

**INTERGOVERNMENTAL OCEANOGRAPHIC COMMISSION  
(of UNESCO)**

**Nineteenth Session of the IOC Committee on International Oceanographic Data  
and Information Exchange (IODE-XIX)  
Trieste, Italy, 12-16 March 2007**

**OceanDocs: Repository Network on Oceanography  
and Marine Science**

(by Marc Goovaerts, Hasselt University Library, Belgium)

The OdinPubAfrica project (FUST-project: - see) developed a repository for scientific literature of African marine science. During the project other Odin groups, in the first place OdinCarsa were interested to develop a similar repository project for their region. As a result the OdinPubAfrica repository was extended to accept other Odin groups and was renamed to OceanDocs (<http://iodeweb1.vliz.be/odin> - <http://www.oceandocs.net>).

This program proposal is based on the experiences of OdinPubAfrica and the interest by new partners.

## **1. GENERAL OBJECTIVES**

OceanDocs has the following objective:

- Development of a network of OAI-compliant repositories (Institutional and the OceanDocs Central repository) providing access to full-text publications created by scientists affiliated to oceanographic and marine institutes and managed by their libraries and information centres.

The creation of the OceanDocs network will have the following immediate and direct advantages:

- Make scientific publications of oceanographic and marine institutes more easily and freely accessible to the research and management community,
- Make local and regional grey literature available on a worldwide scale,
- Enhance the internal scientific communication;
- Facilitate publishing of research findings (e-journal as well as e-archive), specifically for scientists in developing countries thereby promoting their research and increasing their access to the international research forum.

The interest of the project is to have as wide a scope as possible. Different sort of documents, grey literature and commercially published papers can be collected as long as they have a scientific significance.

- Articles: an article evolves in his lifecycle from the author's version (preprint) to a journal version (refereed) to a postprint version (adapted author's version). The journal version is not available most of the time, because of copyright limitations. Therefore pre- or postprint versions of an article will be archived in the repository in the first place. If possible a journal version can also be included. Institutional journals can be integrated as a separate collection in the repository, if they are free of any copyright restriction.

- Scientific reports (annual, regular) - project reports
- Theses
- Technical progress notes (unrefereed)
- Conference papers

The sort of documents which have to be excluded are administrative reports, PowerPoint presentations, etc.

The goal of the project is to make the documents in the repository freely available for the general public. Only in specific cases a limitation of access can be installed.

## 1.1 REPOSITORY

A repository deals with the scientific production of a(n) (group of) institution only. In the first place it is a way to make the own scientific publications available to the own community and to other interested scientists.

A repository is not an electronic library. In an e-library all material relevant to a community is collected. The storage of all these documents on a library server cannot be realized as a result of the limitation enforced by copyright and author right laws. An electronic library will use search tools for electronic accessible documents (commercial publishers and open access document collections). Through the repository the own scientific output can be managed and made electronically available.

Universities and research centres all over the world creates their repository to manage and present their scientific output in Open Access. The OpenDOAR directory (<http://www.opendoar.org>) refers to 844 official repositories. The largest one DSpace@Cambridge contains already 179.719 documents (Feb. 5 2007). In the field of oceanography and marine science the number of repositories is growing: e.g. [Archimer](#), [Archive Institutionnelle de l'Ifremer](#), [ePrints Soton \(Oceanography\)](#), [Woods Hole Open Access Server](#), [Marine & Ocean Science ePrints Archive @ Plymouth](#), [DRS at National Institute Of Oceanography](#) and of course [OceanDocs](#).

## 1.2 NETWORK - HARVESTING

Repositories are local web databases containing the scientific output of an institute. They are part of a distributed network. Every repository is OAI-compliant. It means that the metadata can be accessed and downloaded through an http query.

The OAI – protocol is a standard for metadata exchange. By conforming to this standard the metadata is available for other services.

Harvesters are using this protocol to collect data and make it available through a search engine. New search engines are developed specifically for OAI-compliant websites, e.g. OAIster. Google Scholar uses the harvester technique to collect data in repositories. Specific content based harvesters are also being developed. [Avano](#), by Ifremer, is an experimental harvester for the marine and aquatic sciences.

## 2. PREVIOUS EXPERIENCES

### 2.1.1 ODINPUBAFRICA – OCEANDOCS (See [Final Report](#))

The goal of this two year project (2004-2006) was to develop an electronic platform to collect scientific documents (articles, conference papers, working papers, ...) produced by members of African research institutes in the field of oceanography and marine science. The Hasselt

University Library (Belgium) coordinated the project and is still strongly involved in the follow up project OceanDocs.

The results of the project are:

- The ODINPubAfrica project has accomplished its objective to establish a central repository of electronic publications. The repository contains 1122 documents related to marine science and oceanography prepared by African authors or authors affiliated with African institutions;
- The ODINPubAfrica project has trained 15 information professionals in Africa and two regional coordinators (one for East and one for West Africa);
- The ODINPubAfrica project has been recognized as a valuable example for the development of electronic repositories in developing countries. Similar repositories are now planned for Latin America and Eastern Europe (ECET). In addition IOC/IODE decided to place all e-repository projects under a new umbrella called OceanDocs;
- The ODINPubAfrica electronic repository is harvested by Google Scholar, providing global exposure of publications by African scientists;
- OdinpAfrica developed specific collections for existing paper journals. The latest release of the OceanDocs Dspace support the creation of e-journals.

But also

- research scientists are unfamiliar with the e-repository concept and see it as a competitor with “traditional journals”. It has been challenging for the African marine information managers to convince scientists;
- entering input into the e-repository is an extra and unrewarded task for the librarians, while they often need to enter the same information in other metadata systems (eg local bibliographic database, ASFA,...);
- Internet connectivity remains a problem in some locations.

OceanDocs is an extension of OdinpAfrica. The software has been upgraded. New features have been developed, but mainly it is a strategic choice to use one platform for different Odinp communities. From the user point of view it is better to have a one point access to the OceanDocs publications instead of being confronted with different repositories..



Fig. 1. The OceanDocs webpage

We note also that in 2007, the project “Development of an “easy-to-install” deployment package of an electronic repository of marine science publications (e-repository in a

**box)**” is being implemented (Project Document: See Annex I). This project, funded by the Government of Flanders within the FUST framework has the following objectives:

- (i) create a turnkey e-repository system based upon the DSpace e-repository application, customized for marine science collections and including clear instructions, standards and guidelines for deployment;
- (ii) train a few core experts from developing country regions in the deployment of the turnkey system focusing on “train the trainer” methodology;
- (iii) provide translations of manuals and guidelines in French, Spanish and Russian to promote adoption of the system in different regions

The results of this project will contribute to OceanDocs.

## 2.2 LESSONS LEARNED FROM THE ODINPUBAFRICA PROJECT

The OdinPubAfrica project makes it clear that there is a need for a dedicated OceanDocs repository, which we can give the name OceanDocs Central. Many institutes do not have the capacity to set up their own repository, surely institutes in developing countries.

But at the same time it is clear that other institutes will have their own repository or be part of a university repository (or other larger unit). Therefore IOC needs to develop a dual strategy: supporting institutes in developing countries with the central repository and developing a network of institutional repositories that are compliant with the OceanDocs metadata standard.

The OAI protocol makes an exchange of metadata possible, but only defines a basic metadata standard, Dublin Core unqualified. For OceanDocs, a metadata structure based on DC qualified compatible with the Agris AP and MODS has been worked out. We propose this as a standard for the OceanDocs network.

OceanDocs will become a distributed network with OceanDocs Central on one side and institutional repositories on the other side. The user is interested in a unique access point to the publication. A OceanDocs harvester will realize that. Can we integrate the harvester function in OceanDocs Central or will it be necessary to set up a harvester besides the repository. This has to be worked out. We prefer the integrated approach.

The development of a repository network is partly a technical task, but the organizational and motivational aspects are at least as important. On every level it is important to develop a policy: on the OceanDocs network, on the participating Odin groups and finally on the institute level. The policy will contain information about responsibilities, collection definition, submission procedures, quality control, copyright, ...

Capacity building is necessary to support the development of the repository network technically and organizationally. The target group is the group of information managers and librarians. The idea of repositories for scientific publications is rather new. New ideas have to be introduced through an information and promotion campaign, not only on the level of the information services but also on the level of the institute management and directly to the researchers.

A repository is in the first place a platform to make scientific documents available. The accessibility of the metadata, through OAI (and other protocols) creates the possibility to develop services. The most important one is the searchability through harvesters. Data can easily be reused in other formats, as a publication list of an author in OceanExpert, as an e-journal page integrated in the institute website, etc. The development of services makes the content of the repository(ies) more visible and therefore more attractive for researchers.

## 3. OTHER PROJECTS

Many new initiatives are developed under the umbrella of Open Access and OAI. Specifically in the field of aquatic and marine science some projects have been announced or are starting up. IAMSILIC proposes the Aquatic Commons project. Ifremer set up Avano and FAO is also working out projects to integrate repository metadata in their AGROVOC system. It is in the interest of the different actors to work together where it is possible.

### 3.1 AQUATIC COMMONS

The Aquatic Commons repository was established by the International Association of Aquatic and Marine Science Libraries and Information Centers (IAMSILIC) in 2006. It is intended to facilitate the exchange of scientific research related to the marine/aquatic environments by providing a searchable, Web accessible repository for digital documents. Initially, it will seek to build content based on born digital and legacy documents in PDF format that have been produced under the auspices of IAMSILIC member organizations or partnered agencies. It is specifically mandated to offer repository services where local, stable IT support is lacking. Future development will explore the serving of additional formats and alphabets. The Aquatic Commons is intended to complement institutional repositories and to collaborate with related subject repositories particularly in developing countries, such as the OceanDocs program of the Intergovernmental Oceanographic Commission.

A long-term goal will be to assist in providing access to legacy collections that have never been easily accessible to researchers and to provide access to the grey literature produced as the technical series of smaller research units in universities, governmental and non-governmental agencies.

The repository runs on the EPrints open access software created at the University of Southampton and is managed by contract with the Florida Center for Library Automation, Gainesville, Florida. The repository is managed for IAMSILIC by an Aquatic Commons Board appointed by the IAMSILIC Executive Board.

The Aquatic Commons repository is part of a larger Aquatic Commons initiative intended to serve as an IAMSILIC infrastructure for resource sharing.

### 3.2 AVANO

Avano is an OAI harvester for Marine and Aquatic Sciences. Therefore, it collects bibliographical data of electronic resources (documentation, images, datasets...) available in a group of Open Archives via the OAI-PMH protocol in order to aggregate them into a centralised database. Its Web interface offers centralised viewing of resources disseminated throughout several servers.

Avano harvests many archives from Marine Science research institutes. All resources stored in those specialised Marine Science archives are systematically and automatically referenced in Avano.

Avano also interrogates a group of open archives not specialised in Marine Science in which are stored, among others, a group of resources linked to Marine and Aquatic resources.

To process archives which are not perfectly categorised within our fields of interest, Avano uploads all of their records in a temporary database. Those databases are indexed and an automatic system isolates records that contain one or several terms linked to Marine or Aquatic Sciences.

Records spotted by this key-word system are then manually validated by librarians before they can be visible via Avano. To validate those records, librarians use a Web site. Key-words found in records are highlighted. This system allows librarians to reject index files when key-words are not related to our field of interest (for example when Fish is used for Fluorescence in situ hybridization).

## 4. TARGETS OF THE OCEANDocs PROGRAM

#### 4.1 INFORMATION PROGRAM AND CAPACITY BUILDING

##### - **Training session of trainers**

Already a training session is scheduled from April 23 to April 28 (funded through the project described in Annex I) where information managers of OdinCarsa, OdinECET, OdinCindio, WestPac are invited. The focus of the training is to prepare them as trainers and coordinators of OceanDocs for their region. This has to be repeated every two years to update their knowledge

##### - **Regional training sessions**

The regional coordinators organize regional training sessions to introduce OceanDocs in their region and to pass through the necessary skills.

##### - **Development of training material**

The material used during the training sessions for odinPubAfrica are made available on a webpage: [http://www.uhasselt.be/bibliotheek/demos/OdinPubAfrica\\_e.htm](http://www.uhasselt.be/bibliotheek/demos/OdinPubAfrica_e.htm). They will be integrated in OceanTeacher. They have to be upgraded constantly. More specific material needs to be developed.

##### - Development of promotional material

- o for information managers
- o for institute managers
- o for researchers

#### 4.2 SOFTWARE DEVELOPMENT AND SUPPORT

- The actual OceanDocs repository is developed by the Hasselt University Library on DSpace software, with major adaptations.
- Other software packages (EPrints, Fedora, CDSWare) are also OAI-compliant and can be used by partners for their institutional repository. But the OceanDocs program can not give support for all these different software.
- An easy-to-install version of OceanDocs-DSpace (for Scientific Linux/RedHat) is being developed by Hasselt University Library and will be available in April as DSpaceBox ((funded through the project described in Annex I)
- Repository software is changing fast. Nearly every year a major upgrade is available.
- Specific tools have to be developed for (semi-)automated submission (e.g.if metadata already exists) or to support institutes with limited internet access.
- Services are also not standard available and therefore some specific software development will be necessary: A priority seems the linking of OceanExpert and OceanDocs.
- Harvester software has to be installed, integrating harvested metadata in OceanDocs Central.

#### 4.3 DEVELOPMENT OF PARTNERSHIP WITH THE ODIN GROUPS AND THEIR MEMBERS

- Development of policies on the different levels:
  - o OceanDocs Network – Odin - Institute
- Creation of communities and collections in OceanDocs Central:
  - o An institute who prefers not to set up their own repository can use the OceanDocs Central infrastructure to create their homepage for their institute and identify the major collections. It is always possible later on to copy/merge the collections and their documents to a local repository
- Setting up OceanDocs compatible repositories
  - o Other institutes are invited to set up their own repository which can be harvested and where the metadata can be integrated in OceanDocs Central.

- They must be OAI compliant and support extra metadata schemes. We propose Agris AP, with specific fields (e.g. ASFA Keywords, Spatial coverage).
  - Coordination and organization of the network
    - What can be done on lowest level has to be done there. But still there is a lot of work on higher levels: access right control – quality control – synchronisation of the different policy levels - ...
- 4.4 INTEGRATION AND COOPERATION WITH OTHER INFORMATION PRODUCTS
- OceanExpert
    - OceanExpert can be used as an authority list for author input. With an OceanExpert unique ID authors and publications will be linked correctly over different repositories and websites. Then we can integrate publication list in OceanExpert.
  - Other databases and portals
    - Institute websites: integration of collections in the regular webpages
    - OceanPortal and other regional portals: availability of data in these portals makes the publications more visible
    - External: ASFA - AGROVOC
- 4.5 HARDWARE NEEDS
- The hardware needs are limited, because OceanDocs is installed on a new dedicated server since 2006. It will be necessary to replace it every 5 à 6 years. The back-up capacity needs to grow in parallel with the growth of OceanDocs Central.

## 5. WORK PLAN AND BUDGET

### 2008

- Creation of a task force
- Identifying Oceanographic repositories or repositories with oceanographic related collections
- Development of the OceanDocs Network
  - New Odin communities on the central server
  - Creation of institutional repositories in the OceanDocs Network
  - Promotional activities targeted to the information services/managers of the IODE linked institutes
  - Development of repository policy on the different levels: OceanDocs – Odin – Institute
  - Training of trainers (end 2008)
- Integration with OceanExpert:
  - OceanExpert as authority list for OceanDocs and related repositories
  - Integration of publication list in OceanExpert
  - Software development to realize this integration

### Expected results:

- Commitment of Odin groups + a policy document
- Set up of minimal 1 experimental repository per Odin group
- OceanDocs Central contains 2500 documents

### 2009

- Development of the OceanDocs Network

- Institutes of participating Odins create their institutional repositories or collections on OceanDocs central
- Development of the OceanDocs harvester function
- Promotional activities targeted to the institute management and the researchers
- Quality control on the submitted metadata: OceanDocs metadata structure will have changed as a result of the changes and new services available (e.g. integration with OceanExpert)
- Training on the Odin level

**Expected results:**

- 50 institutes will have their Inst. Repo. or their collections on OceanDocs central
- OceanDocs central contains 2500 documents
- OceanDocs harvesting will contain 10000 docs
- Visitor statistics will be available

**Long-term: 2010-13**

- Development of the OceanDocs Network
  - IODE national centers will invite the marine and oceanographic institutes in their country to participate in OceanDocs
  - Two-yearly update training session for coordinators
- Development of services based on OceanDocs:
  - Publication lists for CV – research group presentations - ...
  - Journal – conference pages
- Development of specific submission tools for OceanDocs-Dspace software:
  - ingest of existing metadata (Catalogue – ASFA-ISIS)
  - submission tool for institutes with limited internet connection
  - integration of web services for linking to different thesauri (developed by FAO)
- Integration with other IODE Information tools
  - OceanPortal

**BUDGET (in US\$)**

	<b>2008</b>	<b>2009</b>	<b>08+09</b>
development of training materials	5,000		5,000
promotion activities and products	5,000	5,000	10,000
training of trainers	40,000		40,000
tools development	5,000		5,000
services development	10,000		10,000
harvester software installation		5,000	5,000
support for institutional repositories		10,000	10,000
quality control		5,000	5,000
<b>TOTALS</b>	<b>65,000</b>	<b>25,000</b>	<b>90,000</b>

[end]



ANNEX I

**UNESCO/Flanders funds-in-trust co-operation**  
**SMALL SCALE ACTIVITY PROPOSAL**

This proposal was prepared for the review meeting scheduled for 24 May 2006, where it was also approved. The project will start in March 2007.

**Project: Development of an “easy-to-install” deployment package of an electronic repository of marine science publications (e-repository in a box)**

BACKGROUND

The Internet has caused a dramatic change in the publishing paradigm: whereas medium to large scale publishing used to be reserved for commercial publishing companies, the Internet has created new opportunities that enable anyone to publish and/or reposit materials cheaply and quickly using the Internet. Universities and scientific institutes are taking the publishing process in their own hands. A new model of publishing is based on local collections (repositories) made accessible through the OAI<sup>1</sup>-protocol.

Scientists in developing countries traditionally found it difficult to publish their findings in international journals and this for various reasons. This often resulted in these findings being published in local or institutional reports and journals. As these were often unavailable outside the institution they were considered as “grey literature”. They were difficult to find and thus rarely cited, creating a vicious circle: scientists who were rarely cited found it difficult to publish in international journals forcing them to publish in the rarely cited grey literature.

The development of electronic repositories (e-repositories) empowers scientists in developing countries to break this vicious circle: already research has shown that e-repositories improve the citation frequency of publications. In addition rare materials are now finding their way to the global science community thereby exposing a huge amount of previously unknown research.

As a complement to ODINAFRICA the project “Development of an African repository for electronic publications: was successfully submitted for funding to the FUST in 2003 (project 513RAF2004).

The creation of this repository made documents more easily accessible to the African Ocean (and coastal area) research and management community, which enhanced the internal scientific communication; and facilitated publishing of research findings by African scientists (e-journal as well as e-archive) thereby promoting African research and increasing access by African scientists to the international research forum.

The project focused on equipment, training and input. The project implementation has resulted in valuable experience and expertise that is now being called upon in other regions: Caribbean & South America, Indian Ocean, Eastern Europe are just a few regions that have expressed strong interest in developing their own e-repositories. However, as DSpace is an open source application without commercial support the deployment and finetuning of the software application is challenging. This will hamper the effective implementation of e-

---

<sup>1</sup> Open Archives Initiative ([http:// www.openarchives.org](http://www.openarchives.org) )

repositories in developing regions. It is therefore recommended to provide a turnkey system that can be easily installed and for which easy to use manuals and guidelines are available.

### OBJECTIVES AND ACTIVITIES

Taking into consideration the above observations the proposed small scale activity will focus on the following objectives:

- (iv) create a turnkey e-repository system based upon the DSpace e-repository application, customized for marine science collections and including clear instructions, standards and guidelines for deployment;
- (v) train a few core experts from developing country regions in the deployment of the turnkey system focusing on “train the trainer” methodology;
- (vi) provide translations of manuals and guidelines in French, Spanish and Russian to promote adoption of the system in different regions

### PARTNERS

In view of their extensive experience in the deployment of e-repository systems using DSpace the Hasselt University, Diepenbeek, Belgium will be the expert Partner in this project.

### TIMELINE

This project is intended to have a duration of 12 months (June 2006 – June 2007)

### BUDGET

The requested budget is USD 33,000

Activity	Budget
Development of turnkey e-repository package	US\$ 10,000
Organization of training course for key trainers	US\$ 20,000
Overhead (10%)	US\$ 3,000
<b>TOTAL</b>	<b>US\$ 33,000</b>

[end]