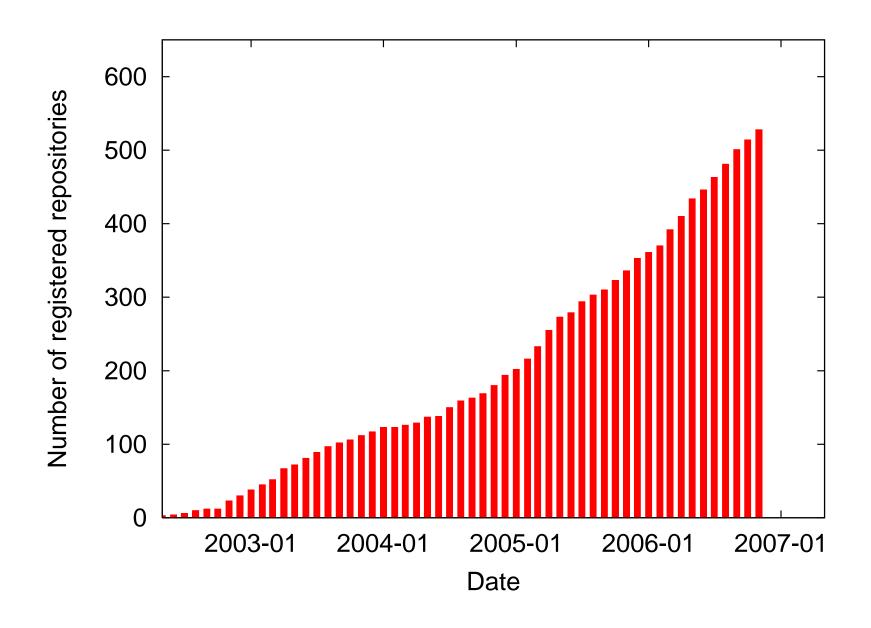
The State of the OAI-PMH

Simeon Warner

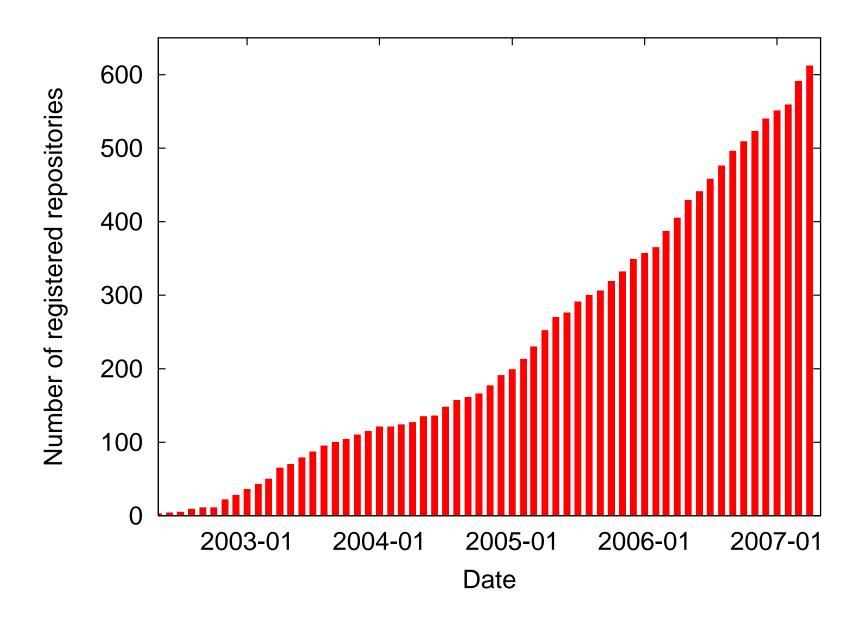
simeon@cs.cornell.edu

OAI5: Innovations in Scholarly Communication CERN, 18–20 April 2007

Registrations at OAI validator



Registrations at OAI validator



Other indicators

Registry	Repositories	
UIUC registry	1365	responding v2.0 repositories
ROAR	862	registered repositories
OAIster	769	OAI-PMH data sources
BASE	409	OAI-PMH data sources

Service	Items	
OAIster	11.3 million	
BASE	5.4 million	OAI-PMH + web
•••		
Avano	61 thousand	in aquatic sciences

Looking back: OAI1, 2001

Results from ListIdentifiers requests made to all 19 registered OAI repositories as of 2001-03-08.

Repository	Items
arXiv	155522
OCLC Theses and Dissertations	102762
NACA	6352
M.I.T. Theses	5196
The Oxford Text Archive	1290
•••	•••
CDLDERM	2
Tobacco Control Digital Repository	1

Notable milestones and developments

- <friends> and ListFriends
- oai_dc vs other metadata formats
- <branding>
- Static repository and gateway specification
- Complex object formats as "metadata"
- OAI-rights effort and specification

Rights

OAI rights effort devised schema and rules to associate rights information with OAI-PMH records. Did not propose a way to associate rights information with resources referenced by OAI-PMH records.

Use for metadata doesn't really buy much as people were sharing metadata anyway.

Use of OAI-PMH to share resources as complex objects somewhat slow to catch on but expect rights information to be more useful in this case.

Technical issues

Sets: the protocol provides no good and reliable way to indicate changes in set membership. Best approach to update datestamp to at least signal entry into set.

Character set limitations in set names have lead to a number of messy workarounds (especially the use of URL encoding).

Ties to and reliance upon HTTP has been a mixed bag. Simple and stable. More subtle flow-control would be nice.

Finally "over the hump" with XML and UTF-8 character encoding issues. Tools have caught up.

A return to scholarly focus

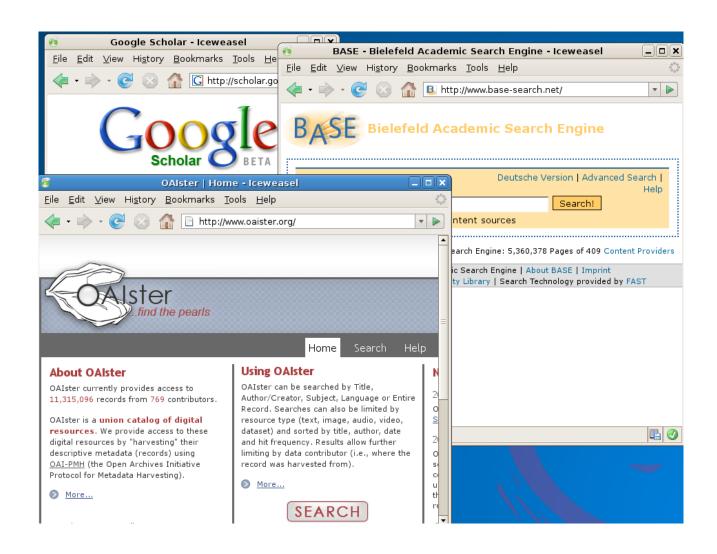
As the OAI-PMH evolved we saw the Item evolve from e-print, through document like object, to resource. The protocol is still written in that way.

The focus of activity around the OAI-PMH is really now about scholarly objects again, with additional use in museums community.

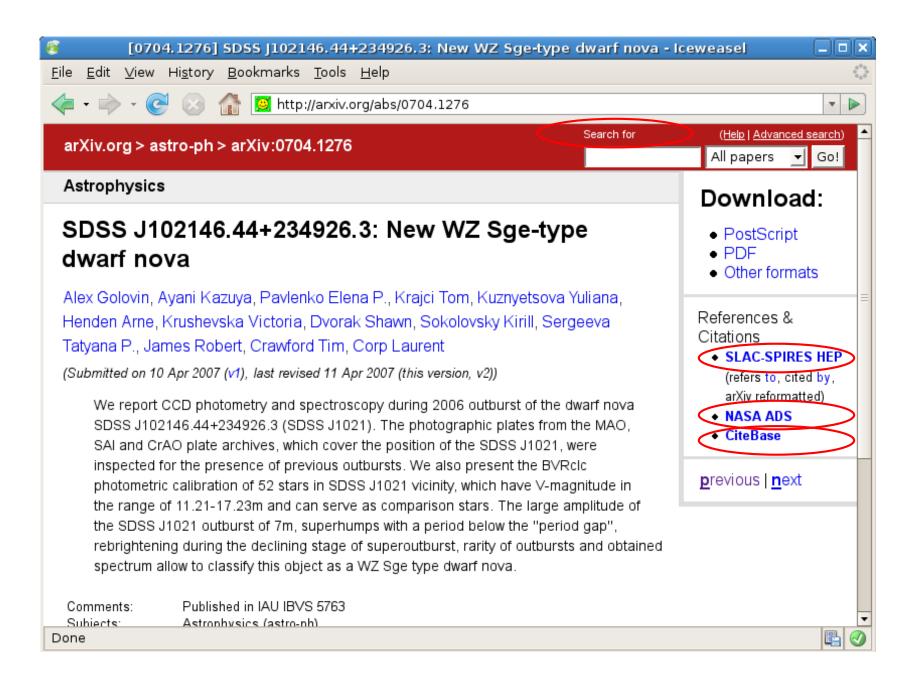
Now general awareness that scholarly objects include more than just articles — data, movies, software, etc. — so generalized notion of **Item** is appropriate.

Cross-archive search

An original motivator for OAI-PMH – necessary for **Institutional Repository** notion. Now here but also Google.

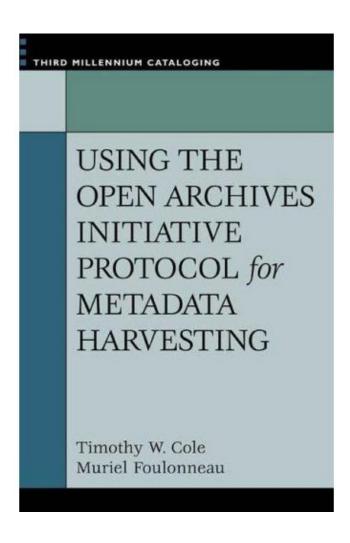


Richer services



Maturity... OAI-PMH, the book

Soon to hit the shelves from Tim Cole and Muriel Foulonneau:



Q. What now?

That's for Herbert...