Joint Proposal to the **Executive Board** of the **International Association of Aquatic and Marine Science Libraries and Information Centers (IAMSLIC)** and to the **Advisory Board** of the **Aquatic Science and Fisheries Abstracts (ASFA)** for the development of the **Aquatic Commons Initiative.**

Written by:

Stephanie C. Haas, Assistant Director, Digital Library Center, University of Florida Libraries
haas@uflib.ufl.edu

Reviewed by: Priscilla Caplan, Assistant Director, Digital Library Services, Florida Center for Library Automation (FCLA) pcaplan@ufl.edu; Pauline Simpson, National Oceanography Centre, University of Southampton, UK ps@noc.soton.ac.uk; Steve Watkins, California State University, Monterrey Bay steve_watkins@csumb.edu

History

For the last three years, IAMSLIC has acknowledged the value of institutional repositories and encouraged members to implement IRs within their own institutions. Repositories are now spreading because they provide stewardship and global access to the digital assets of institutions and organizations. Impetus for the development of IRs has also come from the Open Access and Open Archives movements. Similarly, under the guidance of the ASFA Advisory Board, the ASFA Partnership has recognized the value of IRs, and Cambridge Scientific Abstracts (CSA) has increased the harvesting of metadata for digital objects for inclusion in ASFA. For many years, the need to provide access to full text documents has been recognized by the ASFA Partnership and IAMSLIC. Technological advances and the development of IRs now make it possible to address access to digitized texts.

Purpose and Stakeholders

The current proposal is intended to establish an Aquatic Commons initiative to provide a central portal to the literature in marine and aquatic sciences. This initiative will provide an Aquatics EPrint repository for the deposit of metadata and digital texts where IT support at the local level is unstable, and a harvester that will harvest and aggregate metadata from digital repositories that serve OAI compliant metadata including the Aquatics e-print repository. Because of the extensive experience of the staff at the Florida Center for Library Automation (FCLA) in developing harvesting functionality, digital collection and metadata development, and collaborative project development and management, they will be responsible for the implementation and ongoing maintenance of the Aquatic Commons model elucidated here.

The model was designed to integrate the efforts of the total community by harvesting metadata where available and by creating repository and harvesting opportunities where needed.

Many larger institutions and agencies are creating repositories related to their own missions. The IOC has created OdinPubAfrica which is aimed at collecting and serving digitized scientific and technical publications on African marine science from Ghana, Indian Ocean commissions, Kenya, Mauritania, Morocco, Mozambiqu, Senegla, Seychelles, Tansania, and Tunisia. Woods Hole, Scripps, Oregon Institute of Marine Biology, University of Oregon and Southampton all have digital collections related to their oceanographic/marine missions. Presently, NOAA is working on a repository, and Jan Haspeslagh is advancing a marine science repository for the

Netherlands. All of these efforts are supported by stable IT infrastructures with both adequate and ongoing financial and human resources.

To reiterate, the main goals of the Aquatic Commons are: 1) to harvest into a searchable database the metadata from the repositories that exist and those that are currently under development, and 2) to create the Aquatic EPrint repository for smaller institutions and research endeavors that do not have access to stable IT support. Existing repositories with OAI capability include Woods Hole, Scripps, Oregon Institute of Marine Biology, University of Oregon, and Southampton; metadata from additional repositories can be added to the system as they become OAI compatible. The Aquatic e-print repository will accept all aquatic related digital texts. It is hoped that an active collaboration with the FAO Fisheries Department and with the ASFA Partners will bring a particular focus to freshwater and estuarine materials from developing countries. Similarly, it is hoped that the IAMSLIC members will also use the Aquatics EPrint repository for the grey literature developed in the hidden centers and smaller academic units of their institutions.

The initial stakeholders that have been identified include:
1) Researchers and research institutions in the marine and aquatic sciences;
2) UN, International, and National ASFA partners;
3) CSA;
4) FAO ASFA Secretariat;
5) FAO Fisheries Department;
6) Other marine research agencies such as IOC, NOAA, etc.;
7) IAMSLIC and its affiliated regional groups; and
8) Florida Center for Library Automation (FCLA).

Justification of Need

One of the recurrent needs expressed by users of the ASFA database (and consistently vocalized by IAMSLIC members over the years) has been the inability to obtain the grey literature documents indexed by the ASFA partners. Because most documents are now created electronically whether on a networked or stand alone computer, these electronic files offer the opportunity to readily share the science globally through repository networks.

While it is difficult to estimate the total number of digital documents that might be deposited in the Aquatics e-print repository, three individuals indicated immediate interest at the IAMSLIC Conference in Rome: Anton Immink, Communications Officer, Aquaculture and Fish Genetics Research Programme, Stirling, UK; Simon Wilkinson, NACA, Thailand; and Catalina Lopez-Alvarez, Universidad Autonoma de Baja, Mexico. As part of the summer testing of the repository software, Guillermina Cosulich, INIDEP, Argentina is actively participating in the planning.

System Architecture

The components of Aquatic Commons include:

an Aquatics EPrint repository,
existing OAI-capable repositories,
a harvester,
an OAI data provider (to interface with the zebra server),
a search and retrieval interface,
a database, and

a zebra Z39.50 server to interface with the IAMSLIC Z39.50 Distributed Library

During the summer of 2005 as part of the planning phase, FCLA implemented and tested all of the component parts. At that time the software used for the search and retrieval interface was proprietary. It was decided that if full implementation was reached that only open source software would be used in the final architecture configuration.

FCLA Qualifications

The Florida Center for Library Automation provides computer services that assist Florida's university libraries in their daily operations and record keeping as well as providing students and faculty with electronic access to scholarly materials. These services include the operation of a shared Integrated Library Management System, licensing of electronic resources, and providing the repository and support for digital versions of library-owned collections. The university system of Florida consists of ten universities, one college, twelve off-campus centers, seven agricultural research and education centers, and sixty-nine county cooperative extension programs located throughout the state. Additionally, FCLA has taken on a leadership role in the development of harvesting initiatives in the state in its work with the State Library to develop Florida Electronic Library (http://www.flelibrary.org/index.cfm).

The staff of FCLA has extensive expertise in mounting standards based digital collections and in creating tools for the management of digital collections over time. It is one of the first U.S. based organizations to have created a viable digital preservation repository (FCLA Digital Archive http://www.fcla.edu/digitalArchive/index.htm) and is a recognized leader in the U.S. for its innovative and collaborative approaches to building digital resources.

Perhaps the greatest benefit of involving FCLA in this project is the assured stability of its continued performance beyond the careers of any individual staff member of FAO, ASFA, or IAMSLIC. As a legal entity, FCLA has the ability to commit to binding agreements for work performance and system support.

The Aquatic Commons development would be supervised by Priscilla Caplan as the Assistant Director for Digital Library Services.

Succinctly, the hosting requirements for the Aquatic Commons model include:
1. Unix systems administration and secure network and computer server facilities,
2. technical knowledge and programming skills needed to develop an integrated repository/harvester system,
3. ability to evaluate and implement open source software for the harvester, EPrint repository, search interface, and database functions,
4. ability to work with existing OAI-compliant data providers to integrate them into the Aquatic Commons,
5. knowledge of different metadata formats,
6. understanding of the server requirements associated with diverse digital formats, and
7. ability to work collaboratively with a diverse clientele.

Development Phases

The development of the Aquatic Commons will occur in three phases.

Phase 1: **Development of the Aquatics e-print Repository**.
(Estimated implementation time: 6 months from funding)

The open source EPrints software is available from the University of Southampton and has a large community of users. The FCLA staff has already loaded and done initial testing of this software. During Phase 1, a dedicated server would be purchased, installed, and the software reloaded. Because of the large number of options offered by the software, extensive configuration is possible, requiring active participation by intended contributors. All setup requirements including appropriate authority listings, preferred formatting for fields, etc. would be determined in consultation with IAMSLIC and the technical staff at ASFA. Training materials will be developed by IAMSLIC members and administrative duties associated with metadata review and object acceptance would be delineated and assigned to IAMSLIC members.

**Phase 2: Development of the Aquatic Commons harvester, an OAI provider,
a search interface, and database**
(Estimated implementation time: 8 months from funding)

FCLA will investigate the open source *Arc* software as a potential product for providing search and retrieval capability, and possibly harvesting. Arc functionality is described in an article by Liu Ziaoming and others, entitled *Arc−An OAI Service Provider for Cross−Archive Searching* [accessed 11/9/05 at http://www.ils.unc.edu/~mln/jcdl-arc.pdf]
         "*Arc* harvests metadata from several OAI compliant archives, normalizes them, and stores them in a search service based on a relational database (MySQL or Oracle)." Because of the rapidity with which software evolves, if a more effective software appears, FCLA retains the right to change open source software packages.

The harvester system can be run on the same server as the Aquatics e-print repository, thus costs will be for staff for development and ongoing maintenance only.

**Phase 3: Implementation of zebra Z39.50 server to interface with IAMSLIC Z39.50
Distributed Library**
(Estimated implementation time: 2 months following establishment of either the Aquatics EPrint repository or Aquatic Commons database)

FCLA has a zebra Z39.50 server in use for another application. It can be configured to work with the Aquatic EPrint repository or with the more comprehensive Aquatic Commons database once it is established.

By implementing this server interface, any search query launched in the IAMSLIC Z39.50 Distributed Library will also be searching the metadata of the repositories. Effectively, this means a user will be searching across the Aquatic Commons, the distributed catalogues of more than 40 libraries plus their branches, and the serials listings from 40 institutions

Pauline Simpson suggests "That the Aquatic Commons is one more building block toward a superb one stop shop with document delivery at the hub. We are building a thematic portal with built in cross searching."

Only funding required is in staff costs of FCLA and Steve Watkins to configure the interface.

BUDGET

The majority of costs associated with Aquatic Commons are related to the hardware and programming to setup and maintain the system. Additional costs are associated with administrative functions related to the submitting of records to the Aquatics EPrint repository, development of training materials, and the promotion and education of potential users of the repository. Because IAMSLIC is a wholly volunteer organization, it would be most appropriate if it could meet its fiscal responsibilities through cost share of members' time. The major activities requiring members' time are delineated below; however, no time has been included for determining and testing configuration options for Aquatic EPrint repository and a minimal amount of time has been included for the development of training materials for Phase 2 and none for Phase 3. Ongoing promotion and education has only been costed in Phase 1, but obviously will be continuous.

Phase 1: Aquatic EPrint Repository	ASFA/FAO Funding	IAMSLIC (cost share)
Personnel		
FCLA Programmers 120 hrs @ $15	$1,800	
Purchase, install and configure server computer 15 hours @ $20	$ 300	
Develop training materials in English, French, and Spanish by IAMSLIC members 20 hrs/ language @ $12		$ 720
Administration of submittals and coordinating setup between ASFA/FAO, IAMSLIC and FCLA 10% of Stephanie Haas' time for one year		$6,700
Consulting on e-repository software and setup with Pauline Simpson, Southampton 10 hrs		$1,000
Promotion of and education about the repository by IAMSLIC Resource Sharing Committee (sessions at 5 regional conferences, etc.)		$1,000
Hardware/Software		
Server, dual cpu, 4GB memory, 156GB internal disk (projected to serve about 75,000 pdf files	$5,000	
Tape cartridge for backup	$ 200	

Red Hat Linux (OS)	$ 50	
Tivoli (backup server)	$ 50	
Tripwire (security)	$ 300	
Total	**$8,900**	**$9,420**

Phase 2: Development of the Aquatic Commons harvester, an OAI provider, a search interface, and database	ASFA/FAO Funding	IAMSLIC (cost share)
Personnel		
FCLA programmers 200 hrs @ $15	$3,000	
Develop training materials in English, French, and Spanish by IAMSLIC members 20 hrs/ language @ $12		$720
Total	$3,000	$720

Phase 3: Implementation of zebra Z39.50 server to interface with IAMSLIC Z39.50 Distributed Library		IAMSLIC (cost share)
Personnel		
FCLA programming 40 hrs @$15	$ 600	
Steve Watkins' time to implement interface estimated at 10 hrs		$1,000
Total	$600	$1,000

The costs listed above represent the start-up costs for the various system functions.

Annual ongoing costs for FCLA system maintenance for the *Aquatic Commons* are:

Hardware / network
      Server maintenance $ 500
      Hardware replacement cost $ 1,000
      Network cost $ 86
Software
      Red Hat Linux (OS) $ 50
      Tivoli (backup server) $ 50
      Tripwire (security) $ 165
Staff
      Ongoing maintenance and
      support (20 hrs/mo) $ 3,600

Total annual ongoing costs   **$ 5,451**

This includes maintenance for the Aquatics e-print repository, a harvester, an OAI provider, interfaces to existing OAI-compliant servers, a search and retrieval interface, a database, and a zebra Z39.50 server to interface with the IAMSLIC Z39.50 Distributed Library. For the Aquatics EPrint repository alone, hardware and software costs would be the same, but annual staff support would be 5 hrs/mo ($900). The hardware replacement cost is intended to provide a sufficient amount over a five year period to replace the server.

Ongoing matching costs from the IAMSLIC membership would be continued efforts to inform all levels of the aquatic community of the availability of the Aquatic e-print repository, to provide guidance in the submittal of items to the repository, and where appropriate to submit digital items from their own organizations and institutions.

Awareness, Education, and Training

As indicated in the budget outline above, the IAMSLIC Resource Sharing Committee will assume the main responsibility for preparing training materials in French, Spanish, and English related to the Aquatics e-print repository and/or the Aquatics Commons database. Additionally, it is expected that all collaborating partners involved in the Aquatic Commons project will inform their colleagues of these initiatives. It is hoped that through colleague networking interest in contributing to the Aquatic Commons will be sparked much as it has been with the IAMSLIC Z39.50 Distributed Library project. Because Phase 3 of this project will also permit direct searching of the Aquatic Commons database through the developed IAMSLIC Z39.50 Distributed Library, those participating in this established program should find this enhancement of particular value in providing more access to relevant full text.

Jean Collins of FAO, who is very supportive of the Aquatic Commons initiative, wrote:

> "I think an IAMSLIC Aquatic Commons is a good idea...I realise that there are constraints and many legal-type issues that would have to be ironed out...It would be the best opportunity for me to help get developing countries involved and their publications included.
>
> It would be good to start with a critical mass of publications - such as those of NACA and MRC. As regional organizations they also have more resources than a lot of national fisheries institutions - they would provide a better testing ground than those who only produce a half dozen documents a

year.

> My role (although I don't always determine that myself!) would be providing contacts and promoting the Aquatic Commons, providing guidelines/training for developing country libraries, getting funding."

As indicated in the Justification of Need section above, NACA has already indicated its interest in participating. Jean's willingness to promote and provide training assistance at the local level in developing countries is a key factor for developing a truly valuable repository.

Preliminary discussions with CSA have already begun about harvesting the records from either the Aquatics e-print repository or from the Aquatic Commons database for inclusion in ASFA. Such inclusion will stimulate increased use of this material and is expected to greatly enhance the value of the ASFA database to all researchers.

ADMINISTRATION

In order to provide oversight and to further the goals of the Aquatic Commons initiative, one member from each organization should be appointed to serve on an Aquatic Commons Steering Committee. The exact charges, duties, and communication avenues will be decided later.

In terms of administration, all aspects of the technical functioning of the system will be the responsibility of FCLA staff. The administrative review of submittals to the Aquatics e-print repository will be the responsibility of IAMSLIC members who will agree to serve gratis for a period of at least one year. Outgoing submittal review administrators will train the incoming administrator and provide consultation as needed.

The Aquatic Commons database resulting from the harvesting of extant repositories and the provision of OAI metadata are all automated functions that will be administered by FCLA. It will be the responsibility of all partners to alert the Aquatic Commons Steering Committee to new repositories coming online.

SUSTAINABILITY AND EVALUATION

In terms of financial commitments, the costs associated with the development and maintenance of Aquatic Commons are minimal. IAMSLIC is an all volunteer organization, but its inkind contributions in terms of the development of the project and the administration of the Aquatic e-print repository are substantial. Initially, Stephanie Haas has agreed to act as the facilitator to coordinate the development of the three phases and to act as the submittal review administrator for the repository. Pauline Simpson has agreed to act as a consultant to Stephanie and FCLA in the initial setup and implementation of the repository system. Steve Watkins will be providing match too in setting up the interface with the zebra Z39.50 server.

The Florida Center for Library Automation projects that $4,500 will be needed on a yearly basis to maintain the Aquatic Commons components. Current proposed changes in IAMSLIC membership dues are likely to affect IAMSLIC's ability to contribute to this maintenance fee. However, it is also likely that the continued e-print repository administrative duties that are the responsibility of IAMSLIC will come close to matching that amount.

It seems logically that the cost-benefit evaluation of this project will have to be measured qualitatively by those contributing to the e-print repository, and quantitatively by measuring use of the e-print repository and/or the Aquatic Commons database. Harvesting from these two sources will also be a measure of value. A critical mass of documents must be deposited before searching the e-print repository attains efficace; the same holds true for the creation of the Aquatic Commons database from harvested metadata. During the start up phase of any of the component, any evaluative measures will be inaccurate, although statistics will be an important functionality of the software. If possible, a commitment to fund the ongoing costs for a two year period to commence after the set up year would be highly desirable. At that point, a more accurate assessment of potential benefit can be determined.

As with all initiatives, the sustainability of the Aquatic Commons project is directly tied to its perceived value by the stakeholders. There has been an ongoing need for full text access to the ephemeral, or grey literature, indexed by ASFA for years. Because more and more scientific reports are born digital, the capture and preservation of these files is becoming increasingly critical. IOC has begun to capture a very small proportion of the digital literature related to marine science. It has plans to create similar repositories in other circumscribed regions of the world. The role of Aquatic Commons is to compliment these efforts by addressing the repository needs of the non-circumscribed marine and aquatic regions. The documents related to freshwater and estuarine systems are more critical in many parts of the developing world than are the marine documents. For many years, the activities of IOC, FAO, FAO ASFA, and IAMSLIC have inextricably tied the organizations together. The joint support of an initiative such as Aquatic Commons will not only formally strengthen relationships but will draw on the expertise of each organization in building digital resources of global value.