



# **El proyecto PADICAT (Patrimoni Digital de Catalunya)**

Biblioteca de Catalunya

con la colaboración de:

Centre de Supercomputació de Catalunya (CESCA)

Departament d'Universitats, Recerca i Societat de la

Informació



# La Biblioteca de Catalunya (I)

- Creada en 1907 por Enric Prat de la Riba
- 3.000.000 de documentos apróx.
- Crecimiento anual de 120.000 documentos
- 230 puestos de consulta en unos 3.000 m<sup>2</sup>
- Datos 2006: 11 M€+ 172 personas + 50 Km de depósito



## La Biblioteca de Catalunya (II)

La misión de la Biblioteca de Catalunya es

- compilar, conservar y difundir la producción bibliográfica catalana y la relacionada con el ámbito lingüístico catalán
- velar por la conservación y la difusión del patrimonio bibliográfico
- mantener la condición de centro de consulta e investigación científica de carácter universal.



# La Biblioteca de Catalunya (III)

Plan estratégico 2004-2008

ABIERTA

FIABLE

ÚTIL

Especial atención al contexto digital  
sin olvidar la producción *tradicional*



# La producción *bibliográfica* en 1971

“Son objeto de depósito legal los escritos, estampas, imágenes y composiciones musicales, producidas en territorio nacional, en ejemplares múltiples, con fines de difusión, hechos por procedimientos mecánicos o químicos. Comprenderá por tanto:

**Libros**, sea cualquiera la índole de su contenido y la forma de impresión y estén o no destinados a la venta.

**Folletos**, o sea escritos cuyo número de páginas sea mayor de cuatro y no exceda de 50, y con características semejantes a las señaladas en el párrafo anterior, incluyéndose en este concepto las separatas de artículos de revista que tengan la acotada

extensión. **Hojas impresas** con fines de difusión y que no constituyan propaganda esencialmente comercial.

**Publicaciones periódicas** (revistas y diarios). **Partituras musicales.** **Grabados:** láminas sueltas,

láminas de calendario, estampas, cromos, "chrismas", anuncios artísticos. **Mapas y planos.** **Carteles** anunciadores de espectáculos, fiestas y demás actos públicos, tanto religiosos como profanos; anunciadores de artículos

comerciales, siempre que lleven grabados artísticos; bandos y edictos. **Postales ilustradas.** **Naipes.**

“**Slides**” destinadas a difusión y venta. **Impresiones o grabaciones sonoras**

realizadas por cualquier procedimiento o sistema empleado en la actualidad o en el futuro. **Producciones**

**cinematográficas**, tanto de tipo argumental como documental, y "filmlets”.

Fuente: Orden de 30 de octubre de 1971, por la que se aprueba el Reglamento del Instituto Bibliográfico Hispánico



## Distribution of Top-Level Domain Names by Host Count Jan 2005

Domain	Hosts =	All Hosts	- Dup Names	Level 2 Domains	Level 3 Domains	
TOTAL	317646084	377387537	59741453	3011940	71168724	
net	139057448	141788322	2730874	217788	41926996	Networks
com	56428268	106658870	50230602	1894844	13345788	Commercial
jp	19543040	19668895	125855	32873	307025	Japan
it	9343663	9398958	55295	32779	662573	Italy
edu	8992398	9238213	245815	4469	2958885	Educational
arpa	6817735	7195504	377769	17	5084	Mistakes
nl	6443558	8458045	2014487	47393	829949	Netherlands
de	6127262	6264784	137522	151368	1103801	Germany
fr	4999770	5127240	127470	14197	424345	France
au	4820646	4906182	85536	45	51193	Australia
uk	4449190	5302225	853035	66	67556	United Kingdom
br	3934577	3998966	64389	402	49785	Brazil
ca	3839173	4260553	421380	21602	1407118	Canada
tw	3516215	3574416	58201	46	15090	Taiwan, Province Of Chir
se	2668816	2683229	14413	9168	506136	Sweden
pl	2482546	2517280	34734	7431	413975	Poland
us	2429244	2488667	59423	14337	38293	United States
be	2012283	2104966	92683	9492	212082	Belgium
fi	1915506	1934620	19114	10766	795865	Finland
dk	1908737	1941838	33101	10484	186622	Denmark
mx	1868583	1913248	44665	117	29248	Mexico
ch	1785427	1918697	133270	31475	636248	Switzerland
mil	1667794	1823785	155991	111	95692	US Military
at	1594059	1645857	51798	13758	235064	Austria
org	1459335	1554092	94757	205537	1135697	Organizations
es	1304558	1352485	47927	4060	347338	Spain

Fuente: Internet System Consortium, 2005



# La producción en Internet en 2005

.es	1.304.558 dominios
.net	139.057.448 dominios
.com	56.428.268 dominios
.edu	8.992.398 dominios
.org	1.459.335 dominios
.info	30.828 dominios

Fuente: Internet System Consortium, 2005



# La producción *bibliográfica* en 2006

Libros, Folletos, Hojas impresas, Publicaciones periódicas,  
Grabados, Mapas y planos, Carteles, Postales, “Slides”,  
Impresiones o grabaciones sonoras, Producciones  
cinematográficas...

+

Libros digitales, Diarios y revistas en línea, Sedes  
web, Weblogs, Webcams, Votaciones en línea,  
Chats, Sedes web corporativas, Comercio en línea,  
Webs personales, Artículos y ponencias digitales,  
Documentación de software...



# ¿La producción *bibliográfica* en 2020?

## Pronostican ocaso de los periódicos de papel en 15 años más



(02/08/2005 08:32 EST): Los periódicos de papel pierden cada vez más lectores, que prefieren las versiones en línea. Expertos británicos prevén que la publicación digital de contenidos habrá desplazado totalmente al papel antes de 2020.

DIARIO TI: Alrededor del 25% de los lectores de periódicos ahora acceden a los contenidos noticiosos únicamente vía Internet, indica la organización Publishing Industry Network (PIN). Las estadísticas de PIN indican que aunque la mayoría de los lectores de periódicos continúan comprando la versión de papel, un porcentaje cada vez mayor migra a Internet.

La Biblioteca Nacional Británica prevé un panorama similar para los libros y estima que la gran mayoría de las producciones literarias serán publicadas digitalmente para el año 2020.

En tal contexto, la Biblioteca Nacional Británica ha anunciado recientemente planes de lanzar una nueva biblioteca digital. Por ahora, la directora de la entidad, Lynne Brindley, ha expresado aprensiones en el sentido que el material digital es más vulnerable que los libros físicos. "Si no tratamos este material de forma adecuada, no sobrevivirá para las futuras generaciones", señaló a BBC.



Año 9 - Edición 1835

**Jueves  
4 Ago  
2005**



# El patrimonio digital



# El patrimonio digital (I)

Cada vez más, los recursos fruto del conocimiento o la expresión de los seres humanos, sean éstos de carácter cultural, educativo, científico o administrativo, o engloben información técnica, jurídica, médica y de otras clases, se generan directamente en formato digital o se convierten a éste a partir del material analógico ya existente.

Fuente: Unesco, 2003



# El patrimonio digital (II)

RACO

(Revistas Catalanas en Acceso Abierto)

ARCA

(Archivo de Revistas Catalanas Antiguas)

CLACA

(Clásicos Catalanes)

Fuente: Unesco, 2003



# El patrimonio *nacido digital* (I)

Los productos *nacidos digitales* no existen en otro formato que sea el electrónico o digital.

Cada vez más numerosos, son a menudo efímeros y se crean y desaparecen en muchos casos con rapidez

Fuente: Unesco, 2003



# El patrimonio *nacido digital* (II)



CANDIDATURA DEL POETA  
EN MIQUEL MARTÍ I POL  
PER A PREMI NOBEL DE LITERATURA



Introducció	Adhesió	Biografia	Bibliografia
<p>CCOO-PV recolzarà en un acte públic el 5 d'octubre i a la ciutat de Castelló la candidatura al guardó de Premi Nobel de Literatura per al poeta Miquel Martí i Pol.</p> <p>A l'acte hi participaran diversos representats del món cultural i sindical de la ciutat de Castelló i del País Valencià i comptarà amb l'actuació poètica i musical de Toti Soler i Cinta Massip amb l'espectacle "Si alguns mots o la veu de l'home".</p> <p>Miquel Martí i Pol va néixer a Roda de Ter (Osona) el 19 de març de 1929. El seu pare era manyà i la mare treballadora del tèxtil. Estudià fins els catorze anys en una escola parroquial i participà en tasques culturals al Centre Catòlic del seu poble. Als catorze anys va haver de deixar els estudis i entrà a treballar a l'empresa Tecla Sala i Fills "La Blava", fins que en 1972 una malaltia el va obligar a recloure's a casa seua.</p> <p>La seva obra poètica i literària és molt ampla i diversa ha estat traduïda</p>			



# El patrimonio *nacido digital* (III)

INICI QUÈ ÉS WWW.TERRICABRAS-FILOSOFIA.INFO? EL WEBLOG [ ESPAÑOL ] [ ENGLISH ]

## www.terrificabras-filosofia.info

INFORMACIÓ L'ARXIU PARTICIPACIÓ DIÀLEG AULA DE SECUNDÀRIA CURSOS Dijous, 4 d'agost de 2005

**INFORMACIÓ**

- Joan Gamper, 75 anys
- Vacances d'agost
- Les llibertats, per Josep-Maria Terricabras
- CATALUNYA. En marxa cap a la llibertat

**DIÀLEG**

- Turismes

**AULA DE SECUNDÀRIA**

- 1r BAT: El plaer de la recerca
- 2n BAT: Cap a la universitat
- ÈTICA. L'ètica en crisi (II)

**FÒRUMS**

XAT VOTA

**EL PORTAL PER REFLEXIONAR EN XARXA**

**USUARIS REGISTRATS**

**USUARI**

**CONTRASEIYA**

 **ENTRAR**

- Us voleu registrar gratuïtament?
- Heu oblidat la contrasenya?
- Us voleu subscriure?
- Coneixeu els avantatges de ser subscriptor?

**CERCA**

**QUÈ VOLEU CERCAR?**

 **CERCAR**

**A QUINA SECCIÓ?**

( tot el portal )

**el weblog**  
reflexions diàries de Josep-Maria Terricabras



# El patrimonio *nacido digital* (IV)

**Plataforma Aturem la Guerra**

**Altres documents**

- [Manifest](#) del 20 de març de 2004 (pdf 66 kb)
- [Manifest](#) del 15 de febrer de 2004 (pdf 94 kb)
- [Manifest](#) (legit durant l'acte dut a terme al Paraninfo el 14 de febrer de 2004, pdf 75kb)
- [Manifest](#) (de la manifestació del dia 27 de setembre de 2003)
- [Manifest](#) (de la manifestació del dia 1 de juny de 2003)
- [Manifest](#) (versió de març de 2003, pdf 56 kb)
- [Manifest](#) (versió d'octubre de 2002)
- [Logotip](#) (pdf 175 kb)
- [Declaració de Londres](#) (1/3/03)

**Materials de venda**

- **Xapes:** es retiren a 50 cèntims i es venen a 1 euro
- **Banderolles:** es retiren a 1 euro i es venen a 2 euros
- **Adhesius:** es retiren els paquets de 500 a 5 euros i es venen a la voluntat
- **Cartells:** de franc

Podreu comprar el material de la campanya a:

**Contacte:**

Centre de Treball i Documentació (CTD)  
c/ Sant Salvador, 96 baixos  
08024 Barcelona  
Telèfon 93 284 65 20

**Entitats Adherides a Barcelona**

AAPG  
ACAT (Acció Cristiana per l'Abolició de la Tortura)  
Acció Ecologista  
Acció Solidària Blanquerna (ASB) de la Universitat Ramon Llull  
ACO  
Agrupament Escolta i Guia Terra Plana de Belianes  
Aixec Comunista per la Independència  
Ajuntament de Barcelona  
Ajuntament de Molins de Rei  
Amèrica Solidària 2002  
Apropar  
ARAGUAIA, amb el Bisbe Casaldàliga per les causes de l'alliberament  
Arenys Solidari  
ARPIJOTS (Grup de defensa del Patrimoni històric i natural)  
Arquitectes Sense Fronteres  
ASDHA (Associació de Drets Humans a l'Afganistan)  
Associació de Miners del Ramon de la Mina

**Actes de les reunions**

**Barcelona**

- 22 de maig de 2003
- 15 de maig de 2002
- 23 d'abril de 2003 (reunió-debat)
- 24 d'abril de 2003
- 27 de març de 2003
- 27 de febrer de 2003
- 28 de novembre de 2002
- 26 de setembre de 2002

**Coordinació amb altres plataformes**

- 17 d'abril 04 - coordinació estatal
- 10 de gener 04 - coordinació catalana
- 27 de març - coordinació estatal
- 9 de març - coordinació estatal
- [Resum reunió internacional a Londres de l'1 de març](#)

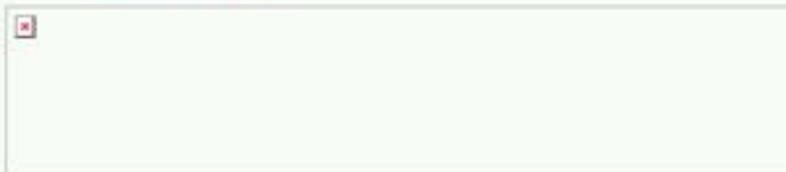
**Premsa**

- [Comunicat 16/07/04](#) No a l'enviament de tropes a l'Afganistan.
- [Comunicat 13/05/04](#) convocant a l'acte públic reivindicatiu del dia 15 de maig i a la manifestació de l'19 de juny contra las



# El patrimonio *nacido digital* (V)

Les Notícies de VilaWeb : milers de mallorquins celebren a Lloc els 35 anys de l'OCB.



## Un cap de setmana força esportiu

Aquest cap de setmana ve marcat per una successió d'esdeveniments esportius

- Comença a Barcelona l'[Open Godó de Tennis](#)
- Els [Dragons](#) han jugat el primer partit de la [Lliga Mundial](#) aquest dissabte, aconseguint una important victòria a Dusseldorf per 12 a 27
- ha començat el [campionat del món de motociclisme](#), amb una segona posició per a Àlex Crivillé (recordeu la web [Els Pòdium](#), on podeu apostar qui el guanyarà)

i, com sempre, seguir el futbol... El València, dissabte, apallissà l'Atlètic de Bilbao 5 a 2 i el Barça ha repetit el resultat però al revés contra l'Atlètic de Madrid guanyant 2 a 5.

Us recomanem visitar el servidor de [Caixa de Terrassa](#)



1997

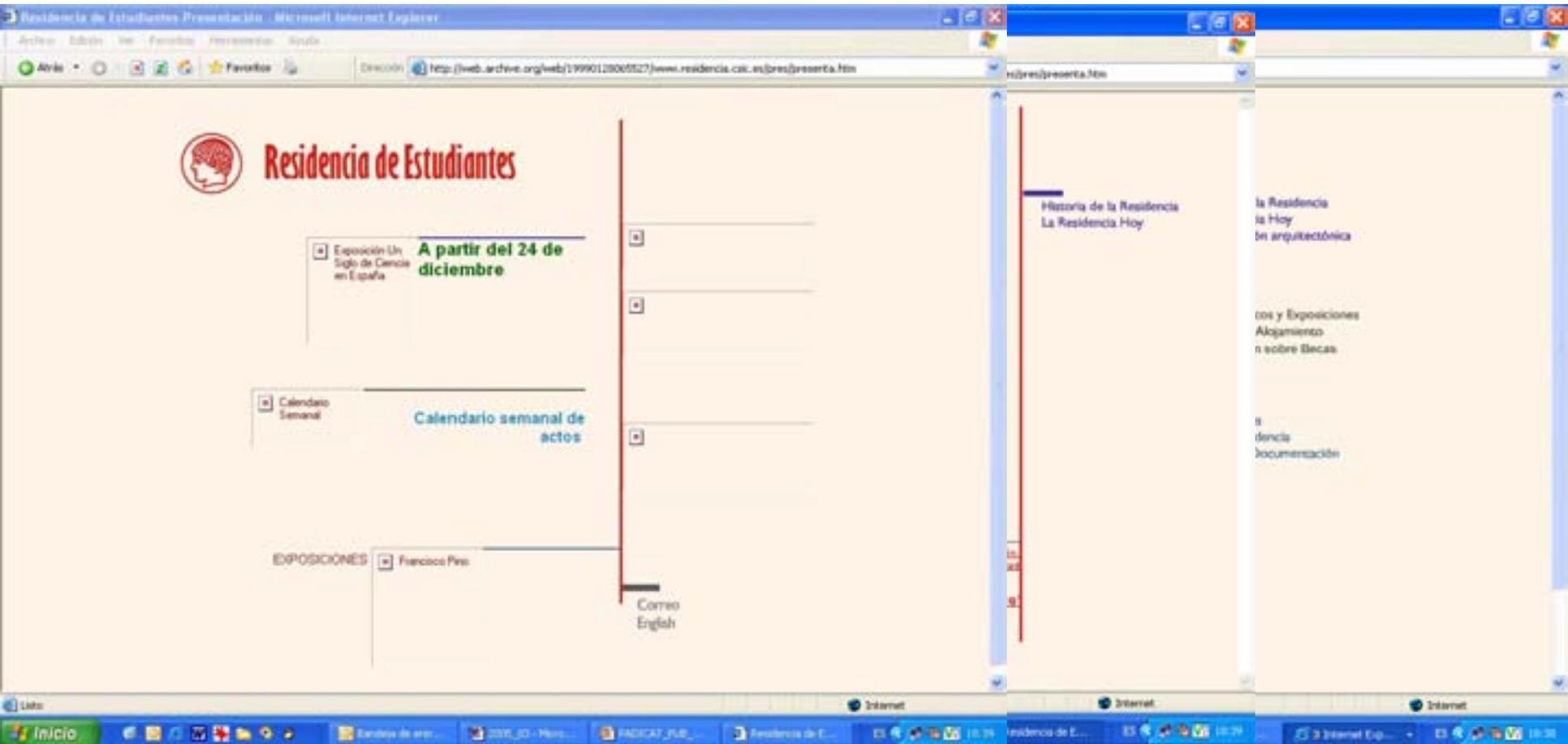
1998

2003

2005



# El patrimonio *nacido digital* (VI)



1998

2001

2005



# Experiencias en preservación digital

**Alemania, 1997**

**Australia, 1996**

**Áustria, 1999**

**Canadá, 1994**

**Dinamarca, 1998**

**EEUU, 2000**

**Estonia, 2004**

**Finlandia, 1997**

**Francia, 2000**

**Grecia, 2003**

**Islandia, 1997**

**Japón, 2002**

**Lituania, 2002**

**Noruega, 2001**

**Nueva Zelanda, 1999**

**Países Bajos, 1995**

**Québec, 2000**

**Reino Unido, 2004**

**República Checa, 2001**

**Suecia, 1996**



# Modelos

## Exhaustivo o integral

Riqueza de la colección y contextualización

Resultados visibles a corto plazo

Coste bajo

Excluye Internet invisible

Colección irregular

Acceso limitado a los resultados

## Selectivo

Colección equilibrada

Máximo acceso a los resultados

Modelo estratégico

Visión parcial

Descontextualización

Coste elevado

## Híbrido

Colección rica y equilibrada

Contextualizada

Modelo estratégico con resultados visibles

Incluye Internet invisible

Máximo acceso (en relación con los derechos de autor)

Coste elevado similar al modelo selectivo



PADICAT  
(Patrimoni Digital de Catalunya)  
Archivando la web catalana



# **PADICAT (I)** **(Patrimoni Digital de Catalunya)**

**OBJETIVO: compilar, conservar y difundir  
la producción digital catalana**

- Producción digital creada en Catalunya
- En catalán
- Sobre Catalunya
- Autoría catalana

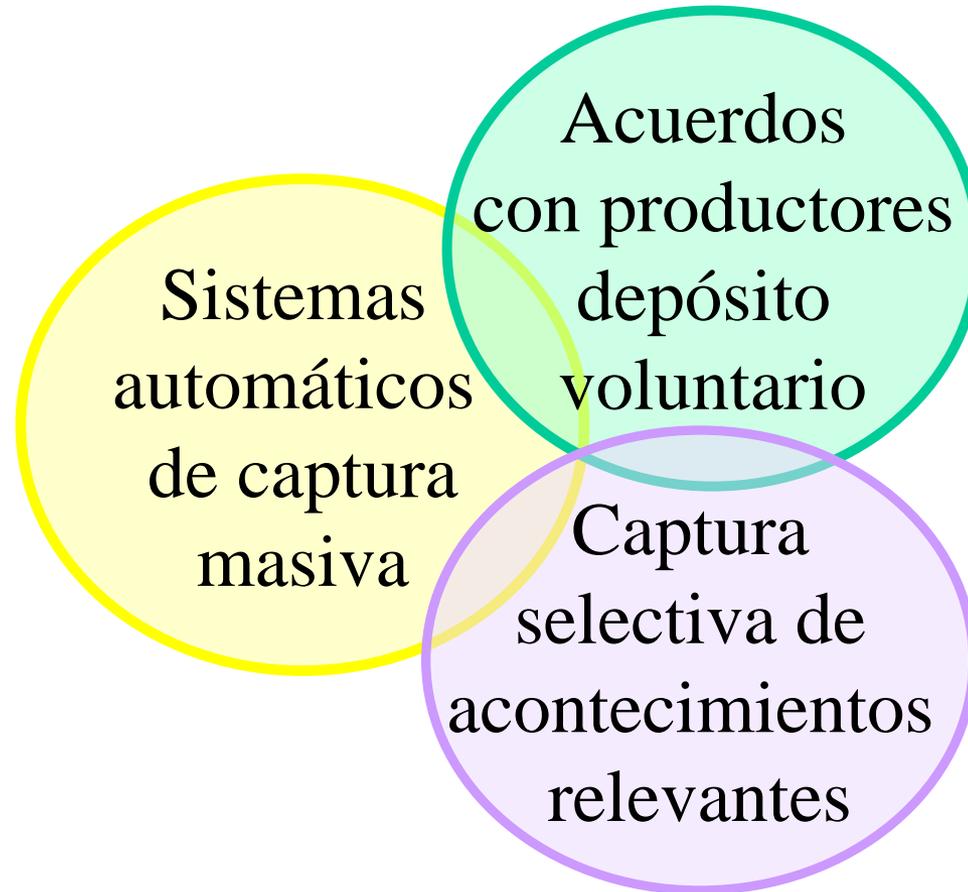


# PADICAT (II)

## (Patrimoni Digital de Catalunya)

### Estrategias combinadas

**Legislación**  
**DL**  
**?**



**Modelo híbrido**



# PADICAT (III)

## (Patrimoni Digital de Catalunya)

### Captura

**Una pàgina web identificable per una URL**

o un conjunt de pàgines web relacionades jeràrquicament a una pàgina principal identificable per una URL que constitueixi **una unitat documental reconeixible** e independent en grau suficient del rest per su temàtica, autoría o representativitat institucional.

En **cualquier formato** (jpeg, pdf, htm, html, gif ...)

Recursos **estáticos y dinámicos**



# PADICAT (IV)

## (Patrimoni Digital de Catalunya)

### Captura

- Webs bajo el dominio .CAT
- Webs en servidores de Catalunya
- Webs en lengua catalana bajo otros dominios (.ES, .ORG, .NET, .COM, .INF, ...)
- Otros webs relacionados temáticamente con Catalunya que no queden recogidos por los criterios anteriores



# PADICAT (IV)

## (Patrimoni Digital de Catalunya)

**Descripción** Identificador permanente y metadatos  
Dublin Core (descriptivos y técnicos)

**Almacenamiento** formato de compresión ARC

**Preservación** Migración, emulación, ...

Ventaja: 96 % formatos estándares (texto/html, imágenes en jpeg, gif o pdf.)

Inconvenientes: Formatos y aplicaciones efímeros, pérdida de contexto, software propietario, ...



# PADICAT (IV) (Patrimoni Digital de Catalunya)

**Acceso:** metadatos + búsqueda por texto libre

Aspectos legales: consulta en abierto o restringido a la red de la biblioteca, derechos de autor, limitaciones de copia y reproducción, ...

La legislación es prácticamente inexistente en relación con el patrimonio digital.



# El sistema de gestión PADICAT (I)

## Hardware

2 servidores (SUN Fire V490 o similar) de 1 GB de memoria y 10 TB de espacio en disco

## Software (código abierto)

Heritrix: robot de captura o crawler

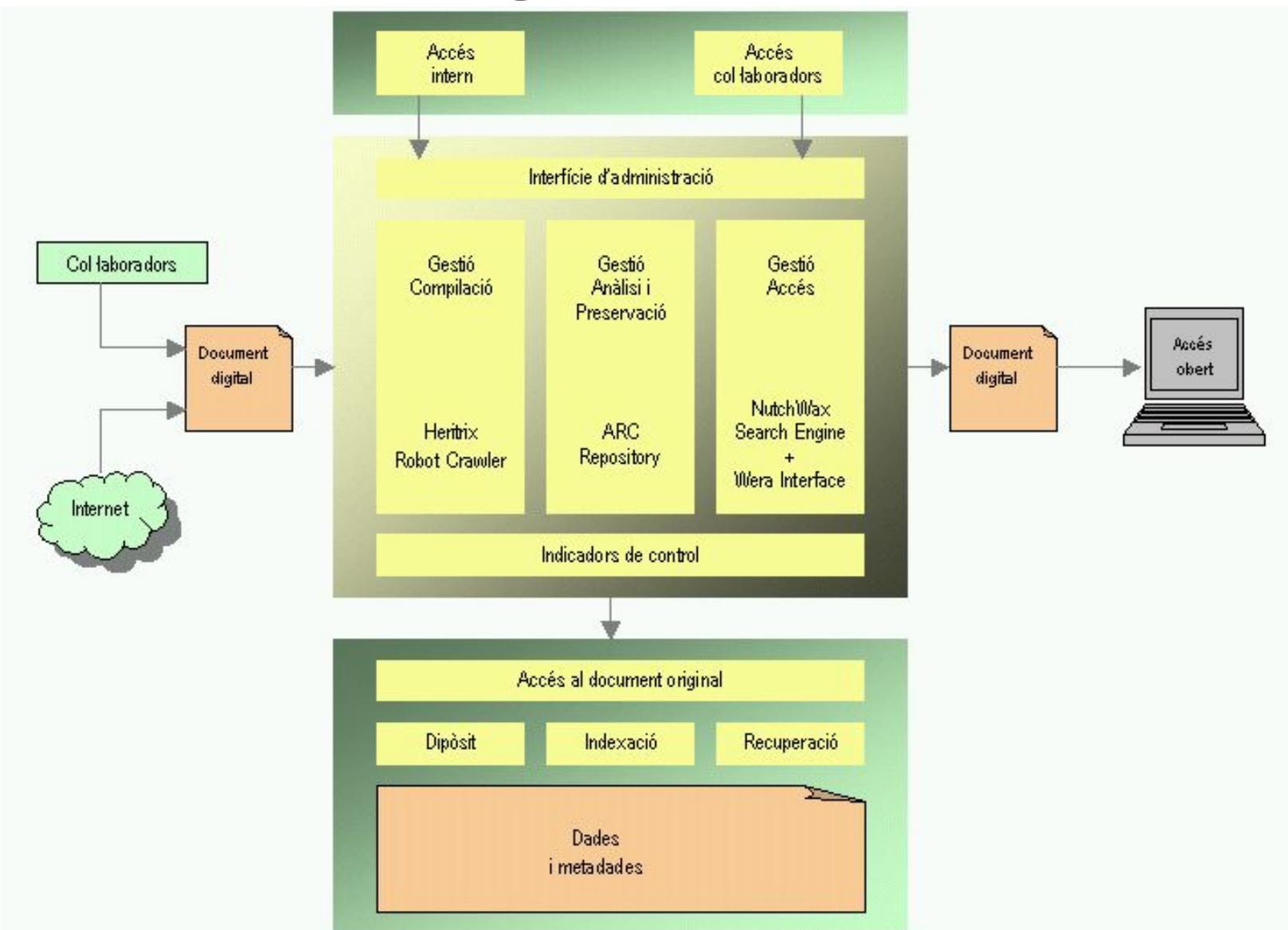
BAT: gestor de archivos

NutchWax: indizador y motor de búsqueda

Wera: interfaz de consulta



# El sistema de gestión PADICAT (II)





# PADICAT en números (I)

**Presupuesto plurianual 2006-2008 :  
766.000 €(hardware + personas)**

Equipo de 7 personas:

- Coordinador del proyecto
- Analista informático
- 2 Bibliotecarios de soporte
- 2 Bibliotecarios especialistas en metadatos
- 1 Administrativo (gestión de trámites y convenios)



# PADICAT en números (II)

## Previsiones para 2009

Proyecto pionero en España y de referencia en Europa (la web *territorial*)

100.000 webs

50 millones de archivos

30 TB de volumen

Acuerdos con 300 instituciones de todo tipo

Acceso en línea a buena parte de la colección



# Cronología PADICAT

- **2005 Análisis y diseño**
  - Estudio de los proyectos y iniciativas existentes, agentes implicados y aspectos legales. **Test de hardware y software.**
- **2006 Producción y plan piloto**
  - Implementación del sistema
  - Primeras capturas y acuerdos
  - Evaluación, rectificación de parámetros y procedimientos
- **2007-2008 Explotación: Oficina PADICAT**
  - Captura sistemática de la web catalana
  - Acuerdos para depósito voluntario regular
  - Acontencimientos



**¡Gracias por vuestra atención!**

Eugènia Serra [eserra@bnc.es](mailto:eserra@bnc.es)