

CENTRE DE SUPERCOMPUTACIÓ
DE CATALUNYA



Catalan Policies and Experiences on Cooperative Repositories

Miquel Huguet
Director

Santiago de Compostela, May 15th, 2007



RECERCAT



JOCS

TAC



TERAFLOP

1. What is CESCA?
2. Introduction: Basic Concepts
3. Policies on Cooperative Repositories
 - Open Access
 - Open Archive Initiative (OAI)
 - Cooperation Benefits
4. Tesis Doctorales en Red (TDR)
5. Dipòsit de la Recerca de Catalunya (RECERCAT)
6. Revistes Catalanes amb Accés Obert (RACO)
7. Patrimoni Digital de Catalunya (PADICAT)
8. Memòria Digital de Catalunya (MDC)
9. Conclusions

What is CESCA?



- ✓ Public consortium, created back in 1991 and integrated by:
 - Generalitat de Catalunya (local government)
 - Fundació Catalana per a la Recerca i la Innovació (FCRI)
 - 9 universities: UB, UAB, UPC, UPF, UdG URV, UdL, UOC & URL
 - Higher Scientific Research Council (CSIC)

- ✓ To manage a large complex of computing and communications systems (e-infrastructures) to provide service to the university and the research community, based on three activity areas:
 - The systems for scientific calculation and university information
 - HPC, SED, SDF, **repositories**...
 - The communications based on the Anella Científica and CATNIX
 - RedIRIS's node in Catalonia
 - The use and promotion of these technologies
 - Difusion, training and editing of the *Teraflop* magazine

- ✓ What are the Institutional Digital Repositories?

The Scholarly Publishing and Academic Resources Coalition (SPARC) defines them as digital collections that **capture and preserve** the intellectual results from **several institutions**

- ✓ Remarkable characteristics:
 - Contain documents created by the institutions
 - Academic content
 - Accumulative and persistent
 - Open access and interoperability

<http://www.sparceurope.org/Repositories>



“A institutional repository is a **set of services** that a university offers to the members of it’s community for the management and dissemination of **digital materials** created by the institution and its community members. It is most essentially and organizational commitment to the stewardship of these digital materials, including **long-term preservation** where appropriate, as well as **organization and access or distribution**”

Clifford A. Lynch, *Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age*, 2003



1. Open Access

To spread the knowledge, not only by the traditional system, but also by permitting the free and open access through the internet

2. Open Archive Initiative (OAI)

Interoperability among repositories

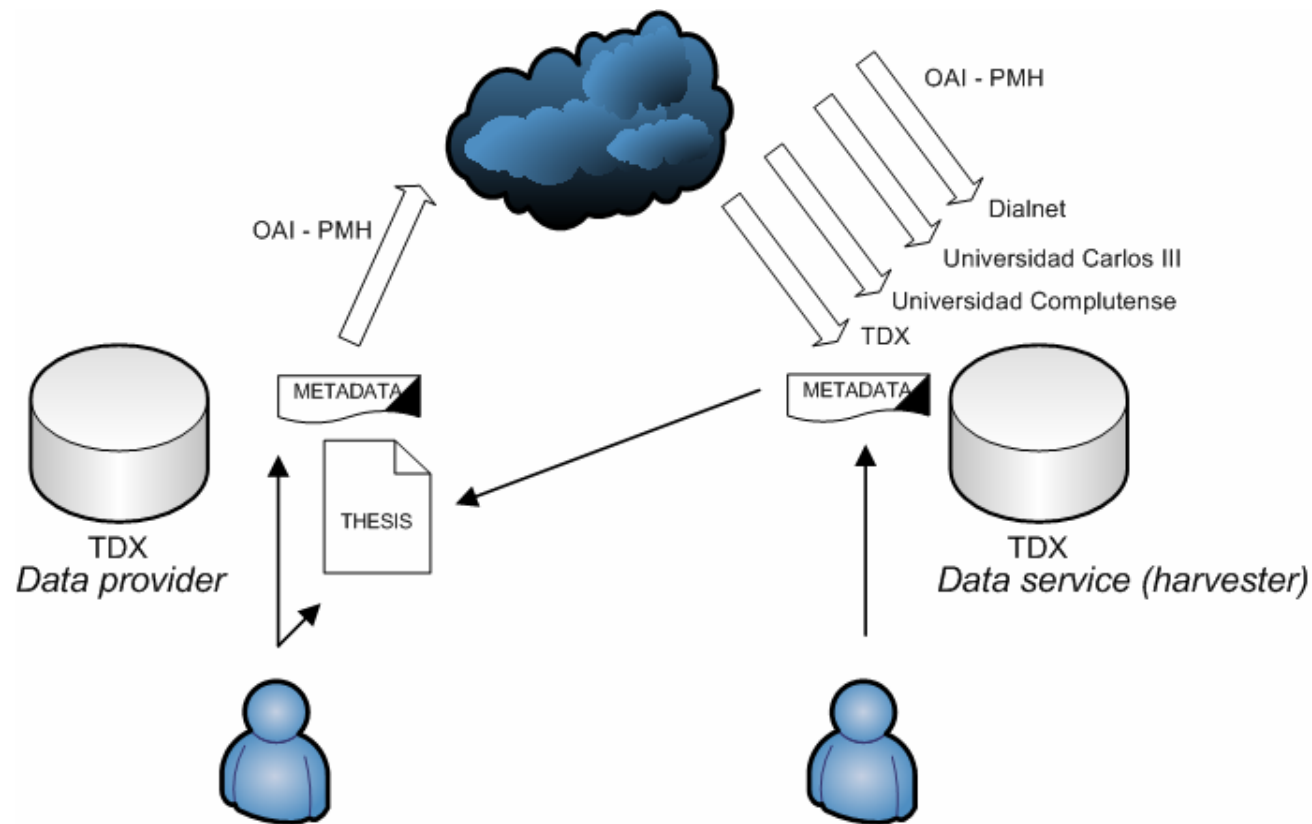
3. Cooperation Benefits

- Adoption of common procedures and sharing the developing and managing costs
 - Task distribution!
- More visibility of the indexed documents
 - Diffusion of the research activity
- Better provision for long-term preservation

- ✓ By the end of 2003, the German Max Planck Society promoted the **Berlin Declaration** on Open Access to Knowledge in the Science and Humanities (<http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>)
 - Supports the wide and transparent diffusion of the scientific knowledge and the human thinking by means of new possibilities provided by internet
 - Proposes to spread the knowledge, not only by the traditional system, but also by permitting the free and open access through the net
 - Provides an alternative to the paradigm of having to pay for accessing information obtained with public funds (generated by the own institution)

- ✓ The author(s) should
 - Guarantee to all the users the right to the open access (...), with permission to copy, to use, to spread, to transmit and to put forward the works publicly (...) with some responsible purpose, and the engagement of naming correctly the responsibility
 - Place online a version of this work (...) in an appropriate electronic format
 - This file will be administered and kept by an academic institution, a research institution, a public administration or an institution that can ensure the open access

- ✓ Interoperability among repositories
 - Sharing metadata via OAI-PMH



1. What is CESCA?
2. Introduction: Basic Concepts
3. Policies on Cooperative Repositories
 - Open Access
 - Open Archive Initiative (OAI)
 - Cooperation Benefits
4. Tesis Doctorales en Red (TDR)
5. Dipòsit de la Recerca de Catalunya (RECERCAT)
6. Revistes Catalanes amb Accés Obert (RACO)
7. Patrimoni Digital de Catalunya (PADICAT)
8. Memòria Digital de Catalunya (MDC)
9. Conclusions

Tesis Doctorales en Red (TDR/TDX)



TDX Theses and Dissertations Online

What is it?

Search

- TDX
- Global

Statistics

Intranet

Contact Us

Català | Castellano | Galego

Author

Advisor

Title

Organization

Department

Subject

Keywords

Free Text

Year of defense

Match Case

SEARCH

Participants

- Universitat de Barcelona
- Universitat Autònoma de Barcelona
- Universitat Politècnica de Catalunya
- Universitat Pompeu Fabra
- Universitat de Girona
- Universitat de Lleida
- Universitat Rovira i Virgili
- Universitat Oberta de Catalunya
- Universitat Ramon Llull
- Universitat de les Illes Balears
- Universitat de València
- Universitat de Vic
- Universitat Jaume I
- Universidad de Cantabria
- Universidad de Murcia
- Universidad de Oviedo
- Universidade da Coruña

Coordination

-
-

Collaboration

-
-

Sponsored

-

<http://www.tesisenred.net>

- ✓ Full-text searches (in PDF files)
- ✓ Search by author, director, title, knowledge area, university, keywords...
- ✓ Indexed by subject (CDU)

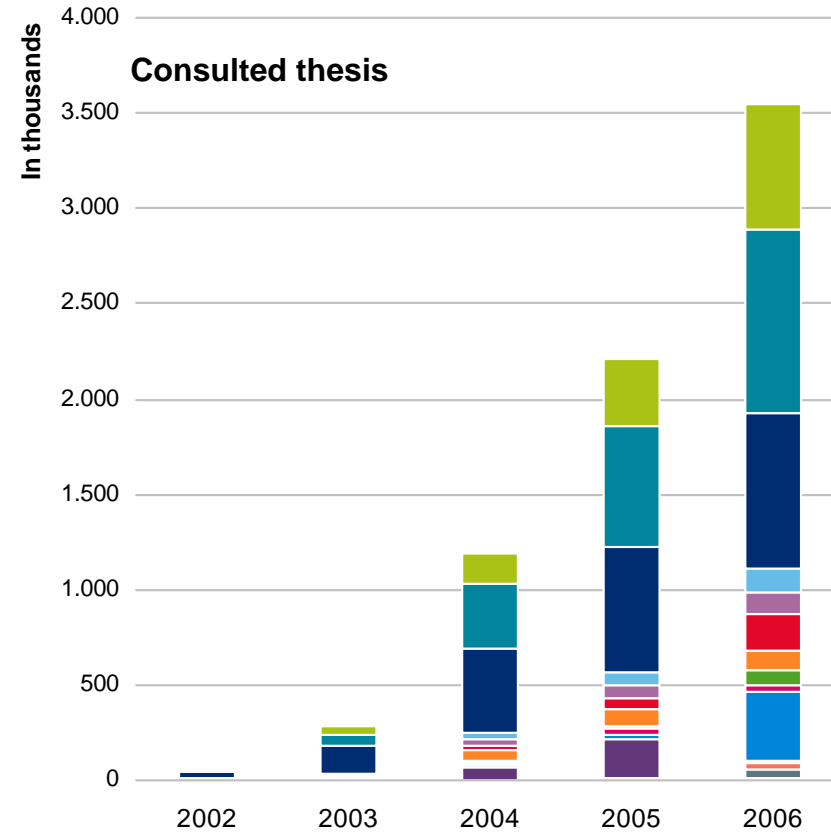
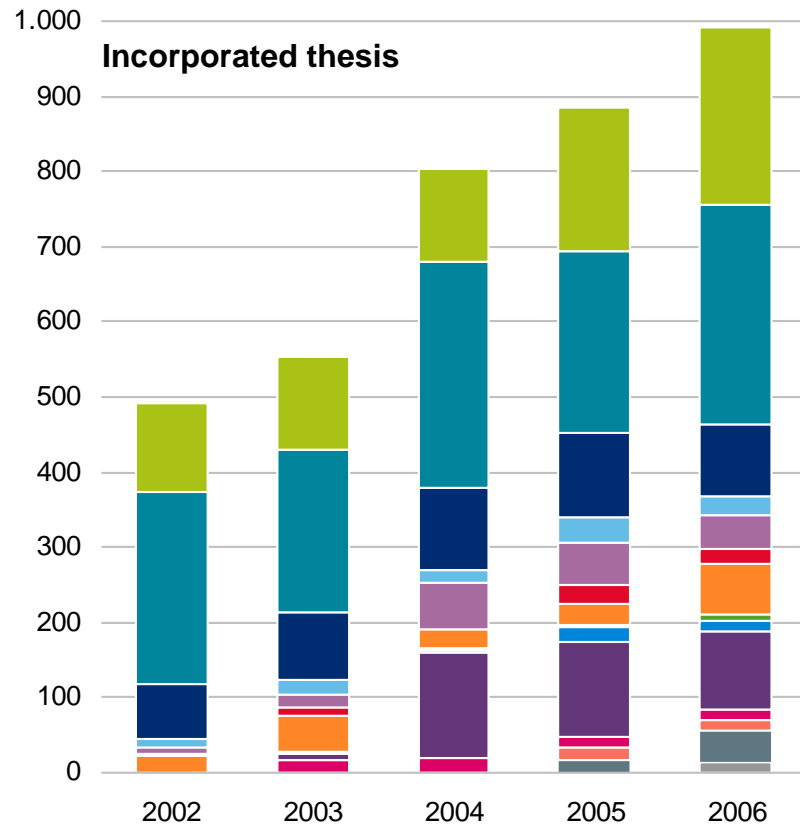
- ✓ Multilingual environment
- ✓ Access and usage statistics
- ✓ Data and service provider
- ✓ 17 universities

Tesis Doctorales en Red (TDR/TDX)



- ✓ First OAI-PMH harvester of Spanish thesis (since May'07)
- ✓ Limited searchable fields
- ✓ For the moment: only Dialnet, e-archivo and own TDR

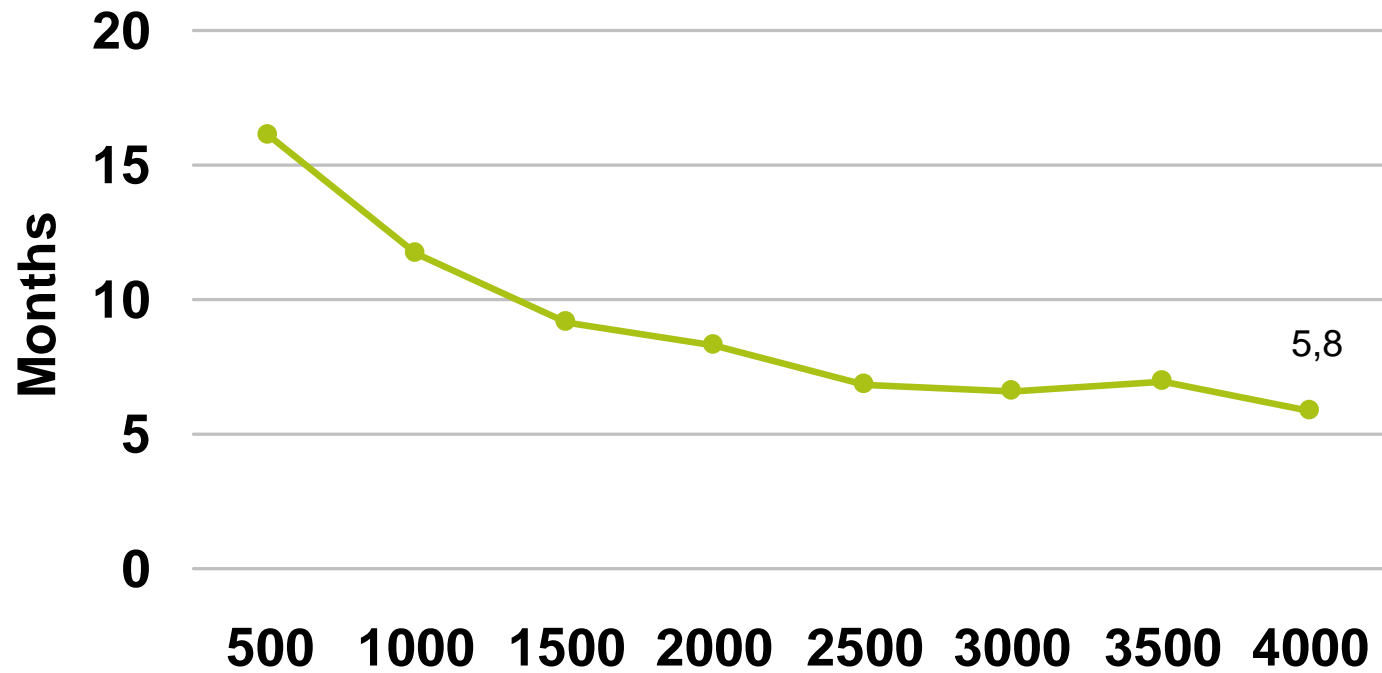
The TDR Evolution



%	UB	UAB	UPC	UPF	UdG	UdL	URV	UOC	URL	UIB	UV	UJI	UC	UM	UDC
Incorporated	20	36	14	3	5	3	5	0	0	1	9	2	1	1	0
Consulted	17	28	29	3	3	3	5	0	0	1	9	2	1	1	0

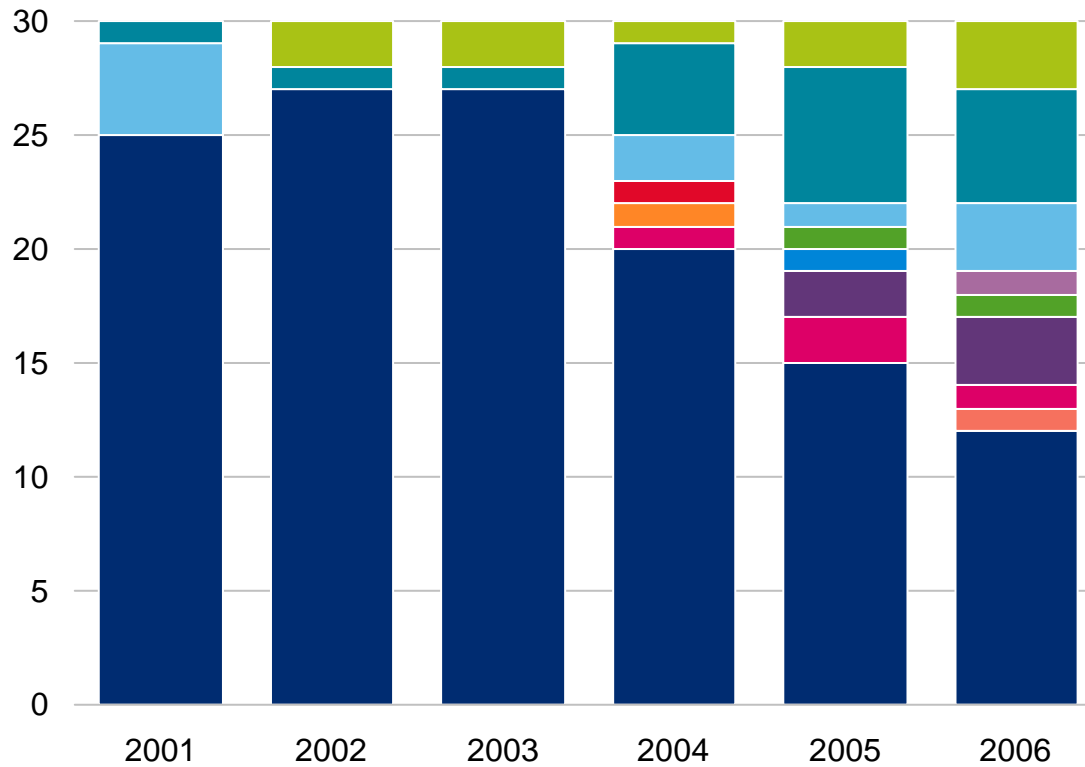


The Downloading Rhythm



The Information Society Penetration Factor

✓ Based on the 30 thesis more consulted:



- ✓ RECERCAT is a cooperative repository of digital documents and it includes the research literature from universities and researching centers of Catalonia, like preprints, communications in congress, research reports, working papers, university final projects, technical reports, etc.
- ✓ Operative since September'05
- ✓ Coordinated by CESCA-CBUC (as TDR)



The screenshot shows the RECERCAT website interface. At the top, there is a search bar and the RECERCAT logo with the text "Dipòsit de la Recerca de Catalunya". Below the search bar, there is a navigation menu on the left with options like "What is it?", "Statistics", "News", "Browse", "Communities & Collections", "Titles", "Authors", "Subjects", "By Date", "Sing on to", "Receive News", "My Repository", and "Edit Profile". The main content area features a search box, a "Communities" list including Universitat de Barcelona, Universitat Autònoma de Barcelona, Universitat Politècnica de Catalunya, Universitat Pompeu Fabra, Universitat de Girona, Universitat de Lleida, Universitat Rovira i Virgili, Universitat Oberta de Catalunya, Universitat de Vic, Agència de Gestió d'Ajuts Universitaris i de Recerca, and Institucions culturals. On the right side, there is a Creative Commons license notice, a "Coordination" section with logos for CESCA and BU, a "Sponsorship" section with the Generalitat de Catalunya logo, and a "Powered by" section with the DSPACE logo. A vertical URL <http://www.recercat.net> is displayed on the right side of the screenshot.

- ✓ 13 institutions
- ✓ More than 3,500 documents
- ✓ Full-text open access
- ✓ Creative Commons Attribution-NonCommercial-ShareAlike
- ✓ OAI-PMH data and service provider



- ✓ RACO (Catalan Journals in Open Access) is a cooperative repository where can be consulted, in open access, the full-text from articles of scientific, cultural and erudite Catalan journals
- ✓ The main purpose of RACO is to increase the visibility and consults of the journals included and to spread the scientific and academic production published in Catalan journals
- ✓ This purpose makes specific in three aims:
 - To encourage the electronic edition of Catalan journals
 - To be the interface that allows the whole search of all the journals
 - To provide the instruments for its preservation
- ✓ Coordinated by CESCA-CBUC-BC

The screenshot shows the RACO website interface. At the top left, the RACO logo is displayed next to the text "Revistes Catalanes amb Accés Obert". A navigation bar contains links for "About", "News", "Participating Institutions", "Subscribe", "Statistics", and "Intranet". Language options for "Català", "Castellano", and "English" are also present. A search bar is located in the top right, with a dropdown menu set to "In the title of the article" and a "Legal note" link.

The main content area is divided into two search sections:

- Article search:** Includes input fields for "Article title", "Article author", "Words", and "Journal title" (with a dropdown menu set to "All journals"). There are checkboxes for "Exact word" next to each field. A "Search limit" field is set to "10". A "Search" button is located to the right of the "Search limit" field.
- Journal search:** Includes dropdown menus for "By subject" (set to "All subjects"), "By title" (set to "All journals"), and "By institution" (set to "All institutions"). Each dropdown menu has a corresponding "Search" button.

On the right side of the page, there is a vertical sidebar with the following content:

- Coordination:** Logos for CESCA, BU, and IRI.
- Sponsored:** Logo for Generalitat de Catalunya.
- Powered by:** Logo for Open Journal Systems.

<http://www.raco.cat>

Last update: 27-04-06

There are 129 journals

- ✓ 129 journals
- ✓ 26 editorial institutions
- ✓ More than 25,000 articles
- ✓ Full-text open access (some they may have a delay before public)

<http://www.padi.cat>

- ✓ To archive the Catalan web sites
 - Webs either with `.cat` domain or under other domain names, but located in Catalonia or related with Catalan culture
- ✓ Hybrid model (selective and systematic)
- ✓ Coordinated by BC

National Digital Archives or Web Archives: a Comparison



The screenshot shows the Internet Archive website. At the top, it says "INTERNET ARCHIVE" and "Universal access to human knowledge". Below that, there are navigation links: "Web | Moving Images | Texts | Audio | Software | Patron Info | About IA". A search bar is visible with "All Media Types" selected. Below the search bar, there is a section for "THE ROYAL LIBRARY National Library of Sweden" with the text "Kulturarw³- Long time preservation of electronic". The main part of the screenshot shows the "PANDORA AUSTRALIA'S WEB ARCHIVE" interface. It has a search bar with "Search PANDORA" and "Search help" buttons. Below the search bar, there is a "Browse subjects:" section with a grid of subject categories:

Arts & Humanities	Health	News & Media
Business & Economy	History & Geography	Politics & Government
Computers & Internet	Indigenous Peoples	Science & Technology
Education	Juvenile	Society & Culture
Environment	Law & Criminology	Sports & Recreation

Below the subject grid, there is a link to "View a complete listing of titles available within PANDORA or browse titles alphabetically" with a list of letters: "1-9 A B C D E F G H I J K L M N O P Q R S T U V W X-Z". At the bottom, there is the "NATIONAL LIBRARY OF AUSTRALIA" logo.

✓ Internet Archive

- 35m webs
- 2 PB

✓ Kulturarw3 (Sweden)

- 350,000 webs
- 13 TB

✓ PANDORA (Australia)

- 9,000 webs
- 1.2 TB

✓ PADICAT today (since July'06)

- 646 versions
- 110 webs
- 200 GB

Memòria Digital de Catalunya (MDC)



memòria digital de Catalunya

Inici : Navegar : Cerca avançada : Preferències : Favorits : Què és la Memòria Digital de Catalunya : Ajuda cercar

seleccionar-ho tot : esborrar-ho tot : afegir a Favorits pàgina 1 de 2 : (<< 1 2 >>) :: anterior : següent

navegació 33 document(s) a: Arxiu de Revistes Catalanes Antiques anar-hi

Imatge:	Títol:	Matèria:	Descripció:
<input type="checkbox"/> 1.	Art: revista de les Arts	Art -- Revistes ; Literatura catalana -- Revistes	Canvi de subtítol i reinici de numeració a partir de 1933
<input type="checkbox"/> 2.	Campana Eulalia: periódico satírico ilustrado	Barcelona -- Revistes satíriques	
<input type="checkbox"/> 3.	Carcajada, La	Humorisme castellà -- Revistes ; Humorisme gràfic -- Revistes	

<http://www.cbuc.cat/mdc>

- ✓ To collect old reviews, photographs, maps, exlibris, etc. related to the Catalan heritage
- ✓ 7 institutions
- ✓ More than 3,700 images (in 19 collections)
- ✓ Open access, coordinated by CBUC

- ✓ Soon...
 - e-learning courses
 - For FECYT, Spanish science and technology magazines



- ✓ Looking for... audiovisual material
 - How to preserve it?



1. By having 5 repositories available, it seems that most of the work has been accomplished

2. This is not true, only the “easiest” part of the work
 - Legal mechanisms have been established to protect intellectual property of knowledge
 - e-infrastructures have been set up to spread the e-information throughout the internet

3. From now on, the “difficult” part has to be accomplished
 - Researchers have to become the best users
 - As well as to contribute their work into them
 - Research centers have to establish the procedures to make it possible
 - Government and financial research institutions have to promote open access to scientific and humanistic information

CENTRE DE SUPERCOMPUTACIÓ
DE CATALUNYA



Thank you for your attention!

Any question?

direccio@cesca.es



RECERCAT



JOCS

TAC

TSIUC

TERAFLOP