Technical Applications Development Strategy And Implementation Plan To Provide Platform Independent Access To StreamNet Data Products Through The World Wide Web (WWW) Service Of The Internet

Submitted By
Pacific States Marine Fisheries Commission

7/1/96

Introduction

This document is submitted as the draft portion of the product deliverable under the StreamNet contract number 95 BI 65130 Modification 002, section 3.5.3 which reads in part, *In collaboration with BPA and participating organizations, produce a draft and final technical applications and development strategy and implementation plan to provide platform independent access to StreamNet data .... through the world wide web service of the Internet.*

Making diverse data types and formats available in an integrated, platform independent environment with a user-friendly interface and multiple output options is a complicated and multi-step process. There are few, if any, existing systems on the WWW with as comprehensive and flexible data content, query, and output options as we are proposing to ultimately achieve in this effort. Furthermore, some technical portions of designing and implementing this system are better understood at this time by project personnel than are others, and the final design and implementation of a system this complex will be an iterative, ongoing process.

Some basic knowledge of the data types we will be working with is fundamental to understanding the key components of this proposal. For the purposes of this discussion, we will define two broad categories of data: **tabular and spatial.**

> **Tabular data** refers to information types that can be stored, searched, reported, and manipulated in a traditional relational database management system. Typical data types include text or character (fixed or variable length), numeric (integer or real), logical (yes/no), and date/time. Typical tabular data outputs include reports, charts, graphs, etc. In the StreamNet scheme, tabular data types are primarily used to store *events or properties* associated with geographic (or spatial) locations or features. For example, information on redd counts or species distribution in the Grande Ronde River can be stored as tabular data, as can property ownership characteristics by county. Existing SQL tabular data models are very fast, and ad-hoc tabular data functions on the WWW have already been prototyped by StreamNet staff and have proven to be feasible.
>
> **Spatial data** on the other hand, is the data type which is used to *store the geometric location of a particular feature.* For StreamNet, this geometric location is typically a digital map or cartographic representation of a feature's location of the face of the earth. Spatial data types are more complex than tabular ones because they are multi-dimensional

and must be geographically referenced with coordinates that accurately place the geographic feature. Typical spatial data types include points, lines, and polygons. These data types can be used to store the feature location of things like streams, lakes, county boundaries, etc. Typical spatial data outputs are geo-referenced plots or maps. Until recently, these data types **could not** be stored in a traditional relational database management system and could only be manipulated in a Geographic Information System (GIS). GIS systems employ special data models which allow for the manipulation (display, plotting, overlaying, etc.) of spatial data. Existing spatial data models are not typically known for their speed, and the feasibility and implementation of ad-hoc spatial data functions on the WWW that performs well need to be carefully examined and prototyped.

Recognizing the difference between these two fundamental data types is critical in this proposal because the two require significantly different technical environments and tools. Therefore, we are proposing a phased strategy/plan to achieve the goals of this project; namely a fully functional WWW site for accessing, displaying, querying, reporting, and downloading all data relevant to the StreamNet project (both tabular and spatial). This phased approach will begin with development of systems aimed at providing access to **tabular data** and then proceed to the development of tools which will be integrated into the system and provide access to **spatial data**. Breaking the development and implementation of this system into discrete parts has several advantages including the following:

- various components of the system will be available in a more timely manner, specifically, we believe that by following this phased approach, we will be able to provide on-line, ad-hoc query and reporting capabilities to **tabular data** within a 3-4 month time span. With this capability in place, we can proceed into the more complex spatial aspects of the project with a fully functional tabular system in place. This would serve the needs of many users in the fisheries community since the majority of data available at this time is of the tabular nature.
- the successful development of this system requires that certain components of the system be in place before others are developed (ie we can't have on-line linkage between an SQL database and a web page without first developing the SQL database). By approaching the development in this pragmatic, step-wise fashion, we can carefully focus efforts at each phase and more systematically and successfully proceed through the effort.
- the phased approach, we believe, more effectively utilizes our limited StreamNet personnel resources.

PSMFC is, therefore, proposing the following four phased strategy to fulfill our contract deliverables. This strategy/plan will cover the time period from July 1st, 1996 through September 30, 1997. Throughout this strategic development PSMFC will rely on the StreamNet Steering Committee to provide input and review on products as they are prototyped, revised, and submitted to BPA for approval.

PHASE I: Provide platform independent access over the WWW to selected tabular data residing on a regional SQL database, enhance availability and selection of pre-prepared geographic products.

Timeline : July 1st - Sept 30

Objective(s):

1) Provide ad-hoc query and reporting functionality to selected tabular data residing on an SQL database through the WWW using an HTML based interface.
2) Provide access to comprehensive set of pre-prepared geographic products through WWW using an HTML based interface.

Task Sequence (key personnel):

1) Port selected StreamNet dataset to Ingress 6.4 SQL environment on PSMFC Sparc machine. Recommendation is to port all adult return data, harvest data, hatchery returns, hatchery releases, reference data and all associated look up tables necessary for fully functional prototype. Throughout FY97 port other StreamNet datasets to SQL environment as appropriate. **(Duane)**
2) Develop, test, and revise as necessary HTML based user interface to query, review, report, or download data from prototype SQL database in a variety of formats. Interface options will include tunnel down as well as boolean approaches to querying the database. **(Duane, Doug)**
3) Select a preferred alternative for user interface based on acceptance criteria (to be determined). The preferred alternative could be comprised of multiple approaches.
4) Expand current pre-prepared map catalog holdings (as GIF images) available on project WWW site to include products that are widely requested and deemed appropriate for our project. Examples might include the following:
   1) Current hatchery location coverage
   2) Current dam location coverage
   3) Updated 1:100K distribution, use type, and barrier coverages by HUC as they become available
5) Explore the feasibility of providing pre-prepared map products in ArcInfo Export format so users could load data directly into their local GIS systems. Examples might include above coverages plus:
   1) Geographic boundary coverages (state lines, HUC boundaries, subbasin boundaries, counties, etc.)
   2) 1:250K river reach coverage
   3) 1:100K river reach coverage

   (Steve, State GIS staff, BPA GIS Staff)
6) Provide interactive query interface to map holdings described above. **(Steve)**

Hardware Resources : Use existing Sparc 1000 at PSMFC

Software Resources : Use existing SQL software at PSMFC (Ingres 6.4)

Cost : No additional hardware or software costs

PHASE II: In a collaborative effort with BPA, prepare for and conduct a 2 pronged prototyping effort to test two specific models of spatial data management and WWW functionality, as well as determining feasibility of using a regional SQL host environment for attribute data used by BPA's GIS shop. The two models to be tested will be 1) the use of an SQL server in combination with ArcInfo software and 2) the use of an SQL server in combination with ESRI's Spatial Data Engine (SDE).

Timeline : July - Feb 28th

Objective(s) :

1) Prepare test environments for spatial data prototyping
2) Determine feasibility of using regional SQL environment for hosting attribute data for use by GIS client at BPA.
3) Determine feasibility of using SQL environment for hosting attribute **and** spatial data for use by multiple clients (SDE).
4) Evaluate 2 technical strategies for providing ad-hoc query and mapping functionality to prototype spatial data (and associated tabular data) through the WWW using an HTML based interface.
5) Based on results of 1-4 above, identify the preferred alternative for the project.

Task Sequence (key personnel

1) In preparation of geographic prototyping prepare test site configurations, databases, and GIS tools, specifically:
   1) Locate and prepare suitable test site for Oracle/SDE testing (Oracle is the only SQL environment available at this time which supports the SDE).
   2) Acquire and install test copy of ESRI's Spatial Database Engine (SDE) on Oracle test site.
   3) Prepare Regional SQL / ArcInfo test site(s). This will include SQL servers at both BPA and PSMFC in conjunction with the ArcInfo installation at BPA. Explore dedicated computer connection between 2 locations.
   4) Select basin for geographic prototyping and acquire and begin porting 100K reach, distribution, use type, political boundaries, and land ownership data to test environments, (Sybase/ArcInfo at BPA, Oracle/SDE at SDE test location). Recommendation is to use a large Columbia River subbasin (John Day, Grande Ronde, Yakima, etc.). **(Steve, Duane, Doug, various GIS and computer resource staff)**
2) Assess performance and feature options of test environments as related to providing remote access to attribute and/or spatial data to a variety of clients, including BPA's GIS shop.
3) Develop, test, and revise as necessary HTML based user interface to query, review, map, or download data from 2 prototype geographic test environments (SQL/ArcInfo, Oracle/SDE). Applications could include ArcInfo AML's, Java Programs, C++

applications, MapObjects, etc.

4) Based on results of task 1-3, prepare and make a recommendation for future project direction in terms of spatial data management, function, and preferred environment.

(**1-4 Steve, BPA GIS staff, Duane, Doug, State GIS Staff)**

Hardware Resources : Use existing Sparc 1000 at PSMFC.
Use existing hardware at BPA.
Hardware details for Oracle to be determined.

Software Resources : Locate existing Oracle installation or procure test license & install.
Install evaluation copy of SDE at test location.
Use existing SQL software at BPA (Sybase).
Use existing GIS software at BPA (ArcInfo).

Cost : Evaluation copy of SDE and mandatory training : $9,000-$15,000 ($5,000 applicable to purchase price in event we purchase SDE), send 2 or more to training. Costs related to Oracle to be determined (could include training).

PHASE III: Based on the experience acquired, function and performance evaluations, and recommendations developed during phases I and II of this proposal, make final recommendations for software/hardware environment for the StreamNet project. Purchase and install appropriate hardware and software based on recommendations.

Timeline : Mar 1st - May 31st

Objective(s) :

1) Determine optimal hardware and software to insure maximum functionality for StreamNet project.
2) Procure, install, and configure optimal hardware and software.

Task Sequence (key personnel

1) Produce detailed report describing outcome of Phases I and II of this proposal. Report will describe results of the various prototyping efforts, pros and cons of various approaches, and make final recommendations for software and hardware environment for the StreamNet project. **(Duane, Doug, Steve, Various GIS personnel)**
2) Subject to approval of report and recommendations, procure, install, and configure hardware and software recommended in report above. **(Computer support personnel)**

Hardware Resources : To be determined.

Software Resources : To be determined.

Cost (non -salary): Depends on outcome of evaluations and prototyping. Could be anywhere from $25,000 - $100,000 for startup costs. Annual maintenance costs to be determined.

PHASE IV: Port, as needed, all StreamNet data to installed, final environment. Complete integrated WWW data access system.

Timeline :	June 1st - Sept 30
Objective(s) :	1) Move all StreamNet data into an integrated environment to provide maximal data system performance. 2) Provide access through WWW interface to all StreamNet data.
Task Sequence (key personnel	1) Port appropriate attribute and spatial data into final StreamNet computer environment. **(Duane, Steve, Various GIS personnel)** 2) Modify, as necessary, WWW access system prototyped in Phases I and II to fully exploit all capabilities provided by final system configuration. **(Doug, Steve)**
Hardware Resources :	No additional
Software Resources :	No additional
Cost :	No additional hardware / software costs