DIS/NW/12

Multi-tiered Peer-to-Peer Data Distribution for Spatially Enabled Data

FY 2005 Proposal to the NOAA HPCC Program

September 13, 2004

| Title Page | Proposed Project | Budget Page |

Principal Investigator:	**Ann Matarese**
Line Organization:	NMFS
Routing Code:	F/AKC2
Address:	NMFS/AFSC/RACE 7600 Sand Point Way NE Seattle, WA 98115-6349
Phone:	(206) 526-4111
Fax:	(206) 526-6723
E-mail Address:	Ann.Matarese@noaa.gov

Tiffany Vance	William Rugen	Nazila Merati	Jason Marshall
Tiffany.C.Vance@noaa.gov	Bill.Rugen@noaa.gov	Nazila.Merati@noaa.gov	Jason.Marshall@noaa.gov

Proposal Theme: **Disaster Planning, Mitigation, Response and Recovery**

Dr. Ann Matarese	Dr. Gary Stauffer	Dr. Douglas Demaster
Fisheries Biologist	Division Director	Center Director
NMFS/AFSC/RACE	NMFS/AFSC/RACE	NMFS/AFSC

Multi-tiered Peer-to-Peer Data Distribution for Spatially Enabled Data

Proposal for FY 2005 HPCC Funding

Prepared by: Ann Matarese and Nazila Merati

Executive Summary:

The general public has used peer-to-peer networking for years with programs such as Overnet or Gnutella. Models of peer-to-peer sharing are being used by the science community for collaboration and data dissemination. Peer-to-peer networking takes advantage of distributed architectures, thus reducing overhead of data distribution. The strengths of this model are allowing for spatially enabled collaboration, reducing infrastructure costs, reducing data redundancy and stewarding the idea of distributed computing and resources. Several open-source and COTS technologies provide mechanisms for peer-to-peer data sharing that will allow GIS data to be shared between collaborators. We propose testing a NOAA implementation of peer-to-peer data sharing both through intranets and the Internet, while adhering to NOAA's security and IT models. Data for testing are ichthyoplankton data generated by the Alaska Fisheries Science Center (AFSC) Resource Assessment Conservation and Engineering Division (RACE), Recruitment Processes FOCI (Fisheries Oceanography Coordinated Investigations) group. These data will be shared within other RACE and AFSC Programs. Data will also be shared with the University of Washington Fish Collection (UWFC). This proposal also supports the concepts of the NOAA Enterprise GIS effort by fostering data sharing between Line Offices (NMFS, AFSC and OAR, Pacific Marine Environmental Laboratories, (PMEL). This proposal addresses the HPCC goal "to improve technology for access to critical data, information and critical resources in a manner that increases mission effectiveness and furthers NOAA's service to the nation." It is part of the "Disaster Planning, Mitigation, Response and Recovery" theme as it creates GIS tools that will expedite the transfer of data to coastal professionals.

Problem Statement:

Data sharing is an integral part of NOAA's mission. However, security and confidentiality issues have required that sometimes datasets be distributed via CD or walked between buildings because of firewalls and security concerns. For day-to-day operations and work-in-progress, many NMFS scientists are still using their feet and slow web connections to collaborate with scientists within their own organizations and in other Line Offices. Programs such as Recruitment Processes rely heavily on collaboration between AFSC and PMEL. These restrictions hinder the collaborative aspects of NOAA's projects. There are alternative technologies that might mitigate these restrictions. While these are applicable to bulk data transfers, they have limitations when used with specific file formats and applications such as a geographic information system (GIS). DODS and the scientific community have developed tools for data sharing using OPeNDAP protocols that are client-server based, but these have not addressed GIS. Currently, GIS and web technology are moving towards allowing rapid

dissemination of GIS data products to NOAA users and constituents. Internet map servers were one of the first of these technologies. While map servers are good for large-scale dissemination, they can be too complicated and expensive for peer-to-peer applications. We are interested in using a new GIS server technology, ArcGIS personal (peer-to-peer) server, to test GIS data collaboration between scientists at NMFS, OAR and their university collaborators. This server includes security aspects that will allow us to take into account NOAA's mandate for secure computing. This project will also contribute to the NOAA Enterprise GIS effort by working towards tools and best practices for data sharing and cross-Line Office collaboration. Additionally, it will be a good exercise for testing long distance and large-scale data transfers.

This proposal addresses the HPCC goal "to improve technology for access to critical data, information and critical resources in a manner that increases mission effectiveness and furthers NOAA's service to the nation." It is part of the "Disaster Planning, Mitigation, Response and Recovery" theme as it creates GIS tools that will expedite the transfer of data to coastal professionals. The proposal also involves the "Technologies for Collaboration, Visualization and Analysis" theme in that it enables distributed data integration. The proposal is designed to use a framework approach that enables the proposed solution to be integrated with other related packages (standardized database applications and the ESRI ArcGIS suite of products), and will be interfaced with major off-the-shelf software products. Our proposed solution can be deployed to other projects and Line Offices that require data sharing and mapping capabilities for collaboration and data analysis. We anticipate using new ESRI technology that will integrate into existing ArcGIS and ArcIMS services. This project is a good test bed for an Enterprise level of data sharing in NOAA between NMFS, OAR and NOS. NOS is currently testing ArcGIS Enterprise Mapping and we can leverage off of the strong ArcIMS engineering work demonstrated at CSC.

Proposed Solution:

We propose to test a peer-to-peer implementation of ArcGIS personal server to share GIS data and maps between various organizations working with fisheries-oceanography data. This implementation will be multi-layered in its approach to security and access to data. We propose to use two test cases – a fisheries information system and a PMEL oceanographic database (EPIC) used by AFSC/FOCI. This work would also dovetail into the e-Foci initiative of streamlining data sharing between organizations.

The main goal of AFSC's Recruitment Processes Program is to understand how biological and physical processes contribute to recruitment variability in fisheries stocks in the Northern Pacific Ocean. This work is performed in collaboration with oceanographers at PMEL/FOCI. The program is responsible for a 20-year database of icthyoplankton data used to map the distribution and abundance of fish eggs and larvae. These data are an important part of ecosystem modeling and essential fish habitat efforts within NOAA. Data are in high demand by AFSC scientists, collaborators at PMEL, and other agencies including ichthyologists at UWFC. Static maps of the data are currently available for browsing using a web accessible interactive Icthyoplankton Information System (IIS), but there is no easy way for updated data and metadata to be shared through the IIS. Confidentiality issues preclude the use of a traditional map server.

With the Recruitment Processes Program at AFSC (FOCI) as one peer, we propose testing a multi-tiered approach to peer-to-peer data sharing of GIS data between three sets of partners. The concept of peer-to-peer data sharing has been used by other organizations, such as the OpenGIS consortium, and will speed data sharing, paper writing and analysis for each program. The first partners would be RACE division scientists and FOCI's ecosystem modeling partners at the Resource Ecology and Fisheries Management (REFM) division at AFSC. The second and third sets would be FOCI's partners at OAR/PMEL/FOCI and external collaborators outside NOAA at the UW Fish Collection. As we move between different tiers of organizations, we will be testing the ability to restrict and allow data sharing and networking capabilities without compromising NOAA's security structure.

We propose implementing and testing various levels of ArcGIS data servers to test the feasibility of peer-to-peer networking for data collaboration between users at NMFS AFSC using different data servers and networks in the same physical building; between OAR and NMFS facilities at the Western Regional Center (WRC) who have a firewall between the buildings; and between NMFS and the UWFC who communicate via the commodity internet and Abilene connections. The proposed solution tests the capabilities of the ArcGIS personal server for data sharing in a collaborative environment, extending the use of ArcIMS, and how we will implement data sharing while adhering to NOAA's IT security model. This multi-tiered approach will be helpful in determining if an Enterprise approach to data sharing is feasible within NOAA.

Give a list of the major activities being considered in this project.

1. Identify databases to be tested and decide functional levels of database information that will be shared.

2. Analyze level of security in the ArcGIS server and decide if it works with NOAA's existing IT security model. Test public/private keys, access control models, authoritative servers, and encryption.

3. Work with IIS Webmaster and Recruitment Processes database manager to determine how authentication protocols will be managed; determine time lag in data distribution for the more public sites.

4. Work with FOCI data manager to establish and identify data types to be shared within AFSC.

5. Install and test an ArcGIS personal server at AFSC for data publishing and data sharing within and between RACE and REFM, work with AFSC's Office of Fisheries Information Services on security model.

6. Install and test an ArcGIS personal server and ArcIMS application using Java Connector at PMEL. Write custom code if necessary to encrypt data. Test network connectivity, bandwidth during peer-to-peer sharing.

7. Test subset of RACE databases for peer-to-peer sharing and collaboration with UWFC.

8. Based on performance of the peer-to-peer connections, fine tune connectors and file transfer options for better performance.

9. Create a "best practices document" for levels of peer-to-peer sharing within NOAA; provide code to other NOAA programs interested in testing data sharing.

Analysis:

This project will allow NOAA to take advantage of new peer-to-peer data sharing methods without deploying a full ArcIMS installation. It would foster collaboration between groups without spending large amounts of time with file transfers, data formatting and networking issues. The test case of data sharing between two programs in the same organization, data sharing between a collaborative NOAA program and one with an outside user, will test NOAA's IT security model and the ability to provide secure transfer of data in progress.

Appropriateness: This project addresses more than one HPCC networking goal as well as the larger goals in the HPCC program. The multi-tiered approach to GIS peer-to-peer sharing can be extended to other groups in NOAA and between Line Offices working together towards NOAA's goals of scientific data stewardship and ecosystem management. The proposal meshes well with NOAA's Enterprise GIS efforts by making geo-spatial data accessible and reducing data redundancy.

Technology: This proposal will utilize very new technologies to create peer-to-peer data sharing using ArcGIS technologies and extend the security model to adhere to NOAA's IT security policies. ArcIMS technology is currently used within NOAA for data dissemination, but ArcGIS personal server for peer-to-peer data sharing has not been tested. The proposed solution will also use NOAA's networks to test data transfer times and data quality between Line Offices.

Scope: The scope of this project is wide; the partners are two NOAA Line Offices and an outside collaborator. The effort is integrated in the Enterprise GIS framework at NOAA. If successfully implemented, the project would be easily transferred to other groups interested in peer-to-peer data sharing for in-process projects as well as collaborating with outside groups. As the technology matures and the security model is finalized, the investigators would like to implement larger scale peer-to-peer sharing of baseline GIS datasets used by NOAA.

Leverage: This project leverages off of research done at ESRI, the developer of COTS technology that is widely used by NOAA to enable peer-to-peer data sharing within an ArcGIS model. We will look at the OpenGIS Consortium model for peer-to-peer sharing as well. NOAA GIS users have implemented ArcIMS applications for data display and dissemination of processed data sets, but not of data that are currently used for collaboration. Security concerns addressed by OPeNDAP projects from FY03 will be helpful. Existing investments in IIS and EPIC will minimize time necessary for data formatting and processing. NOAA West Network security and network efforts will provide the investigators with a backbone infrastructure to test the peer-to-peer applications.

Cost Benefit: The benefit-to-cost ratio of this project is high. Implementation of a peer-to-peer network to share GIS data will enhance the management and transfer of data between collaborators. Testing the multi-tiered approach to security between programs, collaborators in different Line Offices and non-NOAA groups will help further the framework to establish a baseline effort of data sharing towards the Enterprise GIS efforts at NOAA.

Compare the selected plan with the other alternatives: The alternative to this proposal would be to continue to set up FTP sites, send collaborators static maps of raw data, create hard copies of data files or continue to not share data between groups. None of these alternatives deliver timely data products to collaborators and in the case of rapid necessity to respond, a time lag that may affect the quality of the data exists.

Performance Measures:

This project will succeed if it creates a successful implementation of peer-to-peer data sharing of identified data sets between groups at AFSC working on integrated ecosystem models, groups working on large-scale data sets at OAR and AFSC, and a successful deployment to AFSC's outside collaborators. Another measure of success will be the implementation of the model to other Line Offices in NOAA. The project will also succeed if it develops new metrics for performance of peer-to-peer interaction.

Milestones

Month 01 – Evaluate databases for use and determine data to be shared at different levels, analyze ArcGIS personal server security model, review OpenGIS Consortium model for p2p and other data serving models

Month 02- Install personal servers between AFSC/FOCI and RACE/REFM, test password authentication using LDAP servers

Month 03 – Install personal server at PMEL, work on key encryption, revise security if necessary, test with WRC facility, test bandwidth and performance

Month 06– Install personal server at UWFC, work on authentication servers, revise security if necessary, and test bandwidth and performance metrics on NWN Abilene

Month 08- Test ArcIMS application for whiteboarding (using secure connection), work with CSC

Month 10- Based on performance of peer-to-peer connections, fine tune connections and file transfer options, work with NOS/CSC senior software engineer

Month 12 – Create "best practices document" for levels of peer-to-peer sharing within NOAA, make written code (Java Connectors) available

Deliverables

- Example of peer-to-peer data sharing between RACE and RACE/REFM Divisions within the AFSC
- Example of peer-to-peer data sharing between AFSC and PMEL
- Example of peer-to-peer data sharing between AFSC and UW
- Best practices document
- Robust peer-to-peer data sharing tools