreamNet data exchange formats were also updated, with new data types added such as barriers. In FY 1998 region wide anadromous fish distribution and use type data were made available and integrated with an enhanced version of the 1:100,000-scale hydrography.

What does StreamNet do today?
The objectives StreamNet has today can be briefly summarized in the list below

Objective 1. Fish and Wildlife Program Data Development
Objective 2. Data Management
Objective 3. Library Services
Objective 4. Fish and Wildlife Program Data Services
Objective 5. Project Administration
----------Additional objectives found in *FY 1999 EPA Statement of Work*----------
Objective 6. EPA Environmental Data Development
Objective 7. NMFS Anadromous Fish Trend Data Development
Objective 8. NMFS Habitat Restoration Projects

The first five objectives are from StreamNet's *Fiscal Year 2000 Fish and Wildlife Program (FWP) proposal*. The last three are from StreamNet's *FY 1999 EPA Statement of Work.*

From a data management perspective, the data centerpiece that many objectives revolve around is the Pacific Northwest River Reach File. A lot of the data development and data management tasks StreamNet performs in achieving these objectives involves integrating data with the PNW River Reach File, most often using the *StreamNet Data Exchange Formats*. StreamNet has defined these data exchange formats, and continues to revise them. StreamNet's *FY 2000 Fish and Wildlife Program (FWP) proposal* also states that StreamNet will *"maintain the 1:100,000-scale hydrography and a regionally consistent watershed, subbasin, and physiographic province referencing system, including*

providing means for integrating FWP-generated data with this system."[2] The PNW River Reach File is often referred to as the 1:100,000-scale hydrography, the 100K Hydrography, and the PNW River Reach System.

From a data management perspective, a functional view of Streamnet looks a bit different.

- First of all StreamNet is a traditional library, with a collection of regional publications relevant to the FWP. It makes a selected portion of these electronically available on the Web. The rest of the collection is housed in a physical library. Apparently StreamNet maintains an electronic library catalog using the off the shelf INMAGIC library catalog program.
- Second, StreamNet maintains a FWP project database for BPA, NWPPC, and CBFWA, and PSMFC is the project administrator for many FWP projects. The FWP project database is part of the StreamNet library available on-line. There is a certain synergy between StreamNet managing the FWP project database and PSMFC being the project administrator for many FWP projects.
- Third, StreamNet maintains the PNW River Reach File, the regionally consistent 1:100,000-scale hydrography mentioned above, and integrates this with FWP project generated data to make various regional data products. Regional operating and natural resource agencies such as BPA, COE, BOR, CRITFC, CBFWA, NWPPC, USFWS, NMFS, EPA, USGS, BLM, and USFS are the primary audience and beneficiaries of these region-wide regionally consistent data products.
- Fourth, StreamNet sees itself as serving "*a data coordination function within the Fish and Wildlife Program…*," and in order to achieve that objective, "*establishing and implementing a more formal long-term working arrangement with other projects is a primary goal.*"[3]

StreamNet's primary means for serving a data coordination function within the FWP can be summarized in this simplistic fashion: *define regional data exchange standards and collect an "exchanged" copy of the data in a central data repository. Fund participating agencies and tribes to prepare data for submission to the central repository in the StreamNet Exchange formats*. If possible, persuade the participating agencies to collect the data in a StreamNet compatible format in the first place. The StreamNet web site then distributes regional data products while avoiding duplication of effort with other FWP regional data projects, although some duplication occurs. The StreamNet web site has hyperlinks to other regional data projects such as FPC, PTAGIS, and CWT. The *StreamNet Data Exchange Formats* and the StreamNet data holdings are maintained and updated on an ongoing basis. Regional data products, especially regional GIS data products dependent on integration with the PNW River Reach File, are developed and made available on the StreamNet web site or through other means. The cost of making the regional data products and distributing them is covered under StreamNet's contract with the FWP, or under outside contracts with other entities.

---

[2] Objective 2, task f, in StreamNet's FY 2000 FWP proposal found at web address ftp://www.streamnet.org/pub/streamnet/projman_files/88108041.doc

[3] StreamNet's *FY 2000 FWP proposal* found at web address ftp://www.streamnet.org/pub/streamnet/projman_files/88108041.doc

StreamNet/PSMFC is also currently involved in the creation of the National Hydrography Dataset (NHD); a national 1:100,000-scale hydrography being assembled by USGS and EPA. StreamNet might be involved in the long-term maintenance and management of the PNW portion of the NHD. Apparently, no decision has been made yet by USGS regarding its long term data management approach[4], but the fully distributed model is the default and is in effect today. Currently, StreamNet is not only maintaining and enhancing the PNW Reach File, it's at some stage of converting and integrating it into the new National Hydrographic Dataset. StreamNet's *PNW River Reach documentation,* (revised Jan. 1999), states that, "*PSMFC is also the lead contractor responsible for supervising the integration and conversion of the PNW Reach Files into the new National Hydrography Dataset."*[5] Streamnet's *1997 Strategy and Procedure for Long-term Maintenance of the 1:100K River Reach System,* states that, *"StreamNet and StreamNet-affiliated agencies believe that there are significant advantages to converting the PNW Reach File into the NHD format and have been active participants in the National Hydrographic Dataset (NHD) project. In cooperation with USGS and the states, StreamNet is conducting the "visual pass" portion of the NHD development for the four Pacific Northwest states." … "However, neither the PNW Reach File, the NHD, or, as is likely to be the case, a merged product, will meet the region's demanding need without near-term enhancement and long-term maintenance*."[6]

StreamNet/PSMFC is also currently involved with the *Inter-organizational Resource Information Coordinating Council* (IRICC), which is charged with developing a seamless, current, and accessible information network to support ecosystem management in support of the Northwest Forest Plan, the Endangered Species Act, National Environmental Policy Act, the Clean Water Act, and Federally Reserved Rights. It is remarkable the IRICC web site has a link to a 100K regional hydrography page, which then has no data. Presently, neither the NHD web site nor the IRICC web site distributes actual 100K hydrographic data for the four pacific northwest states.

StreamNet's Current Management Structure

StreamNet's management structure is ever so-slightly-vague on one point. It's hard to determine who has the final say in regards to non-FWP projects, PSMFC or the Steering Committee. StreamNet is administrated by PSMFC, and its current management structure is synopsized in StreamNet's *FY 1999 EPA Statement of Work*:[7]

> "StreamNet is managed by the Pacific States Marine Fisheries Commission (PSMFC), under contract to BPA, NMFS, and EPA. The Northwest Power Planning Council (Council)

[4] USGS paper regarding a qualitative comparison of two data management models for the NBII found at http://www.nbii.gov/about/architecture/architur.htm

[5] Streamnet PNW River Reach File Documentation, updated January 1999, found at web address http://www.streamnet.org/pnwr/PNWNAR.html

[6] *StreamNet Strategy and Procedure for Long-term Maintenance of the 1:100K River Reach System within the Pacific Northwest StreamNet White Paper - September 1997 - final draft*, found at web address http://www.streamnet.org/pnwr/pnw_maint.html

[7] *StreamNet's Fiscal Year 1999 Environment Protection Agency Statement of Work*, found at http://www.streamnet.org/99EPA.htm

provides policy oversight for the Fish and Wildlife Program portion funded by BPA. Other regional agencies with a contractual commitment include:

- California Department of Fish and Game (CDFG)
- Columbia River Inter-Tribal Fish Commission (CRITFC)
- Idaho Department of Fish and Game (IDFG)
- Montana Fish, Wildlife and Parks (MFWP)
- Oregon Department of Fish and Wildlife (ODFW)
- Shoshone-Bannock Tribes (SBT)
- U.S. Fish and Wildlife Service (USFWS)
- Washington Department of Fish and Wildlife (WDFW)

A StreamNet Steering Committee consisting of representatives from BPA, CRITFC, IDFG, MFWP, ODFW, PSMFC, SBT, USFS, and WDFW oversees the project.

StreamNet's FY 2000 FWP proposal mentions three specific non-FWP contracts, two with NMFS and one with EPA, which will provide 9% of StreamNet's FY 2000 budget. It also mentions unspecified contracts with participating agencies, which will provide another 6% of StreamNet's FY 2000 budget.

The Present State of Technology and National Templates for Data Management

National Data Management Templates

Whether Al Gore invented the Internet or not, President Bill Clinton has changed all the rules with regard to federal agencies and information technology. A national template for spatial data management was laid out in Executive Order 12906[8], "*Coordinating Geographic Data Acquisition and Access: the National Spatial Data Infrastructure*", published in April of 1994. This order announced the beginning of a national effort to coordinate spatial data acquisition and access, focusing on the GIS user community. A *National Information Infrastructure* (NII) was created along with an oversight committee, the *Federal Geospatial Data Committee*[9] (FGDC). The NII presently consists of two main components, the *National Spatial Data Infrastructure*[10] (NSDI) and the *National Biological Information Infrastructure*[11] (NBII). The NBII is a national effort to coordinate biological data acquisition and access. The two primary search sites for these efforts are the *National Spatial Data Clearinghouse*[12] (NSDC), and the *National Biological Information Infrastructure Metadata Clearinghouse*[13] (NBIIMC). The hydrological component of the NII is the new *National Hydrographic Dataset*[14] (NHD), created and currently maintained by USGS and EPA. This is going to be a nationwide 1:100,000-scale hydrography dataset when it is completed.

[8] *Executive Order 12906* was published in the April 13, 1994 edition of the Federal Register, Volume 59, Number 71, pp. 17671-17674
[9] The Federal Geospatial Data Committee is found at http://www.fgdc.gov
[10] The National Spatial Data Infrastructure is found at http://www.fgdc.gov/nsdi/nsdi.html
[11] The National Biological Information Infrastructure web site is http://www.nbii.gov
[12] The *National Spatial Data Clearinghouse* is found at Web address http://www.fgdc.gov/clearinghouse/index.html
[13] The *National Biological Information Infrastructure Metadata Clearinghouse* is found at Web address http://www.emtc.usgs.gov/http_data/meta_isite/nbiigateway.html
[14] The *National Hydrographic Dataset* Web address is found at Web address http://nhd.usgs.gov

The National Hydrography Dataset will replace or combine with the PNW Reach File and become the main regional 1:100K hydrography. StreamNet's Data Exchange Formats will be affected by this regional change. FWP project generated data has been integrated with the PNW Reach File using Streamnet's Data Exchange Formats to generate many of the GIS based data products produced by StreamNet. Back in Sept. 1997[15] StreamNet wrote that the "*StreamNet Steering Committee,… has officially adopted the PNW Reach File (1:100,000-scale) as the standard building block for regional data development*." The NHD does not currently endorse the PNW Reach File as the standard building block for regional data development, and does not currently endorse the StreamNet data exchange formats for use with the NHD. Two data exchange formats are currently defined for the NHD, *NHDinARC* and the *Spatial Data Transfer Standard* (SDTS). NHD metadata is formatted and exchanged using the Federal Geospatial Data Committee's (FGDC) *Content Standards for Digital Geospatial Metadata*[16]. NHDinARC is a variation of the ARC/INFO data export, which allows transfer of NHD data into off-the-shelf GIS software packages such as ARC/INFO and ARCView. The Spatial Data Transfer Standard (SDTS) is a national data exchange standard for moving spatial data between software systems. In the future, integrating FWP project generated data with the NHD rather than the PNW Reach File is likely to become the dominant method for developing regional GIS based 1:100,000-scale hydrological data products. If the PNW Reach File and the NHD eventually merge into one product, FGDC compliant metadata, NHDinARC and the SDTS will certainly become the dominant standards for exchanging data with the NHD.

The NHD and the PNW River Reach File are not the only hydrography data sets in the region that FWP project generated data can be integrated with. The states of Oregon and Washington are presently developing statewide 1:24,000-scale hydrological datasets, (Oregon's is available), which are specified to the 5th hydrological level, the watershed level, as opposed to the PNW Reach File and the NHD, which are specified to the 4th hydrological level, the sub-basin level. Presently Streamnet's 100K hydrography has more detail than Oregon's or Washington's 24K hydrography in at least one way, StreamNet's 100K has ARC View GIS shapefile format data files for streams and stream banks. Western Oregon's present 24K doesn't have ARC View GIS shapefile format data files for streams and stream banks. However, Oregon's does have border hydrological units and can be found at the State Service Center for Geographic Information Systems[17]. The spatial data library on this site exchanges GIS data with users by using the ARC/INFO single precision export format and the ARC View GIS shapefile format, and it uses the national FGDC Content Standards for Digital Geospatial Metadata.

On-line searchable "metadata", which is data that describes the content and quality of data, is a key component of Executive Order 12906. "Metadata clearinghouses" contain standardized metadata-based descriptions of biological and spatial data sets and

[15] *StreamNet Strategy and Procedure for Long-term Maintenance of the 1:100K River Reach System within the Pacific Northwest StreamNet White Paper - September 1997 - final draft*, found at web address http://www.streamnet.org/pnwr/pnw_maint.html

[16] The Federal Geographic Data Committee (FGDC) Web site is found at http://www.fgdc.gov/

[17] (Oregon) State Service Center for Geographic Information Systems is found at http://www.sscgis.state.or.us/

information products. Spatial metadata follows the *Content Standards for Digital Geospatial Metadata*[18] established by the Federal Geospatial Data Committee (FGDC). Biological metadata descriptions used in metadata clearinghouses follow the *Biological Profile*[19] of the National Biological Information Infrastructure, which is overseen by the USGS BRD. Metadata is explained on the National Hydrography Dataset Web site as:

> 'Metadata, or "data about data," are data that describe the content, quality, condition, and other characteristics of data. Metadata answer questions such as "how current are these data?"; "how accurate are they?"; "are there any restrictions on their use?"; "what is their coordinate system?"; and many others. Metadata help organizations manage data, advertise and share data, and make informed use of data. Metadata for the National Hydrography Dataset use data elements from the *Content Standards for Digital Geospatial Metadata*[20]" (FGDC, 1994). The standard allows the identity, quality, spatial data organization and reference, entity and attribute definitions, distribution sources and forms, and metadata of the data to be documented. The metadata are provided as text files. For the *National Hydrography Dataset*, a general set of metadata accompanies each set of data. This "NHD" metadata provides general information that applies to all data.'[21]

In the long run, if one is going to do any analyses with any particular data set, it can save a lot of time and effort if one first examines the content, quality, condition, media, and limitations of a particular data set before actually obtaining the data and trying to use it.

The Washington State Geospatial Clearinghouse[22] and the Clearinghouse for the Olympic Peninsula[23] are two more good examples of regional data clearinghouses for biological and geospatial data. They not only use FGDC compliant metadata and the SDTS for translating geospatial data, they are National Biological Information Infrastructure Clearinghouse Nodes[24]. It is a straightforward relatively simple process to become a node on the National Biological Information Infrastructure Clearinghouse. The step-by-step instructions to do so are found in Appendix B

The Inter-Organizational Resource Information Coordination Council has overseen the development of a test GIS server in SW Oregon called the *Internet Map Server for the Southwest Oregon Province*[25]. This GIS map server serves up 50 different GIS layers using the ArcView GIS shapefile format. This site is remarkable because at least eight different federal agencies contributed GIS layers, as well as PSMFC.

Another national data management template applicable to our region is the *Integrated Taxonomic Information System*.[26] The Integrated Taxonomic Information System (ITIS) is a partnership of U.S., Canadian, and Mexican agencies, other organizations, and

---

[18] The Federal Geographic Data Committee (FGDC) Web site is found at http://www.fgdc.gov/
[19] http://biology.usgs.gov/nbii/current.status.html
[20] The *Federal Geospatial Data Committee's* Web address is http://www.fgdc.gov/
[21] Metadata explained on the National Hydrographic Dataset web site http://nhd.usgs.gov
[22] Washington State Geospatial Clearinghouse is found at http://wa-node.gis.washington.edu/
[23] Clearinghouse for the Olympic Peninsula is found at http://cathedral.cfr.washington.edu/~chouse/
[24] NBII Clearinghouse Nodes are listed at http://www.umesc.er.usgs.gov/http_data/meta_isite/nodes.html
[25] Southwest Oregon Province Resource Information is found at http://208.147.49.8
[26] The Integrated Taxonomic Information System (ITIS) is found at web address http://www.itis.usda.gov/

taxonomic specialists cooperating on the development of an on-line, scientifically credible, list of biological names focusing on the biota of North America.

The Most Advanced National Data Management Template

The most advanced, most up to date, most efficient national data management template is the NOAA Server Project,[27] which provides access to distributed NOAA data and information. NOAA Server is mainly an FGDC compliant metadata clearinghouse, which contains standardized metadata-based descriptions of biological and spatial data sets and information products, but there are a few twists. First, the metadata on NOAA server contains contact information for obtaining the data, or it contains a Web site URL (Uniform Resource Locator) where the data can be obtained on-line. Existing web user interfaces to existing data systems are therefore incorporated into the NOAA server system without modification. Second, NOAA Server is made of many metadata clearinghouses, all searchable from a single location. It's based mostly on existing technology, using standard search engines such as Z39.50 and WAIS to search the distributed databases from a single point on the web using a standard web browser. Third, it uses a new technology named CORBA, which stands for Common Object Request Brokered Architecture, along with the Java programming language to create a single web interface to search distributed metadata. This is all only phase one of the NOAA Server Project, which is completed. Phase two, which has completed a proof of concept stage, is to use the Java language in combination with CORBA to build a web user interface that accesses and combines multiple distributed databases from one access point into one data product. I.E. one can draw a graph or a map or download a spreadsheet using data from multiple distributed sources simultaneously.

The NOAA Server Project solved many of the same data management problems faced in our region. First of all, each of our region's biological and environmental observations and derived information products have been made and recorded in response to specific mission objectives of individual agencies and tribes, so information management systems have developed over time independently of one another. For the most part, these information systems utilize a broad range of data formats, data media, and overall different approaches to managing and preserving the regional biological and environmental record. The NOAA Server Project made use of existing data systems and user interfaces, and the investment in them, without additional expense. By using existing off the shelf technology whenever possible, by gathering searchable metadata instead of complete datasets, by using existing national metadata standards, and by using existing data system interfaces, NOAA server minimized the cost of coordinating and distributing information. NOAA server is geographically distributed allowing data providers to control the content of their information offerings and to update and expand their offerings as they see appropriate. Data providers are kept accountable for their offerings. NOAA Server allowed more funds to be expended on the quality of the data, at the source of the data, and less funds to be expended on translating, converting, and copying the data. It also allowed more funds to be spent on documenting the quality, content, condition, and limitations of the data, which made the data more scientifically useful. NOAA Server allowed funds to be spent on the original source and on the original organization of the data, not on the job of translating it.

[27] NOAA Server is found at http://www.epic.noaa.gov/cgi-bin/NOAAServer

There are a few inherent problems with copying data sets from one format to another, one of which is that errors do occur, and sometimes information, or context is lost. There is extra work and extra probability of error. Try to make something foolproof and a fool will prove you wrong. Translation protocols always involve the additional time and expense of validation processes. NOAA Server did not expend funds copying data sets to a central repository, instead centralized access was provided to geographically distributed data providers.

An additional benefit of NOAA Server is that searchable metadata gave management a tool to quickly and cheaply identify data holdings, duplication of effort, or duplication of data, across a vast agency.

Finally, NOAA Server lived up to it's responsibilities under another President Bill Clinton executive order, Executive Order 13011 "Federal Information Technology", from July of 1996, which placed a number of new responsibilities on federal agencies in the region. NOAA Server also lived up to it's responsibilities under an OMB memorandum of October 25, 1996, which added eight criterion for the funding of federal information technology investments. Rather than discuss the details of those responsibilities now the reader is referred to the synopsis of these two documents found in Appendix A.

Data Management Needs of Agencies and Tribes

The present and future data needs of the agencies and tribes are varied and broad in scope. They also quickly and often change to meet management needs. As examples, the agencies and tribes need real time smolt and hydrological data during the fish migration season in order to make recommendations regarding operation of the hydrosystem. This is the main purpose of the Smolt Monitoring Program and FPC data system managed by the Fish Passage Center. The tribes and agencies need smolt to adult survival information for hatcheries, ESU's, and sub-basins. New initiatives continually arise which require specific data. As an example, the anadromous production review will require future data products, as will sub-basin management plans, ESA monitoring, ecosystem monitoring, hatchery evaluations, ESU evaluations, resident fish evaluations, hatchery genetic management plans, and regional, ESU, and sub-basin recovery plans. The recent ISRP report raised issues regarding duplication and coordination, addressing the efficient expenditure of limited fish and wildlife program funds.

Future Data Management Template for StreamNet

Our review of the historic and present StreamNet development and present structure and review of the present national database technology templates led us to conclude that this is an appropriate and timely point to consider a re-design of StreamNet and the regional database concept. We developed the following conceptual design for regional database management.

Our concept is based upon the following fundamental principals:

1. Databases must be designed to meet the dynamic and changing management needs of agencies and tribes.
2. The data must be of the highest quality and accuracy.

3. Data is most accurate maintained close to its source without translation, re-formatting and re-structuring. Data should be entered once.
4. Data should be available globally.
5. Duplication and redundancy should be eliminated.
6. Resources are best expended to strengthen the data at its origin and acquisition, to improve the quality, consistency and storage of the data at its source.
7. The highest efficiency results from using off-the-shelf software and avoiding custom development.
8. Data is compiled on an as needed for management basis and should be of an immediate benefit. Compilation of data for the sake of compilation should be avoided.

Our concept is based upon our following conclusions from our review of historical activities and the present and future state of data management technology.

1. Regional data standards are obsolete. The national standard is moving away from this concept.
2. The National Biological Information Infrastructure (NBII) is the most efficient way to coordinate and distribute biological and geo-spatial information in our region.
3. The NOAA server project improved upon the NBII by incorporating web links to existing data systems into the metadata descriptions.
4. The National template for regional databases has moved away from copying data into central databases from distributed sources, and moved toward the use of searchable, standardized metadata.

Concept for A Future Regional Data System

Our proposed concept follows the NOAA server template. The proposed concept is comprised of two major components. In the first component, a central search engine is established and is the only central regional data activity. Data is not copied from individual sources or individual databases to a central repository. Instead funding investment is concentrated in the individual databases at their source where data is initially compiled. As an example, funding investment would be concentrated on state and Tribal databases. Translating and copying data is not necessary and is eliminated. Individual databases then follow the NOAA server template and provide a metadata description, with web links, to the central search engine. The present Pacific Northwest Reach data is finally included in the National Hydrography database.

In the second component, data may be compiled and combined for specific management purposes or specific analysis by those entities conducting the specific analysis. This is most efficient and then relates directly to specific management actions or purposes. These databases can be made available to others through web sites.

- First of all, the StreamNet data exchange formats will soon be at the end of their effective lives, without more revision. The GIS community at large has adapted defacto the ARC INFO export format, the Spatial Data Transfer Standard (SDTS) of the National Spatial Data Infrastructure, and the ARC View shapefile format,

and the FGDC metadata content standards. The National Hydrography Dataset has adapted the spatial data exchange standards SDTS and NHDinARC, and uses FGDC compliant metadata. The National Biological Information Infrastructure and the National Spatial Data Infrastructure have established standards for the transfer of data and metadata.

- StreamNet should get out of the business of defining regional data exchange standards and use national standards defined under the National Information Infrastructure instead. In the case of GIS data, StreamNet should use the SDTS and defacto standards in use by the GIS community today, most of which are based on ARC INFO or ARC View formats.
- StreamNet's PNW River Reach File is at some stage of being integrated with the National Hydrography Dataset. When this project is completed, maintaining and enhancing the PNW River Reach File will be a duplication of effort. The duplication of effort should end as soon as possible. Sometime in the future, FWP project generated data should be integrated with the NHD instead of the PNW River Reach File to take advantage of "*an enhanced data model"* with "*better attributing*"[28] as StreamNet itself explained. Use of national and defacto data exchange standards will also facilitate integration of FWP project data with larger variety of regional environmental datasets, especially for GIS purposes. Environmental data from Canadian, oceanographic, and satellite sources will be more accessible. For these reasons, StreamNet should focus on expediting the integration of the PNW River Reach file with the NHD, and expedite a regional move toward using the NHD instead of the PNW River Reach File. StreamNet should not invest any more effort into developing a 1:24K-hydrography dataset. 1:24K hydrography efforts should be funded and driven by tribal, state, regional, and federal interests as needed.
- The NHD is managed in all other regions besides the four states of the Pacific Northwest by the USGS and EPA. To be consistent nationally, this should be implemented here in the Northwest. The NHD's *flow validation* process is centrally performed by the USGS. The USGS will continue to perform this validation subsequent to the initial release. The NHD has been designed to accommodate the incorporation of user-supplied updates as well. Users submitting updates must also provide metadata documenting their changes to the data. The status and plans for accepting user-supplied updates is discussed under NHD Data Maintenance.
- With the redesign of StreamNet to be more consistent with the NBII and NOAA Server template, the need for a high level of effort in a central data source is not necessary.

The central role is redefined as:

- Phase 1 of the new StreamNet is for StreamNet to become a regional node on the National Biological Information Infrastructure Metadata Clearinghouse. All the software to do so is free, and available for download through the NBII. Step by step instructions are found in Appendix B.
- By becoming a node on the NBII, regional users could search our regional metadata clearinghouse, or search all the metadata clearinghouse nodes on the

[28] StreamNet Strategy and Procedure for Long-term Maintenance of the 1:100K River Reach System within the Pacific Northwest StreamNet White Paper - September 1997 - final draft, found at web address http://www.streamnet.org/pnwr/pnw_maint.html

NBII. Search one metadata clearinghouse, search them all, the existing free NBII software allows this function when one builds a participating node.

- Being a node on the NBII should become StreamNet's main data management function. There is **no development cost for becoming a node on the NBII**, the software is free, all that is required is a Windows NT Server or UNIX computer with a permanent connection to the Internet and a person who can install and configure Operating system and web server software. It would take about a week to install and configure everything.
- There should initially be three levels of voluntary participation. Level one participants will submit FGDC compliant metadata containing a phone number, fax number, or email address of a contact person who can supply the requested data. Level two participants will submit metadata with a URL for an agency or tribal web site where instructions for obtaining the data can be found. Level three participants will submit metadata with a URL where the data can be immediately obtained on-line from an existing data system. Other regional data centers such as FPC, CWT, PTAGIS, will begin as level three participants, leveraging existing on-line data systems at those data centers.
- Many agencies and tribes maintain their own web sites and have incorporated web sites into their routine operations. This conceptual design capitalizes on all the existing tribal and agency web sites, allowing them to be used and coordinated without change.
- StreamNet's metadata clearinghouse function will coordinate the databases FPC, CWT, PTAGIS, and shall display the content, quality, condition, and limitations of each data set in advance, as described by the individual database managers.
- Funds should be expended at the data source and original data compilation points for most efficient use of funds. Funds should go to enhance the quality of the data and not to translating the data. Funds previously expended to provide copies of data sets to StreamNet should be provided directly the agencies and Tribes to provide FGDC compliant metadata.
- Data providers need to control the content of their information offerings and to update and expand their offerings as they see appropriate. Those who are responsible for the data should provide the data. Tribal sovereignty needs to be respected, and in the information age, tribal sovereignty includes the right of determining which data gets shared beyond the tribe. Funds for data collection should be expended at the source of data collection, the agencies and tribes.
- Funds spent at the data source on the creation of metadata enhances the quality and scientific usability of the data.
- Metadata that is compliant with FGDC standards and data compliant with national or defacto data transfer standards will only have to be created one time. Of course, one can always update or expand one's data offerings, but the Tribes and agencies won't have to translate their data and metadata into a myriad of formats as future needs arise. The job of making the data available regionally will only need to be done once. The metadata can be submitted to numerous search engines using the same formats.
- Metadata Clearinghouses are a much more cost effective means of providing centralized regional data coordination and access rather than duplicating and integrating regionally distributed data holdings into a central StreamNet repository.

- Eventually BPA or NPPC or CBWFA could require that projects funded by the BPA money submit metadata describing the data to be collected to a regional metadata search engine. It could further be required that the search engine be used before a project is funded to determine if the data had already been collected by another project.
- The StreamNet library can be treated as an individual tribal or agency database submitting metadata to the search engine.
- The Integrated Taxonomic Information System should be used for consistent definitions of biota.
- Phase 2 of the new concept implementation will be to follow the lead of NOAA Server, and use CORBA (Common Object Request Brokered Architecture) or DCOM (Distributed Common Object Model), to develop a web interface for combining distributed data in real-time from geographically distributed data providers. There is some development cost associated with this phase. This process should be defined carefully, data element by data element, and defined in small achievable chunks. Other regional data centers such as FPC, PTAGIS, and CWT can become phase 2 participants in the metadata clearinghouse by supplying selected distributed data via CORBA or DCOM to a centralized web interface.
- In the exceptional cases where there is clearly a regional management need to collect specific centralized data from many distributed providers, the data collected should be as narrow in scope and collected for as brief a duration as practicable. Each data element collected should solve a specific part of an overall management problem and should deliver a measurable product such as supporting a specific regional analysis. The data should be compiled for the specific analysis and by the individuals conducting the analysis. By following this philosophy with all region wide centralized data collection efforts, federal agencies in the region will live up to their responsibilities under the OMB memorandum of October 25, 1996, and Executive Order 13011.
- Funds should be expended on combining FWP project generated data only when regional management needs justify the cost. The cost of making a particular distributed data set available on a unified web interface using CORBA or DCOM will have to be offset with the risk of management error due to inadequate information.
- The use of CORBA or DCOM for data management will avoid data duplication, yet allow distributed data to be combined when the benefit justifies the cost. The expenditure of funds to maintain duplicative databases will be at an end.
- Off-the-shelf software and hardware should be used whenever possible. For example, a free GIS tool is available on the National Hydrography Dataset web site to assist in using NHD cataloging unit (CU) workspaces that are in NHDinARC format.[29] This tool is named NHDView, which is designed to run under ArcView 3.0a or 3.1 for Windows 95/98/NT.
- Compatibility with NOAA Server and the EMAN Data Set Library (Environment Canada Server)[30] will facilitate the combination of FWP project generated data with oceanographic data and Canadian data.

[29] NHDView is found on the web at http://nhd.usgs.gov/viewer.html
[30] http://www.umesc.er.usgs.gov/http_data/meta_isite/#EMAN

- StreamNet, the regional metadata clearinghouse on the Internet, should have a mirror site somewhere else on the Internet, for reliability.

Appendix A, Executive Order 13011

The White House issued an Executive Order 13011, Federal Information Technology in July 1996 to implement the provisions of the ITMRA, which places on agencies:

"clear authority and responsibility to make measurable improvements in mission performance and service delivery to the public through the strategic application of information technology.."

Under the order, executive agencies are responsible for:

- Significantly improving the management of their information systems, including the acquisition of information technology
- Refocusing IT management to support directly strategic missions, implement an investment review process that drives budget formulation and execution for information systems, and rethink and restructure functions before investing in information technology.
- Establishing clear accountability for information resources management (IRM) activities by creating agency Chief Information Officers (CIOs) with the visibility and management responsibilities necessary to advise the agency head on the design, development, and implementation of information systems.
- Cooperating in the use of information technology to improve the productivity of federal programs and to promote a coordinated, interoperable, secure, and shared government-wide infrastructure.

In addition, agency heads are tasked to *"strengthen the quality of decisions about the employment of information resources to meet mission needs through integrated analysis, planning, budgeting, and evaluation processes,"* including:

- Determining, before investing in information technology, whether the government should perform the function, if the private sector or another agency should support the function, and if the function needs to be redesigned;
- Establishing mission-based performance measures for information systems investments, aligned with agency performance plans prepared pursuant to the Government Performance and Results Act of 1993; and
- Establishing agency-wide and project-level management structures and processes responsible and accountable for managing, selecting, controlling, and evaluating investments in information systems, with authority for terminating information systems when appropriate.

The order also promotes current "best practices" strategies of structuring information systems investments "*into manageable projects as narrow in scope and brief in duration as practicable*" and of providing for multi-agency acquisitions of information technology. Another trend, "interagency cooperation, collaboration, and review", is reflected by the order's formal establishment of three interagency groups:

- The *Chief Information Officers Council* will recommend overall federal IT management policy, procedures, and standards; share information and experiences and provide advice; identify opportunities for and sponsor cooperation in using

information resources; and address the hiring, training, classification, and professional development needs of personnel in federal IRM.

- The *Government Information Technology Services Board*, to be referred to as the "Services Board," will support continued implementation of the recommendations of the National Performance Review; identify and promote the development of innovative technologies, standards, and practices; and create opportunities for cross-agency cooperation, intergovernmental approaches, and multi-agency projects.
- The *Information Technology Resources Board*, to be referred to as the "Resources Board," will provide independent assessments to assist in development, acquisition, and management of selected major information systems; review, at the request of an agency and OMB, specific information systems proposed or under development and make recommendations; and publicize lessons learned and promising practices.

OMB Memorandum

On October 25, 1996, OMB Director Franklin D. Raines issued a memorandum alerting federal agencies that future funding for major information systems investments will be contingent on meeting eight criteria specified in the memorandum. The memorandum *defines "major information system"* as *"a system that requires special management attention because of its importance to an agency mission; its high development, operating, or maintenance costs; or its significant role in the administration of agency programs, finances, property, or other resources."* The criteria will also apply to large infrastructure investments such as major purchases of personal computers or local area network improvements.

The memorandum requested the heads of the 28 major federal executive departments and agencies listed in Section 3 of Executive Order No. 13011 to provide, by November 12, 1996, a list of their agency's major information systems investments for which new or continued funding is requested for FY 1998, and an evaluation of the extent to which each investment satisfies the memorandum's investment criteria.

The eight criteria for future funding of major information systems and other large information technology infrastructure investments relate to capital planning, information architecture (i.e., the alignment of technology with mission goals), and risk management principles. *"As a general presumption,"* the memorandum states, *"OMB will recommend new or continued funding only for those major system investments that satisfy these criteria"*:

1) support core/priority mission functions that need to be performed by the Federal government;
2) are undertaken by the requesting agency because no alternative private sector or governmental source can efficiently support the function;
3) support work processes that have been simplified or otherwise redesigned to reduce costs, improve effectiveness, and make maximum use of commercial, off-the-shelf technology;
4) demonstrate a projected return on the investment that is clearly equal to or better than alternative uses of available public resources. Return may include: improved mission performance in accordance with GPRA measures; reduced cost; increased quality, speed, or flexibility; and increased customer and employee satisfaction. Return

should be adjusted for such risk factors as the project's technical complexity, the agency's management capacity, the likelihood of cost overruns, and the consequences of under- or nonperformance.

5) are consistent with Federal, agency, and bureau information architectures which: integrate agency work processes and information flows with technology to achieve the agency's strategic goals; reflect the agency's technology vision and year 2000 compliance plan; and specify standards that enable information exchange and resource sharing, while retaining flexibility in the choice of suppliers and in the design of local work processes;
6) reduce risk by: avoiding or isolating custom-designed components to minimize the potential adverse consequences on the overall project; using fully tested pilots, simulations, or prototype implementations before going to production; establishing clear measures and accountability for project progress; and securing substantial involvement and buy-in throughout the project from the program officials who will use the system;
7) are implemented in phased, successive chunks as narrow in scope and brief in duration as practicable, each of which solves a specific part of an overall mission problem and delivers a measurable net benefit independent of future chunks; and
8) employ an acquisition strategy that appropriately allocates risk between government and contractor, effectively uses competition, ties contract payments to accomplishments, and takes maximum advantage of commercial technology.

Other relevant documents for reference:

- OMB Circular A-130, "Management of Federal Information Resources" is currently under revision to incorporate more detailed guidance to the agencies on ITMRA.
- OMB, "Evaluating Information Technology Investments, A Practical Guide, Version 1.0", Office of Information and Regulatory Affairs, OMB, November 1995.
- GSA, "An Analytical Framework for Capitol Planning and Investment Control for Information Technology", U.S. Department of Commerce, April 1996.

Appendix B, How to Establish an NBII Metadata Clearinghouse Node
is found on the web at
http://www.umesc.er.usgs.gov/http_data/meta_isite/howto.html, and is attached here.

304-99.doc