

# FEDORA @ AWI

Fedora User Meeting  
Copenhagen, Denmark  
28 September, 2005

**Ana Macario, Computer Center**  
Alfred Wegener Institute for Polar and Marine Research  
Germany

# Overview

- AWI and its research scope
- SOA at AWI
- Rationale for choosing FEDORA
- Long-term issues



# About AWI

- *1980*  
Establishment of the institute in Bremerhaven as a foundation under public law; AWI is one out 15 centers belonging to Helmholtz Society
- *To date*
  - Budget: 103 Mill. Euro
  - 800 Employees
- *Funding*
  - 90% Federal Ministry of Education and Research (BMBF)
  - 8% Bremen state
  - 1% Brandenburg and Schleswig-Holstein states
  - external funds



-3-

Ana Macario, Computer Center  
Alfred Wegener Institute, Bremerhaven, Germany  
European Fedora User Meeting, Copenhagen, Denmark, 2005-09-28



# Our mission



**To contribute to polar and marine research in order to advance insights into the changeability of the global environment and the earth system**

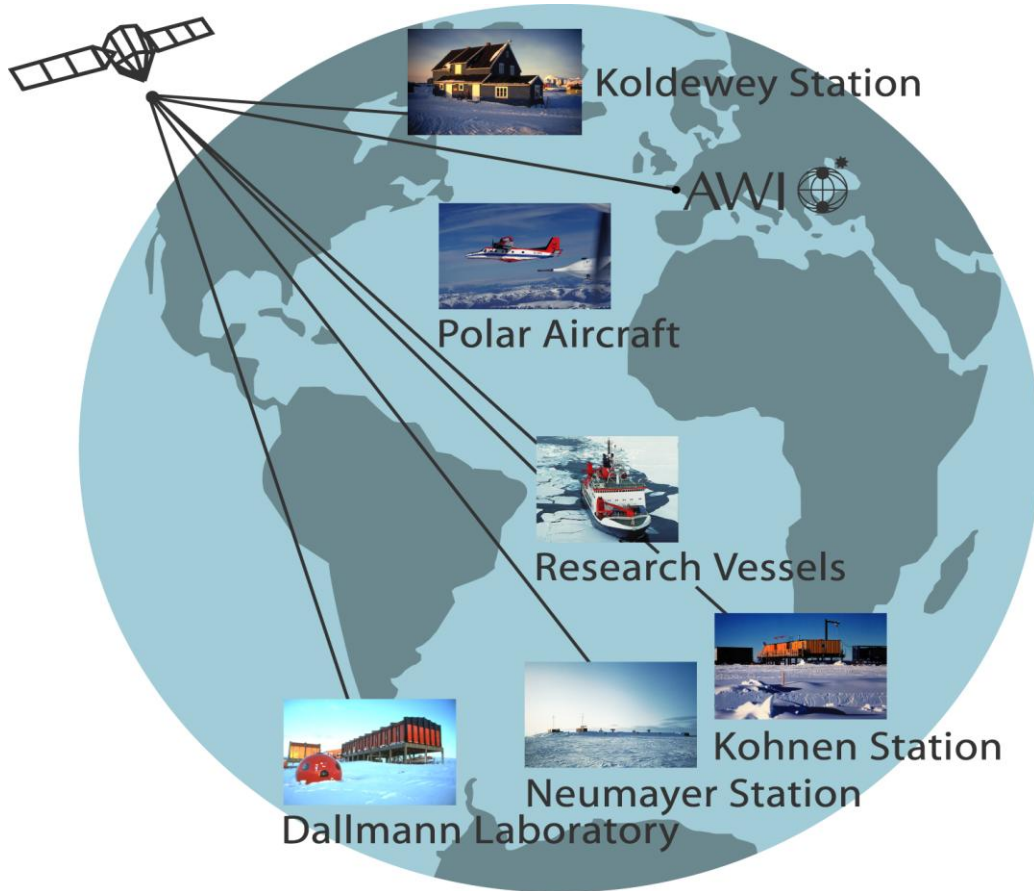


-4-

*Ana Macario, Computer Center  
Alfred Wegener Institute, Bremerhaven, Germany  
European Fedora User Meeting, Copenhagen, Denmark, 2005-09-28*



# Research platforms



## Primary data:

- observations acquired in diverse research platforms, long-time series monitoring (observatories)
- numerical models
- lab. experiments
- photographs, maps/charts

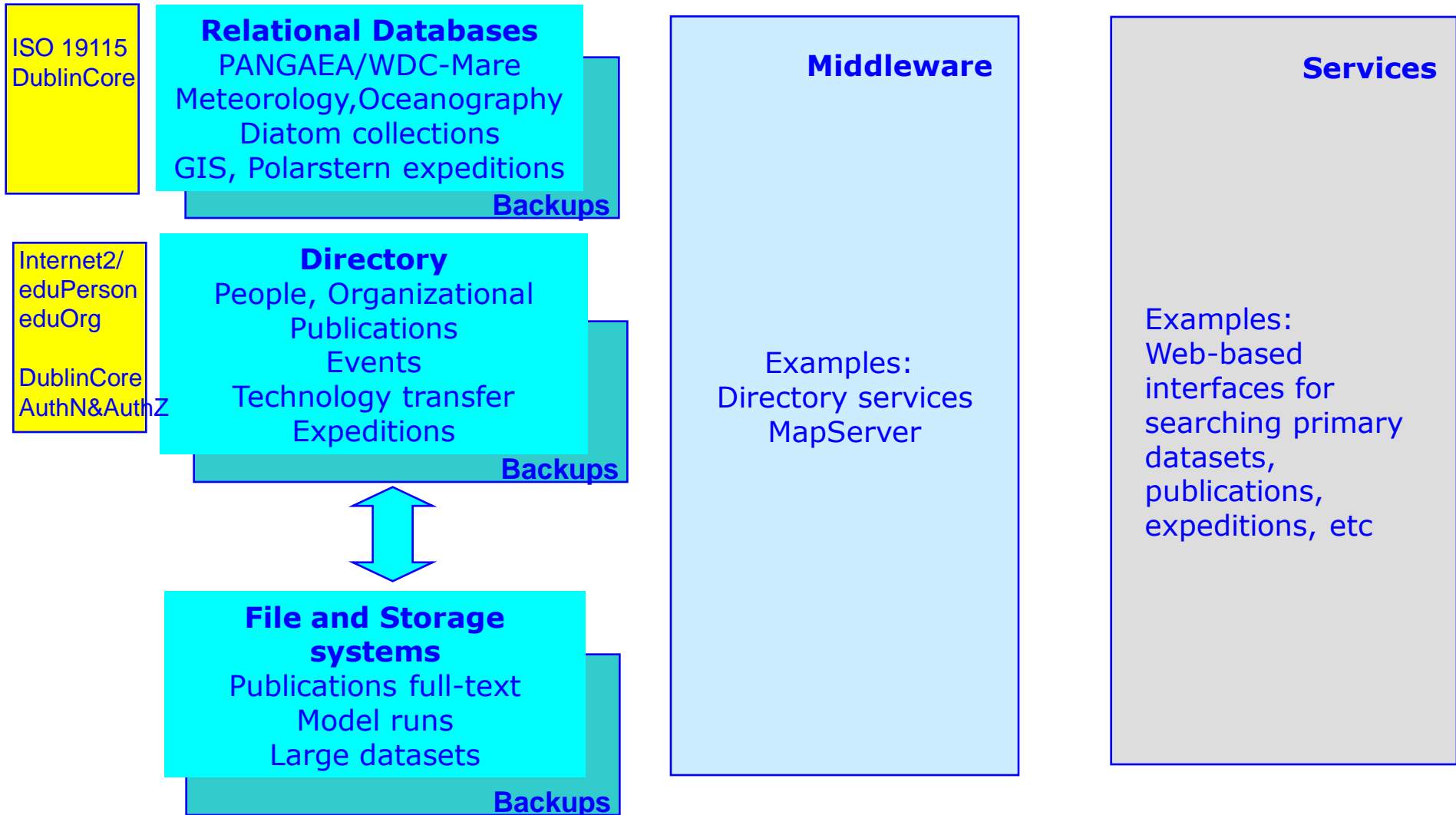
## Publications

## Events

Intellectual property rights –  
Technology transfer



# Simplified Overview (2004)



# In practice...

## "Staging"

Versioning and  
relevant to so  
calibration,  
process

Distributed

"Role"  
access policy  
rig

Spatial, te  
ther  
search/vis  
(GIS mapping services)

## "Publication"

al of quality-  
objects in IR

AI-PMH and

onality to  
ncies (GCMD,  
GBIF, etc)

Fedora  
as "active workspace"



# Why AWI chose to test FEDORA?

- Flexible, extensible digital object model
- Open source; good documentation and tutorials
- Allows for metadata description other than Dublin Core record; relevant for geo-referenced objects (ISO 19115), bio-diversity objects (Darwin Core), objects of type people (Internet2/eduPerson), organizational units (Internet2/eduOrg),etc
- Able to distribute load and object storage among several IR instances („Virtual Repository“ concept)
- Standards compliant: XML storage, OAI-PMH and web services



-8-

*Ana Macario, Computer Center  
Alfred Wegener Institute, Bremerhaven, Germany  
European Fedora User Meeting, Copenhagen, Denmark, 2005-09-28*





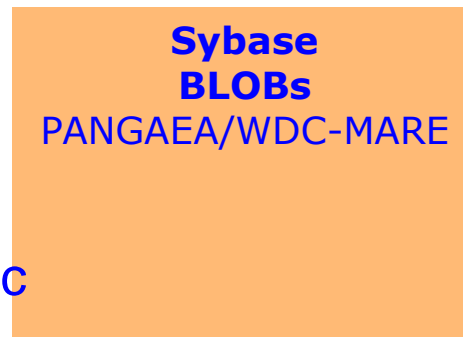
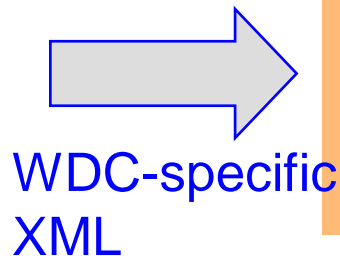
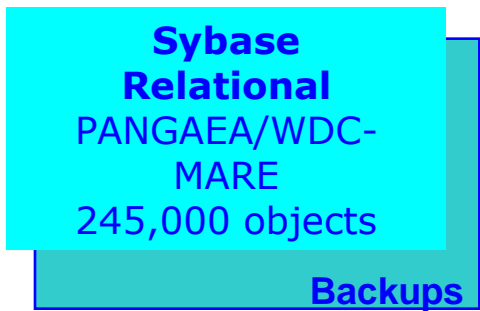
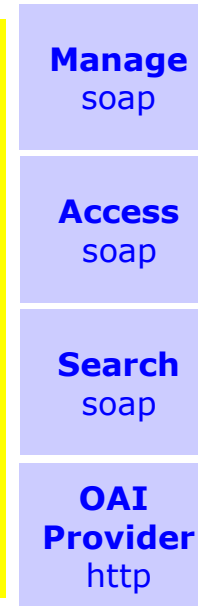
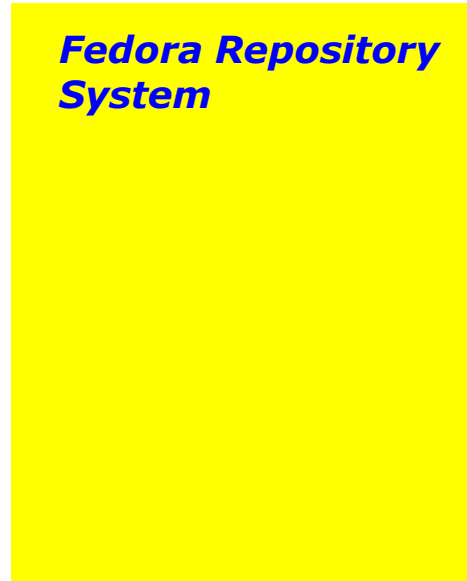
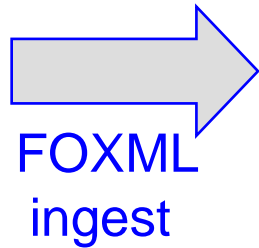
# Why AWI chose to test FEDORA? – *cont.*

- Promising scalability; Fedora@AWI currently archives 15,000 objects
- Object preservation through content versioning; includes audit trail record for preserving event history
- XML ingest/export assures interoperability with existing in house information systems

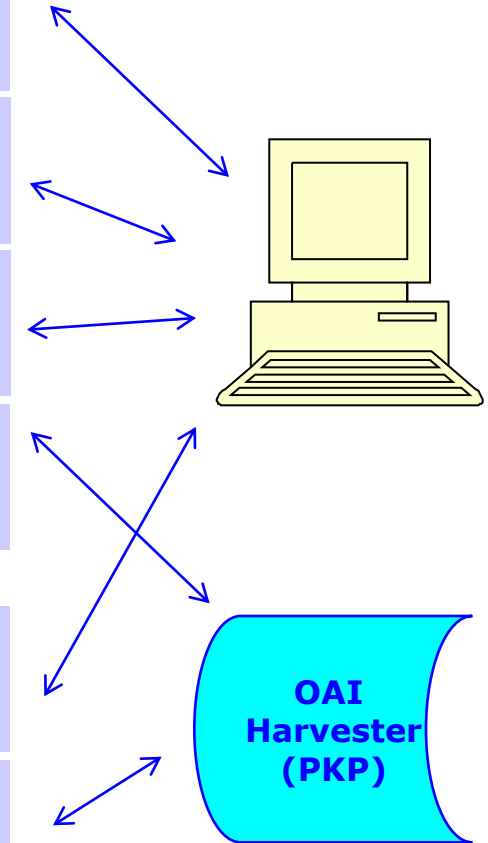


# Simplified Overview (2005)

## Backend



## Frontend



# SOAP client



Welcome to **Helmholtz Web Services** for primary data, publications and personal portfolio.

Search:

Select Repository:

- Fedora at AWI
- Pangaea

**Search**

---

Results 1-5 for '**macario**':


[Show Results 6-10 ->>](#)

- 1** [Personal Homepage of Dr. Ana Macario](#) [text, people]  
(2005) Ana Macario
- 2** [A Discovery Service for Knowledge Related to Research Platforms](#) [event, international talk]  
(2004) Macario, A.; Pfeiffenberger, H.; Reinke, M.
- 3** [Portal for Earth Sciences in Polar Regions](#) [event, international talk]  
(2003) König-Langlo, G.; Macario, A.; Olbers, D.; Pfeiffenberger, H.; Reinke, M.; Thiede, J.
- 4** [Research platforms in polar regions - a portal approach](#) [event, international talk]  
(2003) König-Langlo, G.; Macario, A.; Olbers, D.; Pfeiffenberger, H.; Reinke, M.; Thiede, J.
- 5** [An homogeneous Directory of People, Publications, and other Resources as a means for IT-based Knowledge Management in Science](#) [event, invited national talk]  
(2000) Macario, A.; Pfeiffenberger, H.


Fedora at AWI Response Time: **0.144s**, 5 Results  
Pangaea Response Time: **0.374s**, 0 Results

[Show Results 6-10 ->>](#)

Script Time: **0.399s**



# SOAP client – cont.






## OPEN ACCESS

**Record Details**

<b>Title</b>	<b>A Discovery Service for Knowledge Related to Research Platforms</b>
<b>Author(s)</b>	<i>Macario, A.; Pfeiffenberger, H.; Reinke, M.</i>
<b>Date</b>	2004
<b>Description</b>	<a href="http://www.awi-bremerhaven.de/php/PublicationAbstracts/abstract.php">uri:http://www.awi-bremerhaven.de/php/PublicationAbstracts/abstract.php</a>
<b>Subject</b>	Information Center; Computer Center; Services; Computing and Data Centre; Koldewey, Kohnen, Filchner, Polarstern, Hausgarten, Polar 2 and Polar 4, ARK, ANT
<b>Publisher</b>	XXVIII SCAR Open Science Conference, 25-31 July 2004, Bremen, Germany., 2004
<b>Type</b>	event, international talk
<b>Rights</b>	<a href="http://creativecommons.org/licenses/by-nc-nd/2.0/legalcode">uri:http://creativecommons.org/licenses/by-nc-nd/2.0/legalcode</a>

Response Time: **0.169s**, 1 Result

| [start](#) | [links](#) | [about](#) | [help](#) |

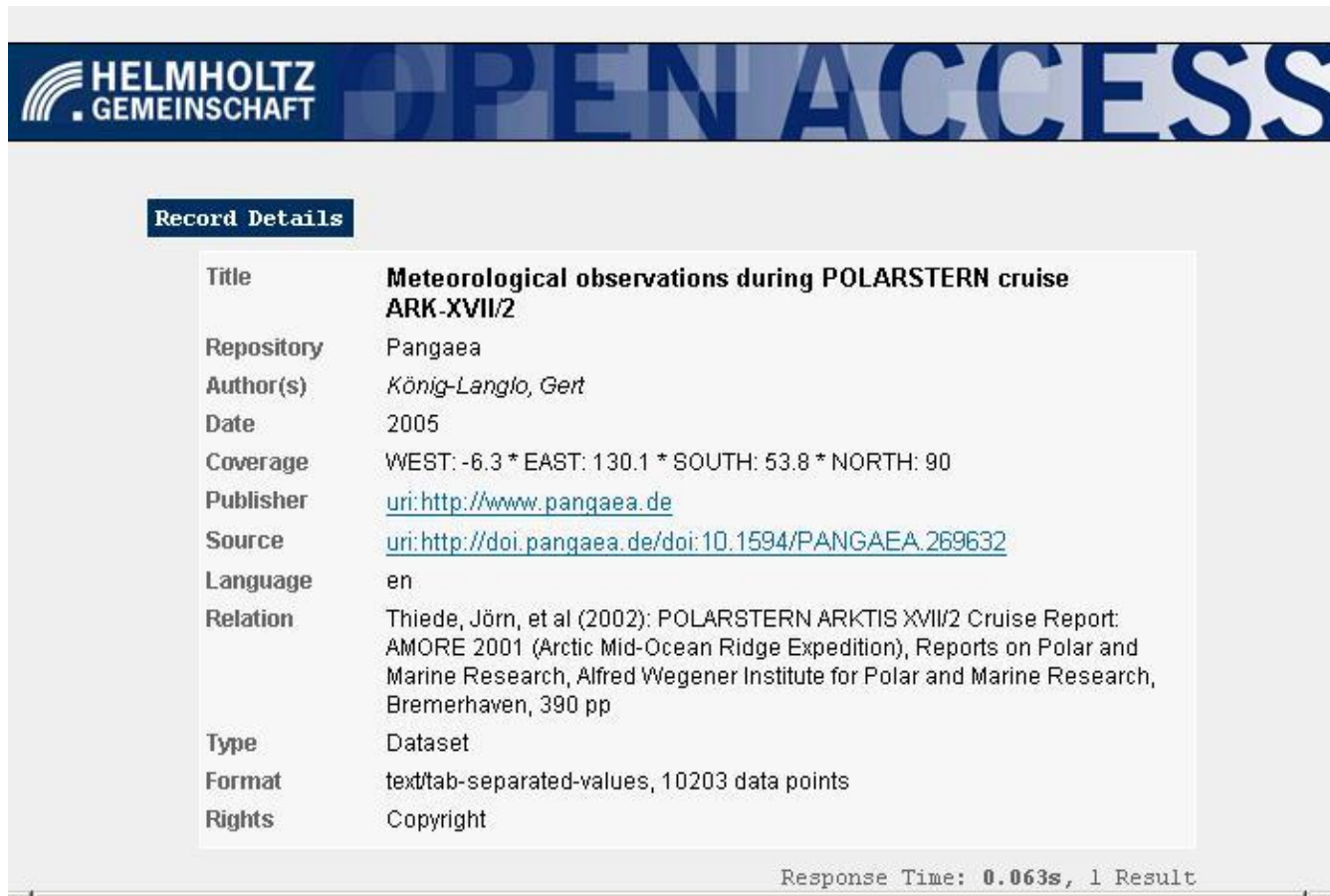
  

-12-

*Ana Macario, Computer Center  
Alfred Wegener Institute, Bremerhaven, Germany  
European Fedora User Meeting, Copenhagen, Denmark, 2005-09-28*



# SOAP client – cont.



The screenshot displays a web interface with a dark blue header. On the left, the logo for 'HELMHOLTZ GEMEINSCHAFT' is visible. To the right, the words 'OPEN ACCESS' are written in large, bold, white letters. Below the header, a tab labeled 'Record Details' is active. The main content area contains a list of metadata fields and their values. At the bottom right of the content area, it says 'Response Time: 0.063s, 1 Result'.

<b>Title</b>	<b>Meteorological observations during POLARSTERN cruise ARK-XVII/2</b>
<b>Repository</b>	Pangaea
<b>Author(s)</b>	König-Langlo, Gert
<b>Date</b>	2005
<b>Coverage</b>	WEST: -6.3 * EAST: 130.1 * SOUTH: 53.8 * NORTH: 90
<b>Publisher</b>	<a href="http://www.pangaea.de">uri:http://www.pangaea.de</a>
<b>Source</b>	<a href="http://doi.pangaea.de/doi:10.1594/PANGAEA.269632">uri:http://doi.pangaea.de/doi:10.1594/PANGAEA.269632</a>
<b>Language</b>	en
<b>Relation</b>	Thiede, Jörn, et al (2002): POLARSTERN ARKTIS XVII/2 Cruise Report: AMORE 2001 (Arctic Mid-Ocean Ridge Expedition), Reports on Polar and Marine Research, Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, 390 pp
<b>Type</b>	Dataset
<b>Format</b>	text/tab-separated-values, 10203 data points
<b>Rights</b>	Copyright

Response Time: 0.063s, 1 Result

-13-

Ana Macario, Computer Center  
Alfred Wegener Institute, Bremerhaven, Germany  
European Fedora User Meeting, Copenhagen, Denmark, 2005-09-28



# A few technical remarks on Fedora 2.0...

- Web services APIs are great; suggested improvements:
  - findObjects: browsing list backwards is not possible yet, totalNumberOfResults is missing
  - addDatastream: file uploads: could it be done with SOAP-attachments?
- Timestamp resolution in miliseconds has raised problems in „conformance tests“ *under* [www.openarchives.org](http://www.openarchives.org)
- „DeletedRecords“ set to „Transient“ in order to allow for incremental harvesting by „modified date“

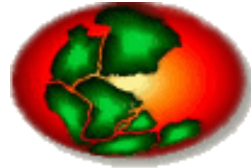


## Next steps ...

- Set up new services: naming, full-text indexing & search, large-scale content ingestion (bulk load) together with metadata
- Metadata transformation services as „disseminator“ – relevant for data supply to external service providers (e.g., NGDC, GCMD, NOAA, GBIF)
- Set up collections (and respective granularity policies) - relevant for object-to-object relationship metadata



# DC-hardwired relation



← **Resource**

Dataset-to-Publication  
relationship metadata  
should be expressed in RDF/XML  
and placed in the  
"Relations datastream"

Dublin Core

Descriptive  
metadata

Metadata

<source>  
<for content

<relation>  
<for  
<relation(s)

-16-

Ana Macario, Computer Center

Alfred Wegener Institute, Bremerhaven, Germany

European Fedora User Meeting, Copenhagen, Denmark, 2005-09-28





# Testing triple store query performance

## Backend

**Directory & File systems**  
People  
Organizational Units  
Publications  
Events  
Technology Transfer  
15,000 records  
Backups

**Sybase Relational**  
PANGAEA/WDC-MARE  
245,000 records  
Backups

FOXML ingest

2006: FOXML ingest

Pangaea-XML

**Fedora Repository System**

**We need the XACML-based module in order to add „live“ data!**

**Sybase BLOBs**  
PANGAEA/WDC-MARE

**Manage**  
http/soap

**Access**  
http/soap

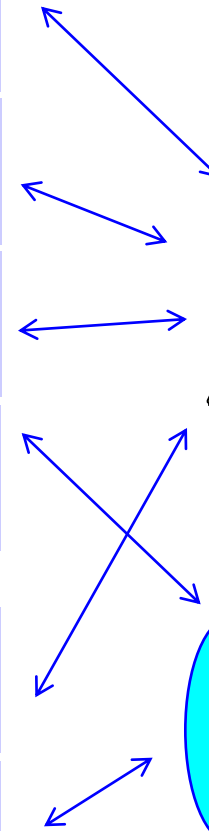
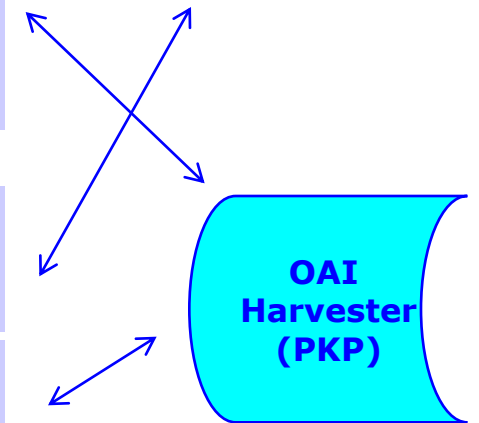
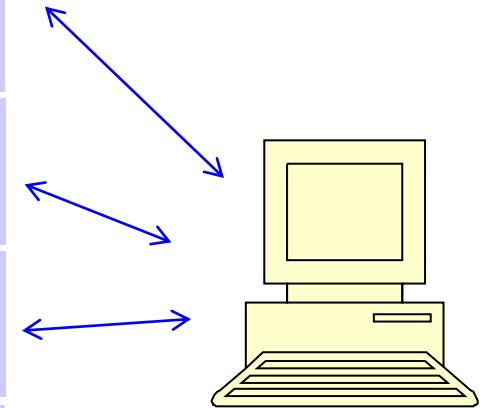
**Search**  
http/soap

**OAI Provider**  
http

**Search**  
http/soap

**OAI Provider**  
http

## Frontend



# Long-term issues for AWI

- Benchmarking for large number of files; we fear scalability breakpoint related to the size of the filesystem-based LLStorage area
- Out-of-box web-based client relevant for „acceptance“ by other Helmholtz centers
- Fine-grained access control policies and Shibboleth based AuthN – relevant in DataGRID context
- Support for sets



-18-

*Ana Macario, Computer Center  
Alfred Wegener Institute, Bremerhaven, Germany  
European Fedora User Meeting, Copenhagen, Denmark, 2005-09-28*



# Long-term issues for AWI – *cont.*

- Federation model
- Collaboration and support infra-structure
  - disseminators for specific visualizations services (e.g. NetCDF data and LiveAccessServer, GIS data and OpenMapServer); relevant for DataGRID
  - ECLIPSE project to facilitate plug-in development?
  - Google strategy
  - Seminars, tutorials for „advanced“ FEDORA users



# Thanks for your attention!

fedora-admin@awi-bremerhaven.de

<http://www.awi-bremerhaven.de>

<http://web.awi-bremerhaven.de/fedora/oai>

**Ana Macario, Computer Center**  
Alfred Wegener Institute for Polar and Marine Research  
Germany