

## Conferenza italiana sul Software Libero 2008

# L'archivio istituzionale Unitn-eprints

Filippo Moser

Dir.Sist.Informativi Servizi e Tecn.Inf.  
Uff.Analisi Svil.e Supporto Applicativi  
Università degli studi di Trento

Francesca Valentini

Progetto Casa editrice  
Sistema Bibliotecario di Ateneo  
Università degli studi di Trento

**Trento, 17 maggio 2008**

# Cos'è un archivio istituzionale

- permette l'**autoarchiviazione** immediata dei **risultati della ricerca** scientifica prodotta in un'università
- permette il deposito di *preprint* o, laddove possibile, di *postprint*
- permette il deposito della "letteratura grigia", di tesi e dissertazioni, brevetti, *working paper*, atti di convegni, materiale didattico ecc.

## **AUTOARCHIVIARE =**

**depositare un documento digitale** in un **archivio** pubblicamente accessibile **conforme al protocollo OAI-PMH** tramite una semplice interfaccia web ove si copiano/incollano i "**metadati**" (nome dell'autore, titolo, nome della rivista, data, ecc.) e si allega il **documento a testo pieno ("dati")**.

# L'Accesso aperto (Open Access)

- accesso **libero** e **senza barriere** al sapere scientifico
- movimento e strategie (nate all'interno del mondo accademico) il cui scopo è **riguadagnare possesso della comunicazione scientifica** offrendo **libero accesso ai risultati della ricerca**
- sfruttando le potenzialità offerte dalla rete, gli articoli vengono **gratuitamente resi accessibili** senza le restrizioni e le barriere previste dalle licenze tradizionali
- autori **mantengono i loro diritti** e **consentono in anticipo alla libera circolazione** del proprio materiale
- la comunicazione scientifica torna a essere libera e globale, favorendo lo **sviluppo anche nei paesi meno ricchi**

# Copyright

- I documenti depositati in un archivio istituzionale sono **tutelati dalla legge sul diritto d'autore** (L. 633/1941 e successive modificazioni)
- L'autore mantiene tutti i diritti morali ed economici sul proprio lavoro, se questi ultimi non sono già stati ceduti a un editore. Qualora ne abbia la facoltà, l'autore può decidere di cedere alcuni dei suoi diritti (Licenze **Creative Commons**)
- **Author addenda** per i contratti editoriali
- Progetto **SHERPA/RoMEO** per le *policies* degli editori stranieri

Ulteriori info: [http://wiki.openarchives.it/index.php/I\\_diritti\\_degli\\_autori](http://wiki.openarchives.it/index.php/I_diritti_degli_autori)

# OA: green & gold ; weak & strong

Strategie di pubblicazione in Open Access:

- **GREEN**: l'*autoarchiviazione* in **archivi aperti**: archivi digitali a carattere istituzionale o disciplinare
- **GOLD**: *pubblicazione* su **riviste ad Accesso Aperto**, che garantiscono la *peer review* ma adottano un diverso modello economico
- **WEAK** : archiviazione/pubblicazione **senza barriere di tipo economico**
- **STRONG**: archiviazione/pubblicazione **senza barriere di tipo economico né legale**

# I vantaggi: quante volte si paga la ricerca?

1

## Modello tradizionale:

1. le università finanziano le **singole ricerche**
2. i ricercatori **pubblicano** sulle riviste scientifiche
3. per accedere ai risultati della ricerca, gli atenei pagano una seconda volta sotto forma di **abbonamento** alle riviste scientifiche
4. i ricercatori, se vogliono **riusare** un loro articolo anche **per scopi didattici**, devono pagare una terza volta per i diritti

- **crisi dei prezzi: + 321% in 20 anni**

- calo sottoscrizioni nelle biblioteche
- contrazione offerta informativa
- limitata disseminazione risultati ricerca

- **crisi delle licenze:**

i contratti editoriali prevedono barriere tecnologiche e legali per la trasmissione e il riuso del materiale già pagato con gli abbonamenti

# I vantaggi: quante volte si paga la ricerca?

## 2

- pubblicare **Open Access** non significa pubblicare gratis. Cambia il modello economico: il costo non costituisce più una barriera all'accesso sotto forma di abbonamento, ma viene coperto a monte, pagando per le spese di pubblicazione.
- archivi istituzionali: i costi rientrano nei costi di gestione dell'ente (evidenti vantaggi di visibilità e prestigio)
- riviste: costi legati al *workflow* editoriale e alla garanzia di una *peer review* di qualità
- tendenza a livello internazionale: ricomprendere i costi per la pubblicazione Open Access nel budget iniziale della ricerca. Sempre maggiore è il numero di enti di finanziamento che richiedono che le ricerche finanziate con i propri fondi vengano rese pubblicamente disponibili in Open Access, impegnandosi a pagare i costi del processo editoriale.

Estratto da "<http://wiki.openarchives.it/index.php/Costi>"

# I vantaggi: l'impatto - 1

L'archiviazione tramite OAI-PMH garantisce

- più vasta **disseminazione**
- migliore **impatto**
- maggiore **visibilità**.

**Più un articolo è liberamente scaricabile,  
più è letto, più viene citato.**

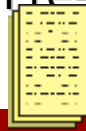
Questo favorisce la condivisione del sapere e quindi un più rapido avanzamento della conoscenza, senza barriere, in tutto il mondo.

by Tim Brody, Stevan Harnad (Southampton University)

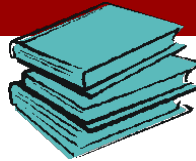
**Impact cycle begins:**  
Research is done



Researchers write pre-refereeing "Pre-Print"



Submitted to Journal



Pre-Print reviewed by Peer Experts - "Peer-Review"



Pre-Print revised by article's Authors

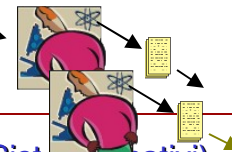
Refereed "Post-Print" Accepted, Certified, Published by Journal



Researchers can access the Post-Print **if their university has a subscription to the Journal**



**New impact cycles:** New research builds on existing



12-18 Months

# Maximized Research Access and Impact Through Self-Archiving

by Tim Brody, Stevan Harnad (Southampton University)

**Impact cycle begins:**



Researchers write pre-refereeing "Pre-Print" self-archived in University's Eprint Archive

Submitted to Journal



Pre-Print reviewed by Peer Experts - "Peer-Review" Pre-Print revised by article's Authors



Refereed "Post-Print" Accepted, Certified, Published by Journal



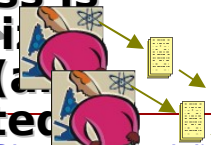
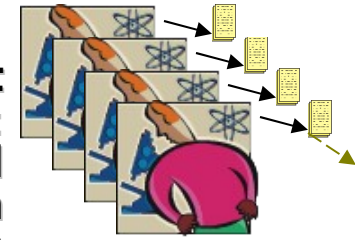
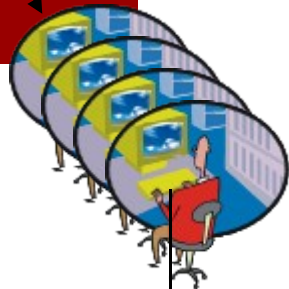
Researchers can access the Post-Print if their university has a subscription to the Journal



Post-Print is published in open-access journal or in toll-access journal and self-archived in University's Eprint Archive

**New impact cycles: Self-archived research impact is greater (and faster) because access is maximized**

**New impact cycles: New research builds on existing**



12-18 Months

10

# L'impatto come valutazione quantitativa

## ● Impact factor:

- si riferisce alla rivista e non all'articolo
- si calcola a partire dai dati raccolti negli ISI Citation Indices (**Web of Science**), “pubblicati” dall'Institute of Scientific Information (**ISI**).

ISI Web of Knowledge<sup>SM</sup>  
Journal Citation Reports<sup>®</sup>

WELCOME ? HELP

Journal Summary List  
Journals from: subject categories **ACOUSTICS** [VIEW CATEGORY SUMMARY LIST](#)

Sorted by: **Journal Title** [SORT AGAIN](#)

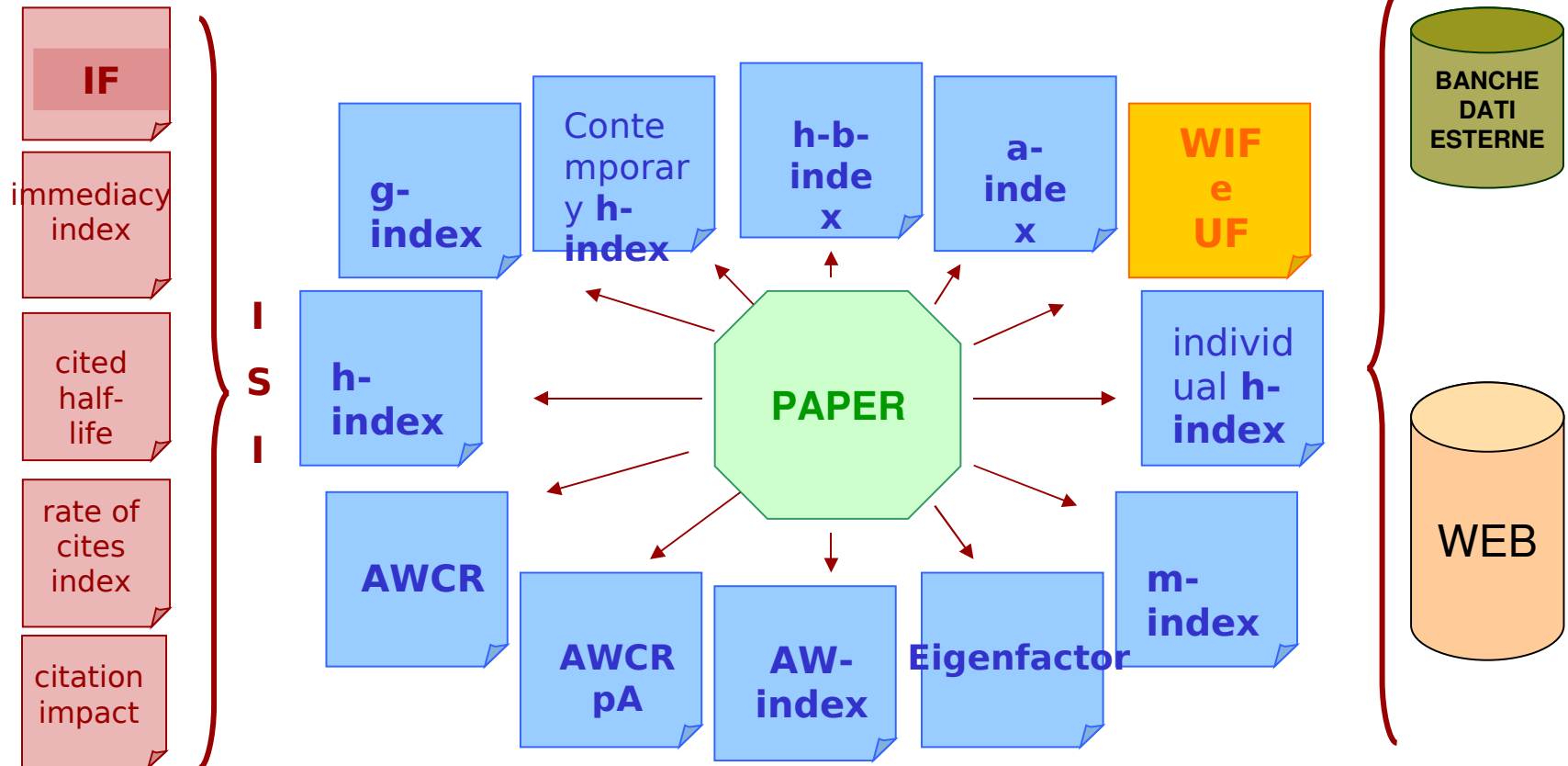
Journals 1 - 20

MARK ALL UPDATE

Mark	Rank	Abbreviated Journal Title <small>(linked to journal information)</small>
<input type="checkbox"/>	1	<a href="#">ACOUST_PHYS+</a>
<input type="checkbox"/>	2	<a href="#">ACOUST_RES_LETT_ONL</a>
<input type="checkbox"/>	3	<a href="#">ACTA_ACUST_UNITED_AC</a>

$$\text{impact factor} = \frac{\text{citations in 1997 to articles published in 1995 and 1996}}{\text{articles published in 1995 and 1996}}$$

# Al di là dell'Impact factor



# La rete Open Access: gli archivi

**986** repositories nel mondo

- **796** istituzionali (**81%**) e **122** disciplinari (**12%**)
- **476** in Europa (**48%**), **297** in America del Nord (**30%**)
- **263** in USA (**27%**), **117** Germania (**12%**), **108** UK (**11%**), ..., **28** Italia (**2,8%**)
- **24%** usa DSpace, **20%** usa Eprints
- **85%** inglese, **13%** tedesco, **6%** francese e giapponese, ....., **2%** italiano (7° posto)

Dati: **OpenDOAR** (31/10/2007)

Multidisciplinary	513	52%
Health and Medicine	77	7%
Computers and IT	76	7%
Science general	73	7%
History and Archeology	73	7%
Social sciences general	65	6%
Business and Economics	56	5%
Geography	54	5%
Law	45	4%
Biology and Biochemistry	43	4%
Physics and Astronomy	42	4%
Mathematics	39	3%
Arts and Humanities general	27	2%
Mechanical engineering and Materials	26	2%
Philosophy and Religion	25	2%
Language and Literature	23	2%

# La rete Open Access: le riviste

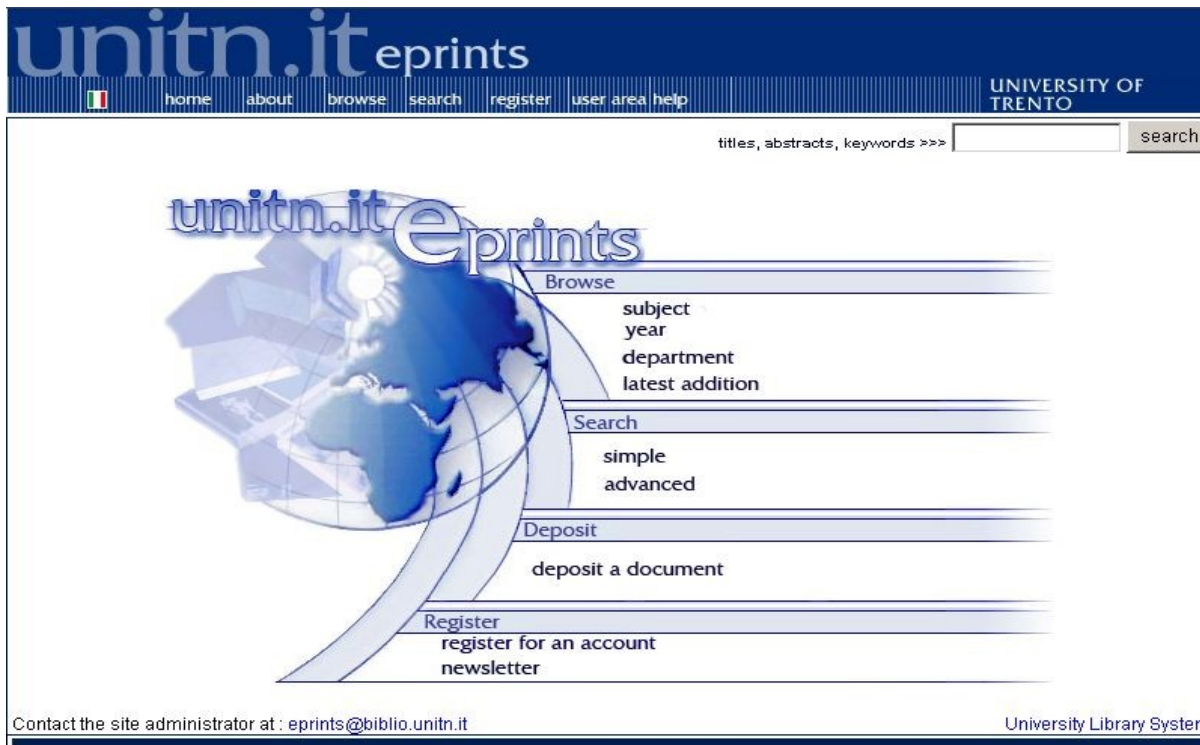
Dati : **DOAJ** (31/10/2007)

- **2897** riviste OA
- di cui circa **1305** (45%) di discipline umanistiche

## Riviste ad accesso aperto in **DOAJ** :

- hanno un **modello di finanziamento** che non pesa sui lettori o sulle istituzioni: : nessun pagamento richiesto per accedere ai testi, i costi di pubblicazione sono coperti da una quota versata dall'autore o dalla sua istituzione (la tendenza è di ricomprendere i costi di pubblicazione nel budget iniziale stanziato per la ricerca)
- offrono il diritto di **leggere, scaricare, copiare, distribuire, stampare, ricercare o collegare al testo pieno** degli articoli
- per essere incluse nella *directory*, devono garantire il **peer-review** o il **controllo di qualità editoriale** sui contenuti

# L'archivio istituzionale Unitn-eprints



The screenshot shows the homepage of the Unitn-eprints institutional archive. The header features the logo 'unitn.it eprints' and navigation links: home, about, browse, search, register, user area, help. A search bar is located in the top right corner with the text 'titles, abstracts, keywords >>>' and a 'search' button. The main content area is divided into four sections: 'Browse' with links for subject, year, department, and latest addition; 'Search' with links for simple and advanced search; 'Deposit' with a link for 'deposit a document'; and 'Register' with links for 'register for an account' and 'newsletter'. The footer contains contact information for the site administrator (eprints@biblio.unitn.it) and the University Library System logo.



UNITN-eprints is running on EPrints2 archive-creating software, which generates eprints archives that are compliant with the Open Archives Protocol for Metadata Harvesting OAI 1.1  
The eprints.org archive-creating software is available for free at <http://www.eprints.org/>



# L'archivio istituzionale Unitn-eprints

- è nato nel **2002**
- accoglie soprattutto *preprints*, *postprints*, *atti di convegno* e *tesi di dottorato* di **10** dipartimenti su 14 e di **3** centri di ricerca
- è amministrato dallo SBA nell'ambito del Progetto Casa Editrice, che comprende:
  - ✓ POLARiS Anagrafe della ricerca (<http://polaris.unitn.it/>)
  - ✓ Unitn-eprints (<http://eprints.biblio.unitn.it/>)
  - ✓ Tesi di laurea-on-line
  - ✓ Procedure editoriali e deposito obbligatorio

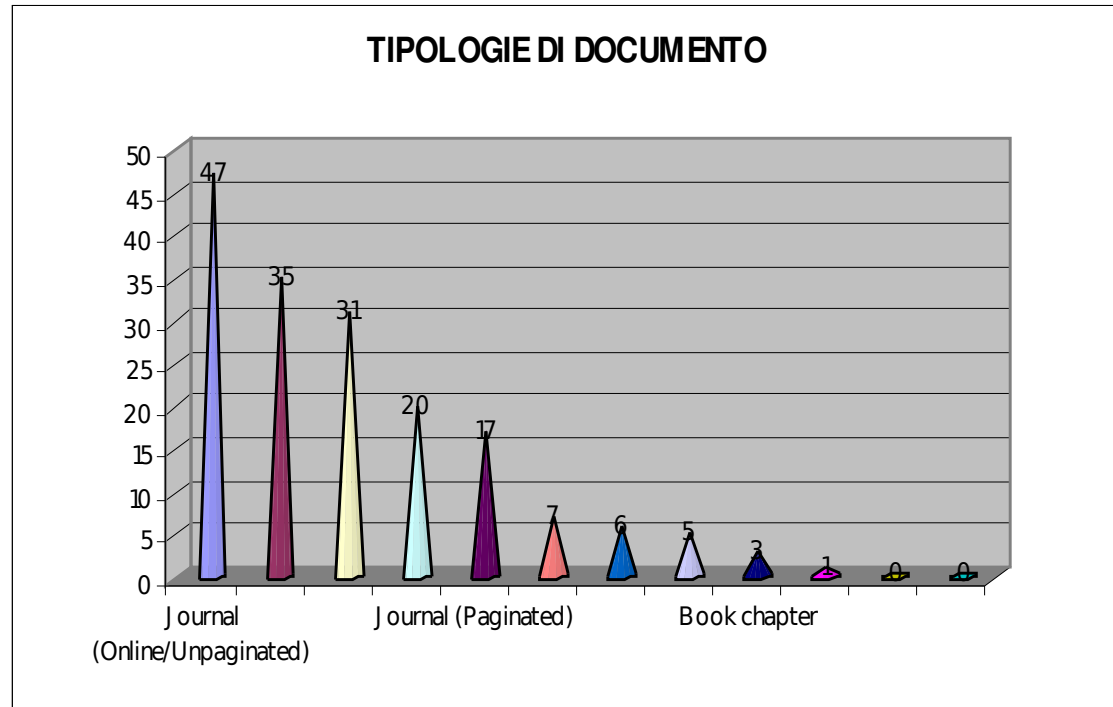
# Unitn-eprints: dati aprile 2008

- **804** utenti registrati
- **929** articoli in archivio
- **118** articoli nelle *workspace* degli autori
- **12** collane dipartimentali

Informatica e Scienza dell'Informazione	523
Matematica	112
Economia	91
CEEL (Computable and Experimental Economics Laboratory)	46
Studi Letterari, Linguistici e Filologici	36
Informatica e Studi Aziendali	33
ISSAN (Istituto Studi Sviluppo Aziendale Noprofit)	23
Scienze Giuridiche	22
Sociologia e Ricerca Sociale	20
Transcrime (Research Centre on Transnational Crime)	18
Filosofia, Storia e Beni Culturali	3
Ingegneria Civile e Ambientale	1
Ingegneria Meccanica e Strutturale	1

# Unitn-eprints - Tipologie di documento

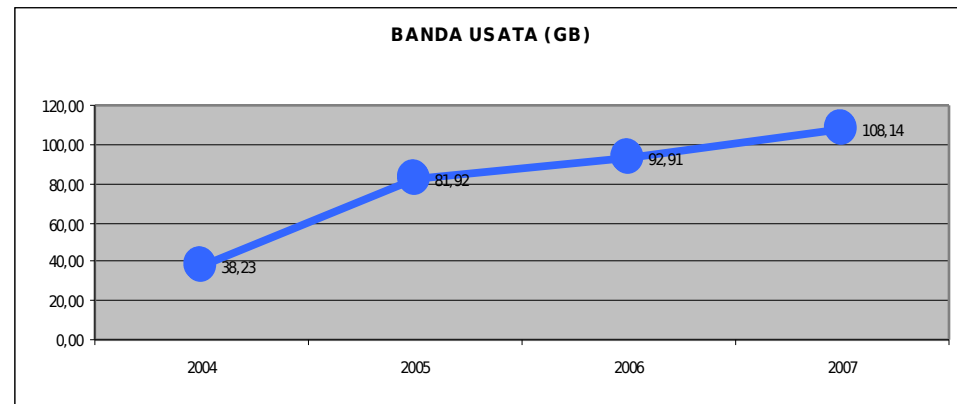
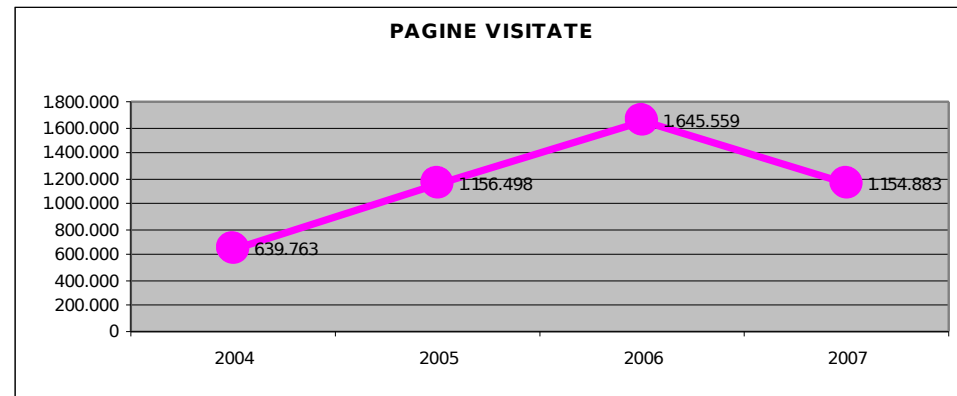
Departmental Technical Report	757
Journal (Online/Unpaginated)	47
Book	35
Thesis	31
Conference paper	20
Journal (Paginated)	17
Conference proceedings	7
Preprint	6
Other	5
Book chapter	3
InCollection	1
Conference Poster	0
Newspaper/Magazine article	0



*Dati all'aprile 2008*

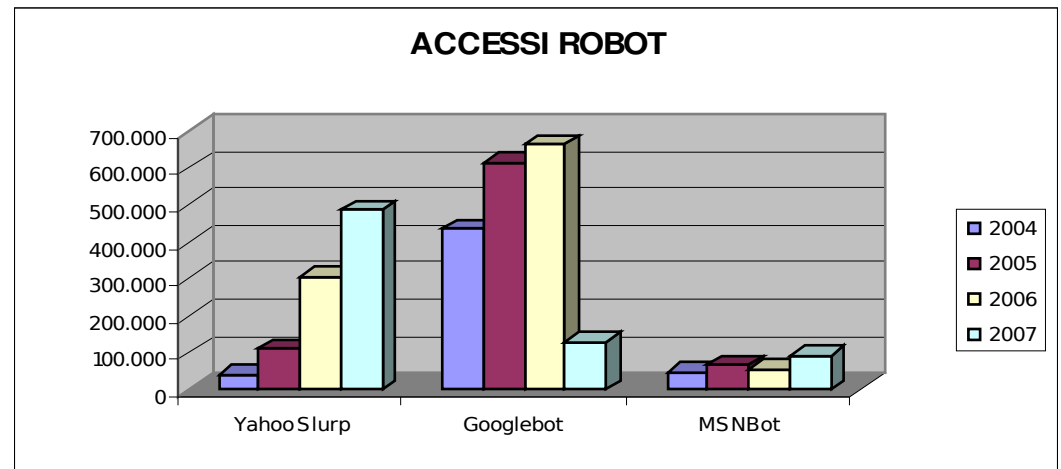
# Pagine viste e banda usata: 2004-2007

- Pagine visitate:
  - 2004 = 639.763
  - 2005 = 1.156.498
  - 2006 = 1.645.559
  - 2007 = 1.154.883
- Banda usata:
  - da 38,23 a 108,14 GB



# Accessi di spider/robot: 2004-2007

ACCESSI ROBOT	2004	2005	2006	2007
Yahoo Slurp	40.665	111.133	306.198	485.982
Googlebot	435.591	616.482	662.001	130.402
MSNBot	49.098	65.838	54.816	91.534



# Accessi da motori di ricerca: 2004-2007

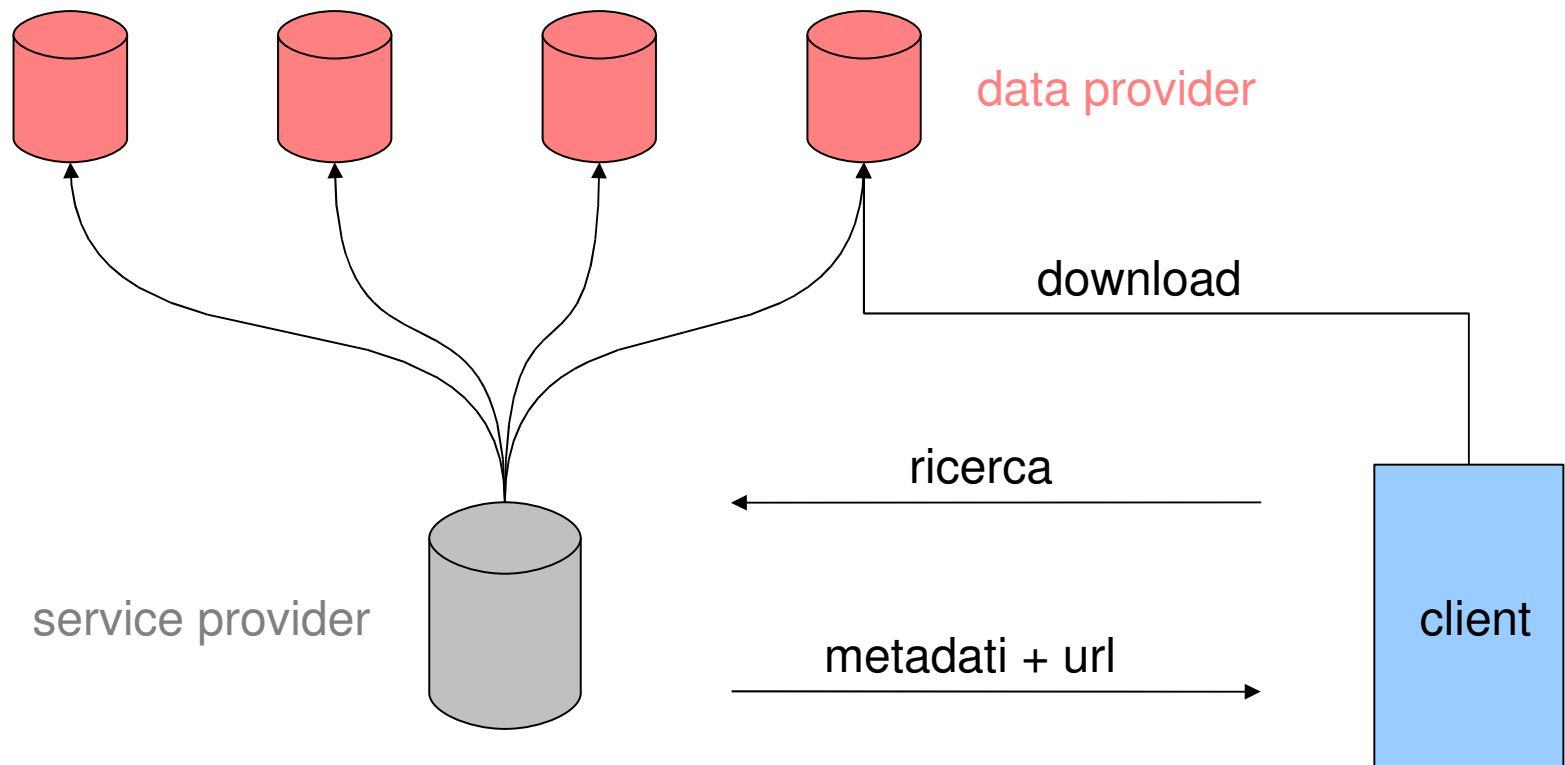
ACCESSI DA WEB O MOTORE DI RICERCA	2004	2005	2006	2007
AltaVista	665	352	187	128
Google	32.304	38.204	54.265	69.933
MSN Search	436	637	234	152
Virgilio	1.127	725	132	0
Yahoo!	3.226	4.305	2.738	1.458

ACCESSI DA WEB O MOTORE DI RICERCA	2004	2005	2006	2007
Accessi diretti o via segnalibro	393.898	477.641	337.629	322.388
Accessi da motore di ricerca	38.935	45.335	58.838	74.390

# OAI-PMH - 1

- Data Provider (es. <http://eprints.biblio.unitn.it/> )
- Harvesting
- Service Provider (es. <http://www.oaister.org/>)
- OAI-PMH “The Open Archives Initiative Protocol for Metadata Harvesting”
- Versione 1.0 del 21/01/2001
- Versione 2.0 del 14/06/2002

# OAI-PMH - 2



# OAI-PMH - 3

- Basato su HTTP (GET/POST)
- Risposte in XML
- Metadati in Dublin Core
- Non e' un protocollo di ricerca
- Identify, ListRecords, GetRecord ...

<http://eprints.brighton.ac.uk/perl/oai2?verb=Identify>

<http://eprints.biblio.unitn.it/perl/oai2?verb=ListSets>

[http://eprints.brighton.ac.uk/perl/oai2?verb=ListRecords&metadataPrefix=oai\\_dc](http://eprints.brighton.ac.uk/perl/oai2?verb=ListRecords&metadataPrefix=oai_dc)

# Eprints

## Perchè Eprints ?

- flessibile/personalizzabile
- OAI - PMH
- molto diffuso
- struttura gestionale orientata ai prodotti della ricerca
- Server LAMP
- GPL

# Eprints: custom

Minority Health Archive  
University of Pittsburgh



Glasgow ePrints Service  
University of Glasgow



UNITN-Eprints  
University of Trento



Altri esempi : <http://www.eprints.org/software/examples/>

# Singolo *e-prints*

Pagina dell' *abstract* :

- miniature (immagini, video/PDF)
- dati per Google Scholar
- richiesta per un *closed-access document*

# Ricerche

Ogni ricerca puo' essere esportata in :

- RSS
- BibTeX
- Dublin Core
- EndNote
- ...

# Tipi di deposito

Supportati molti tipi di deposito:

- Documenti
- Multimedia
- Dati scientifici
- Tech rep
- ...

tutto configurabile da file XML

# Upload

Per ogni documento posso scegliere:

- visibilità ( Open Access vs. Closed Access )
- licenza
- periodo di embargo

# User Interface

## Interfaccia utente :

- **molti miglioramenti rispetto alla versione 2** (meno click, scroll ...)
- **alert su possibili duplicazioni** (titolo)
- **completamento automatico** (nome utente, nome rivista, ISSN, editore)
- **l'autocompletamento puo' usare fonti esterne** (SHERPA/RoMEO)

# Community

- Wiki ( <http://wiki.eprints.org/> )
- Mailing-list
- Supporto comunità + pagamento
- Sito web ( <http://www.eprints.org/software/> )

# Eprints: installazione

## pacchetti per le varie distro:

- Debian/Ubuntu (.deb)
- Red Hat (.rpm)
- Fedora ( .rpm )

**altrimenti ci sono i sorgenti**

# Eprints: requisiti

## Requisiti

- Apache
- Mysql
- Perl
- Mod\_perl
- Gdome
- Pdftotext
- WvText
- Lynx
- latex and dvips

# Eprints su Debian

## Aggiungo al `/etc/apt/sources.list`

```
$> deb http://deb.eprints.org/ unstable/  
$> deb-src http://deb.eprints.org/ source/  
$> apt-get update  
$> apt-get build-dep eprints  
$> apt-get source eprints  
$> dpkg -i ../eprints-X.X.X_all.deb
```

*oppure*

```
$> ./configure --with-smtp-server=smtp.yourdomain --quello_che_mi_serve  
$> sudo ./install.pl
```

# Webografia - 1

- <http://wiki.eprints.org/w/Documentation>
- <http://www.openarchives.org/OAI/openarchivesproc>
- <http://www.oaforum.org/tutorial/english/>

# Webografia - 2

- **METAMOTORI**  
([PLEIADI](#), [NCSTRL](#), [OAIster](#), [Scirus](#), [ScientificCommons](#), ...)
- **MOTORI CITAZIONALI**  
([The citebase Search](#), [Google Scholar](#), [Citeseer](#), ...)
- **REGISTRI**  
([Openarchives.eu](#), [OPENDOar](#), [DOAJ](#), [ROAR](#), ...)
- **NEWS e WIKI**  
([Open Access Directory](#), [OAD Wiki](#), [OA-Wiki Italia](#), ...)