The Digital catch: an integrative role for IAMSLIC in the worlds of metadata, harvesters and repositories. *Extended Abstract*

SIMPSON, PAULINE National Oceanography Centre Southampton (U.K.) ps@soc.soton.ac.uk and HAAS, STEPHANIE C. Digital Library Center, University of Florida (U.S.A.) haas@uflib.ufl.edu

INTRODUCTION
From the electronic file of a single researcher's paper to the digitization of thousands of photographs and maps, the creation of digital objects is ubiquitous. Larger marine and aquatic institutions have the resources to create and share open archive compliant metadata and globally serve their digital collections. Smaller institutions and researchers may create electronic documents but have no means of sharing their knowledge over the Web. In support of the Open Access Movement, repositories of many flavours are now being implemented around the world and researchers are confronted with choices: this presentation reviewed some of the many options open to researchers to archive their digital documents, and presented some of the options that IAMSLIC is pursuing to facilitate the exchange of metadata and to provide access to digital objects from the aquatic community.

Part 1 : Species Diversity (Pauline Simpson)

IAMSLIC has acknowledged the importance of the growing global Open Access and Institutional Repository movement over the years with the inclusion of presentations on these topics in the last three annual conferences.

Since the year 2000 the implementation of Institutional Repositories, has complemented the already existing subject based repositories, the most successful being ArXiv , http://arxiv.org/, but others have had a more cautionary success. With the support of funding from such organizations as the Andrew Mellon Foundation USA and the JISC UK and powered by the Information Community, the number of repositories has risen from 112 in 2002 to over 460 in Sep 2005.

However, repositories are not just confined to subject or institutional content but now a list would include : National; National / Subject; International; Regional; Consortia; Funding Agency; Project; Conference; Peer to Peer and Static; Media Type e.g., Theses; and even Publisher – journal archives and Data Repositories/Archives both of whom are now making their repositories OAI compliant. The Institutional Archives Registry http://archives.eprints.org/ maintained at Southampton reflects the substantial growth in repository 'flavours' on a truly global scale. The problem is that the repositories are not mutually exclusive and already there are real tensions between Institutional Repositories and Designated Data Centres concerning dataset curation. The multiplicity of repositories presents a dilemma for the researcher; the Funding Agencies are mandating deposit in repositories and he/she wants to enter metadata and the full text deposit only once. The logical route is to deposit in their Institutional Repository because institutions

can offer to centralise a distributed activity; provide a framework and infrastructure; has the permanence that can sustain changes; should provide Stewardship of Digital assets with an interest in preservation; but most importantly want to offer a digital showcase for the research, teaching and scholarship of the institution.   At present it is possible for metadata to be harvested by aggregator services like the service provider, OAIster, but this may not be the repository target of choice for the researcher.

Whilst many are now implementing institutional repositories already there is a move to build on the repository movement. With the budgeoning e-Research agenda it is seen as an important component contributing to the Knowledge Cycle of experimentation, analysis, publication, research and learning.   Essential to e-Science is the concept of joined-up- research providing the ability to follow an audit trail from raw data through to information or learning object at any point of discovery.

New projects are now being funded in which the automated linking of text and data is being investigated.   Southampton is involved in the CLADDIER Project (Citation, Location And Deposition in Discipline and Institutional Repositories).   The CLADDIER system will be a step on the road to a situation where (in this case, environmental) scientists will to be able to move seamlessly from information discovery (location), through acquisition to deposition of new material, with all the digital objects correctly identified and cited. Perhaps serendipitously, one of the outcomes of this project will be the ability for researchers to push their metadata to repositories of their choice, i.e. deposit in their Institutional Repository and target ('push') their metadata to other repositories, of whatever flavour, or to service providers.

At the last three IAMSLIC conferences the call has been for IAMSLIC members to implement repositories within their institutions and it is encouraging that IAMSLIC now includes some 20 repositories amongst it members. The vision included that IAMSLIC would need to host a repository for those who did not have the support to set up their own. Thus members repositories and IAMSLIC's repository would provide the aquatic research content that could be harvested by an IAMSLIC service provider offering discovery and location of aquatic and marine science research through a one search interface: in Part 2 of this paper proposed to be called the Aquatic Commons.

Part 2 : The Aquatic Commons (Stephanie C. Haas)

IAMSLIC has benefited greatly from Pauline's involvement in the repository movement since its inception. She has continually updated the membership on the latest developments and Southampton continues to be on the forefront of integrating digital information.

Over the last few years, Pauline and others have developed several models to envision how IAMSLIC members could share digital metadata and objects. The most comprehensive and successful has been the implementation of the IAMSLIC Z39.50

Distributed Library by Steve Watkins. This project has proven invaluable for sharing holdings information to facilitate interlibrary loans.

While institutional repositories are beginning to be developed within the IAMSLIC membership, the desire to share digital documents has become an ongoing need for institutions that do not have stable IT support. Because of the discrepancies in system capabilities among IAMSLIC members, a model that could incorporate all levels of technological capability was needed.

The *Aquatic Commons* model discussed here was developed to address this situation.

Aquatic Commons is a model for digital resource sharing between stakeholders in the marine/aquatic information world. Its integrative architecture accommodates researchers and research institutions at all technological levels. The model includes repositories, harvesting functions, searchable database creation, and integration with IAMSLIC's Z39.50 distributed library and the ASFA database. The Florida Center for Library Automation (FCLA) provided technical expertise, computer hardware/software, and the programming to develop the proof-of-concept functionality of the model.

Aquatic Commons is being developed to:
1) Create a central metadata and digital document reservoir related to marine and aquatic science information worldwide.
2) Support IAMSLIC's long term goal of helping researchers and the public freely access needed information.
3) Integrate the efforts of the total community by harvesting metadata where available and by creating repository and harvesting opportunities where needed.
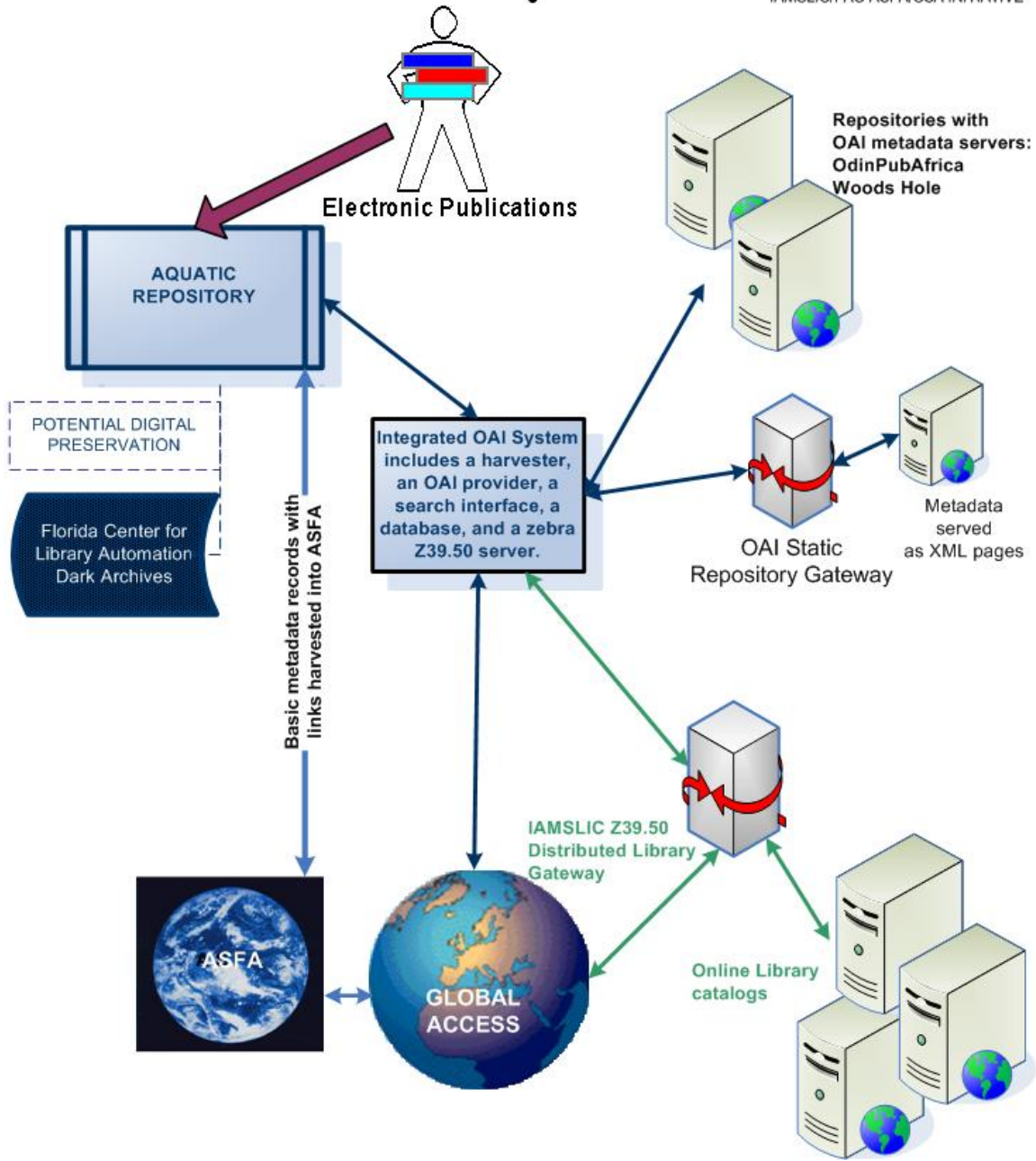
The identified stakeholders in the development and management of the Aquatic Commons include: 1) researchers and research institutions in the marine and aquatic sciences, 2) UN, International, and National ASFA partners, 3) CSA, 4) FAO ASFA Secretariat, 5) FAO Fisheries Department, 6) Other marine research agencies such as IOC, NOAA, etc., 7) IAMSLIC and its affiliated regional groups, 8) the University of Florida Libraries, and 9) Florida Center for Library Automation (FCLA).

As presently envisioned, the Aquatic Commons architecture consists of an integrated Open Archive Initiative (OAI)* System that includes: a harvester, an OAI provider, a search interface, a database, and a zebra Z39.50 server. If the project moves to a production level, the system will be based on Open Access software and will be scalable to accommodate new repositories coming online. A diagram of the model is shown on the next page.

Aquatic Commons:
a networked reservoir to digital documents

IAMSLIC/FAO ASFA/CSA INITIATIVE

Electronic Publications

AQUATIC
REPOSITORY

Repositories with
OAI metadata servers:
OdinPubAfrica
Woods Hole

POTENTIAL DIGITAL
PRESERVATION

Florida Center for
Library Automation
Dark Archives

Basic metadata records with
links harvested into ASFA

Integrated OAI System
includes a harvester,
an OAI provider, a
search interface, a
database, and a zebra
Z39.50 server.

OAI Static
Repository Gateway

Metadata
served
as XML pages

IAMSLIC Z39.50
Distributed Library
Gateway

ASFA

GLOBAL
ACCESS

Online Library
catalogs

Overview of the Components

Aquatic Commons is designed as an OAI integrated system that will functionally:
- ❖ Harvest and create a searchable database of OAI compliant metadata from extant repositories or OAI static repositories including the Aquatic repository developed as part of this model,
  and in turn
- ❖ Serve OAI complaint metadata to other services.

It will also create:
- ❖ An *Aquatic eprint Repository* to house digital works and metadata created by researchers or institutions that don't have stable IT support.

and
- ❖ A zebra Z39.50 server that will interface with the IAMSLIC Z39.50 distributed library.

Optional functionality may include digital archiving at the FCLA Digital Archives of publications submitted to the Aquatic e-print Repository server, and CSA's harvesting of metadata from the Aquatic e-print Repository for inclusion in ASFA.

During the pilot phase, FCLA successfully implemented a test bed of the e-print repository software. Guillermina Cosulich of Argentina tested the metadata creation and document uploading capabilities of this Aquatic e-print repository. Additionally, FCLA harvested and made searchable metadata from six collections including the *Aquatic eprint Repository*. The repositories harvested included: Baltic Marine Environment Bibliography 1970- , W. M. Keck Laboratory of Hydraulics and Water Resources Technical Reports, Oregon Institute of Marine Biology, Woods Hole Oceanographic Institution, and ODINPubAFRICA.

The harvested metadata resides in an umbrella Aquatic Commons repository. The interface is shown below:

Results from executed searches indicate the number of records from each repository harvested: [This search was done on the term "ocean".]

Digital archiving, metadata harvesting by CSA, and implementing the zebra Z39.50 interface that will be used to link the Aquatic Commons repository to the IAMSLIC Z39.50 Distributed Library were not included in the proof-of-concept of the model.

The estimated budget to create a production level model is:

SET UP COST ESTIMATES (year 1)

Hardware / network
       Server, dual cpu, 4GB memory,

156GB internal disk	$	5,000
Tape cartridge for backup	$	200
Software		
Red Hat Linux (OS)	$	50
Tivoli (backup server)	$	50
Tripwire (security)	$	300

Staff

Development and setup (320 hours)	$	4,800

Total one-time costs	$	10,400

ANNUAL ONGOING COST ESTIMATES (starting year 2)

Hardware / network
Server maintenance	$	500
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 $ 4,451

This model was presented for consideration to the IAMSLIC membership and also to the ASFA Board. The IAMSLIC Board indicated that they would like a fuller proposal to be compiled and have charged Stephanie Haas (chair), Simon Wilkinson, Peter Bruggeman, Guillermina Cosulich, and Marcel Brannemann with completing this task by November 30, 2005. I will be asking Pauline Simpson and Craig Emerson, CSA, to be joining the discussions.

Before the closure of the conference, three individuals indicated interest in submitting digital documents to the Aquatic e-print Repository: 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.