

10 años de repositorios cooperativos

Ricard de la Vega

Jefe del Servicio de Portales y Repositorios

Centre de Supercomputació de Catalunya

4as. Jornadas OS-Repositorios

Barcelona, 3-5 marzo de 2010



Un taller en 5 partes



- ✓ Las diferentes fases que han vivido los repositorios cooperativos → Lluís M. Anglada
- ✓ Actuaciones llevadas a cabo en estos 10 años para llenarlos de contenido → Sandra Reoyo
- ✓ Experiencia de la UPF como participante en estos repositorios y con uno institucional → Anna Casaldàliga
- ✓ Infraestructura técnica en que se han desarrollado y han evolucionado → Ricard de la Vega
- ✓ Los retos a los que se enfrentan los repositorios cooperativos → Lluís M. Anglada





- ✓ Consorcio público
- ✓ Creado en 1991
- ✓ Formado por:
 - Generalitat de Catalunya
 - Fundació Catalana per a la Recerca i la Innovació
 - 9 universidades catalanas
 - Consejo Superior de Investigaciones Científicas
- ✓ Anella Científica creada en 1993



Nuestros servicios



Comunicaciones

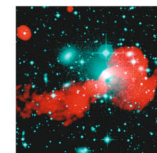


Portales y Repositorios



Cálculo y Archivo

CAP

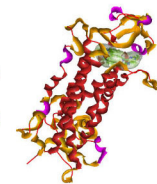


SED

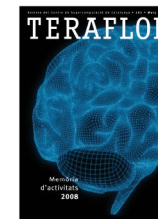


SDF

AUC



Promoción



JOCS

TAC

TSIUC

Operaciones y Seguridad



SAH
EC-UR
ERAC
S24x7

✓ Introducción

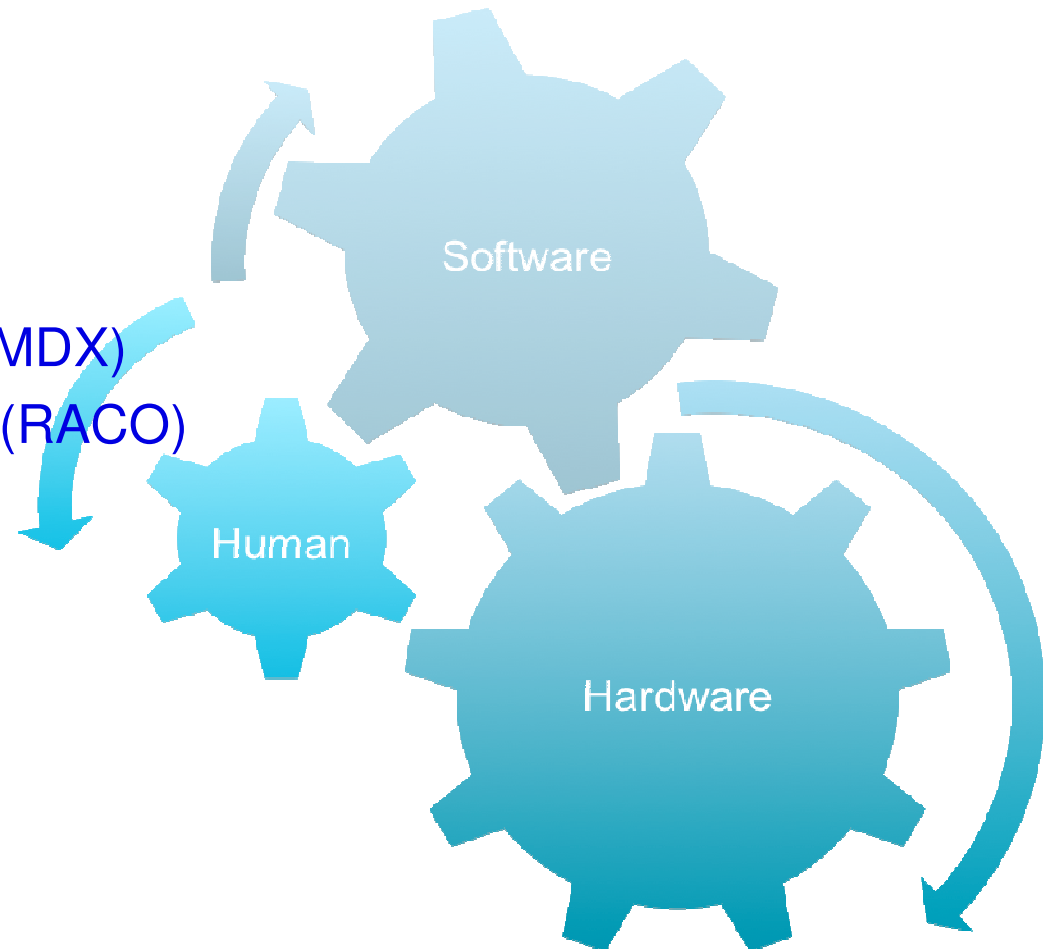
✓ *Software*

- ETD-db (TDR)
- DSpace (RECERCAT, MDX)
- Open Journal Systems (RACO)
- CONTENTdm (MDC)

✓ *Hardware*

- 2001 → 2010

✓ Conclusiones



- ✓ Se ha hecho un estudio inicial para cada repositorio con el objetivo de conocer qué *software* sería el más adecuado en cada caso
- ✓ La mayoría de repositorios son *software* de código abierto, hecho que ha permitido, entre otras ventajas, la adaptación a las necesidades concretas de cada proyecto
- ✓ Interoperatividad a través de OAI-PMH
 - Proveedores de datos
 - Proveedores de servicios



✓ Gestión de contenidos → ETD-db

- Proyecto de colaboración entre:
 - Graduate School at Virginia Tech
 - Digital Library and Archives (DLA)
 - National Digital Library of Thesis and Dissertations (NDLTD)
- 1 instancia de ETD-db para cada institución



✓ Indexación y publicación → Glimpse y Webglimpse

- Consultas globales al repositorio



✓ Módulo para la exportación por OAI-PMH y de estadísticas

✓ Desde 2007, recolector de tesis españolas

→ Módulo MetaIndex + librerías X-Server de Ex Libris



DSpace (2005, RECERCAT; 2009, MDX)



- ✓ *Software* para repositorios digitales
- ✓ Desarrollado por MIT i HP
Actualmente bajo DURASPACE
- ✓ *Software* de código abierto
- ✓ Unas 700 instalaciones
- ✓ OLD Dominion University *plugin* para la recolección por OAI-PMH
- ✓ Desarrollos propios
 - Módulo de estadísticas
 - Herramienta de importación desde REPEC



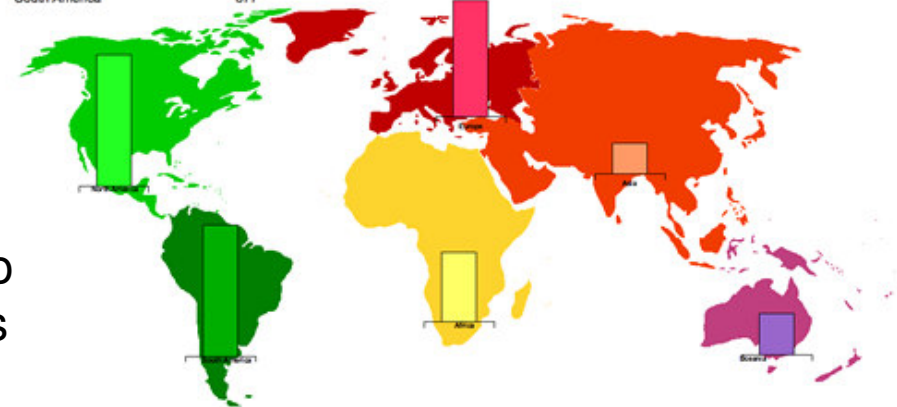
Open Journal Systems (2006, RACO)



- ✓ Herramienta para la gestión editorial + repositorio
- ✓ Public Knowledge Project (PKP)
 - Simon Fraser University Library
 - School of Education at Stanford University
 - Simon Fraser University
 - Faculty of Education at the University of British Columbia
- ✓ *Software* de código abierto
- ✓ Casi 3.000 instalaciones (1-N revistas)
- ✓ Desarrollos propios
 - Módulo de publicación simplificado
 - Módulo de gestión de instituciones
 - Módulo de estadísticas

PKP
PUBLIC
KNOWLEDGE
PROJECT

Journals Using OJS by Continent March 2009
Number of Journals



- ✓ Gestor de colecciones digitales
- ✓ Online Computer Library Center (OCLC)
- ✓ Licencia privativa basada en el número de documentos
- ✓ Unas 1.400 instalaciones
- ✓ Las colecciones pueden estar formadas por diferentes tipos de ficheros: imágenes, pdf, audio, video...
- ✓ Extensión OCR para colecciones digitalizadas. Permite incluir el *transcript*



CONTENTdm



- ✓ Adaptación del *software* de base
 - Diseño gráfico, traducciones y adaptación de textos
 - Configuración de parámetros (*stop words*...)
 - Corrección de errores
 - Ampliación de funcionalidades (diacríticos...)

- ✓ Instalación de módulos o *plugins*
 - Para la exportación o importación OAI-PMH, Manakin...

- ✓ Desarrollo de módulos propios
 - Estadísticas, ingesta masiva de datos, módulos de RACO...

- ✓ *Software* complementario
 - Intranets, páginas de novedades, herramientas de monitorización...

✓ Introducción

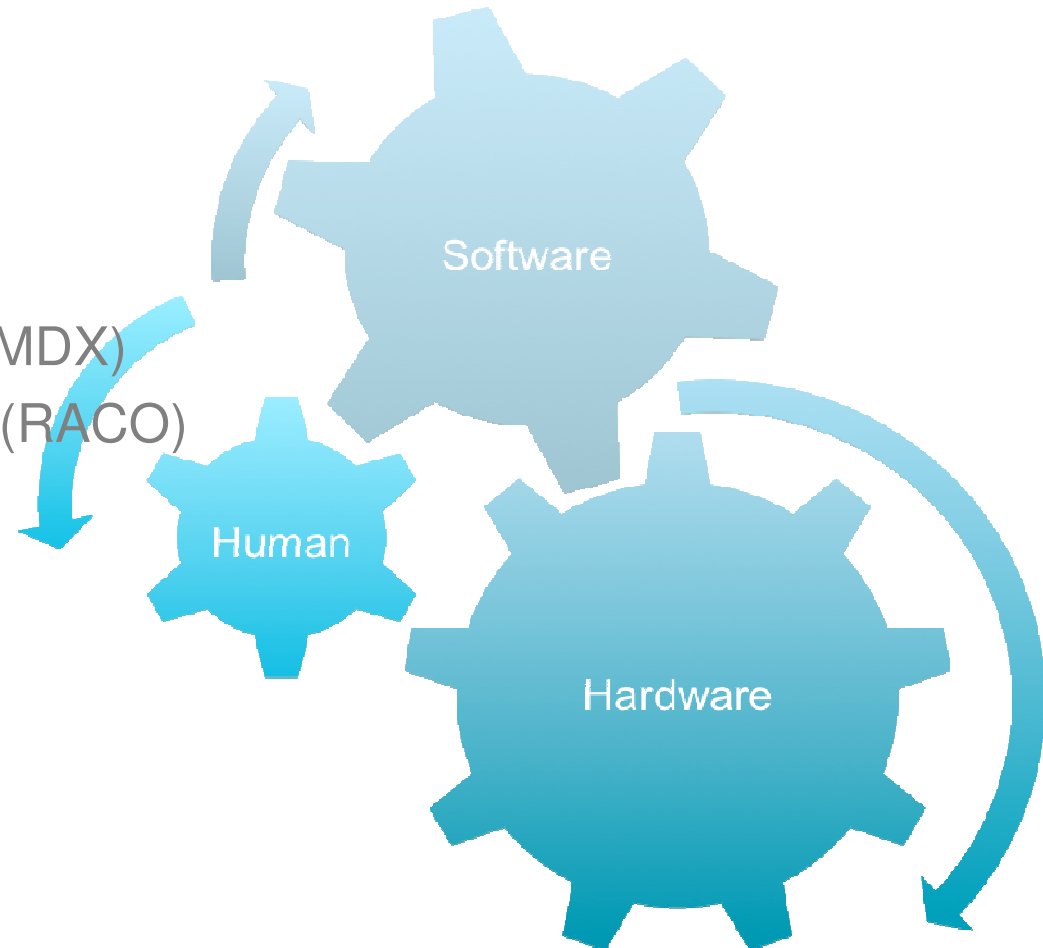
✓ *Software*

- ETD-db (TDR)
- DSpace (RECERCAT, MDX)
- Open Journal Systems (RACO)
- CONTENTdm (MDC)

✓ *Hardware*

- 2001 → 2010

✓ Conclusiones



2001



2005

RECERCAT

2006



2006



2009



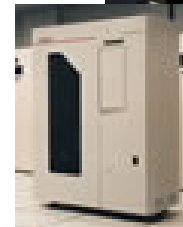
✓ ~80 GB y 7.690 consultas

✓ HP V2500 (cálculo)

- 16 procesadores
- 4 GB de memoria
- 227 GB de disco

✓ StorageTek TimberWolf 9740

- 2,7 TB en cintas



2001



2005

RECERCAT

2006



2006



2009



✓ En 2003 ~100 GB y 278.710 consultas

✓ HP rp5430 (servidor)

- 2 procesadores PA8700
- 704 GB de memoria

✓ HP EVA V.2 (cabina de discos)

- 2,8 TB de disco

✓ StorageTek TimberWolf 9740

- 5 TB en cintas



2001



2005



2006



2006



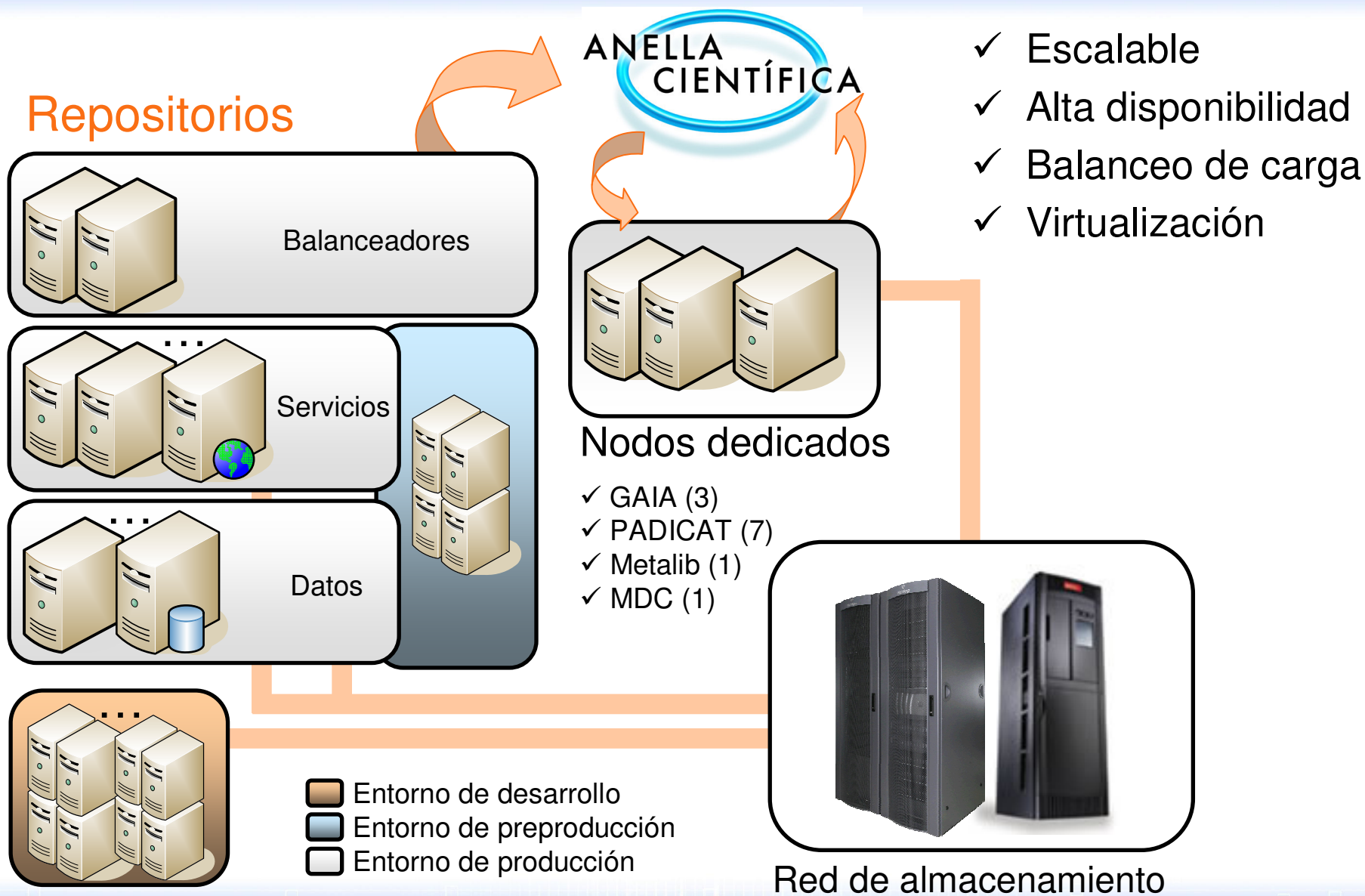
2009



- ✓ En 2010 ~2 TB y 8.019.930 consultas (TDR+RECERCAT+RACO)
- ✓ Clúster de e-Información
 - 32 nodos Proliant DL360
 - 291 GB de memoria
- ✓ NetApp FAS3140
 - 100 TB de disco
- ✓ ADIC Scalar i2000
 - 156 TB en cintas



Hardware en 2010



- ✓ Escalable
- ✓ Alta disponibilidad
- ✓ Balanceo de carga
- ✓ Virtualización

✓ Introducción

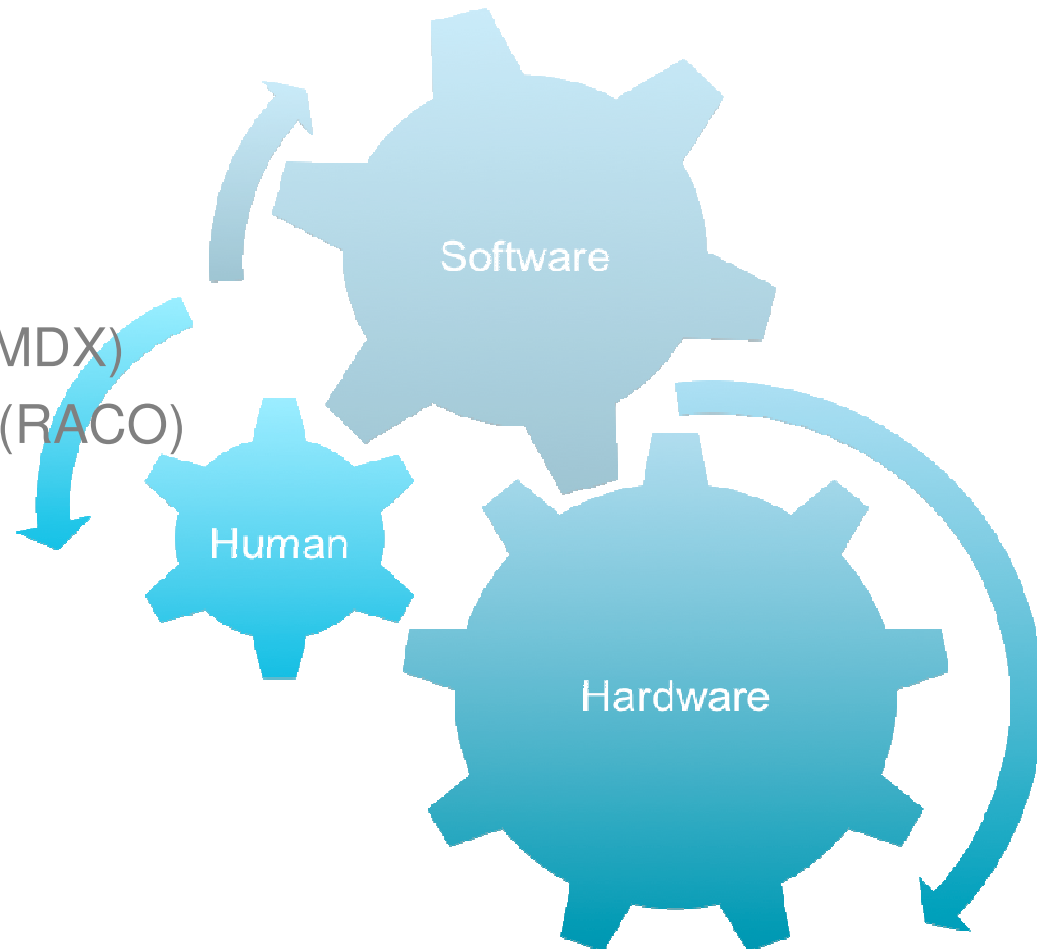
✓ *Software*

- ETD-db (TDR)
- DSpace (RECERCAT, MDX)
- Open Journal Systems (RACO)
- CONTENTdm (MDC)

✓ *Hardware*

- 2001 → 2010

✓ Conclusiones



Con la **cooperación** se consigue:

- ✓ Adopción de procedimientos comunes
- ✓ Mayor visibilidad
- ✓ ...
- ✓ **Economías de escala**
 - En *software*
 - En *hardware*
 - En el equipo humano

