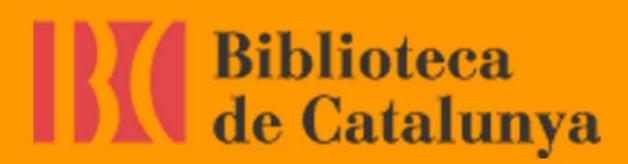
Patrimoni Digital de Catalunya, experiencias del primer año

N. Torres, J. Cambras, J. Prats, X. Torelló, R. de la Vega Centre de Supercomputació de Catalunya



Patrimoni Digital de Catalunya (PADICAT) es un proyecto de la Biblioteca Nacional de Catalunya. Este consiste en Capturar, procesar y dar acceso permanente a toda la producción cultural, científica y de carácter general

catalana producida en formato digital. En definitiva, el objetivo de PADICAT es archivar la Web catalana. El proyecto, iniciado en junio de 2005, cuenta con la colaboración tecnológica del Centre de Supercomputació de Catalunya

Storage Area Network

(CESCA) y con el soporte de la Generalitat de Catalunya. En algunos países se llama "repositorios digitales nacionales" o "archivos

Web" a los proyectos similares, siendo los más conocidos el gigante Internet Archive,

el australiano Pandora, el sueco Kulturarw3 o el Netarchive danés.

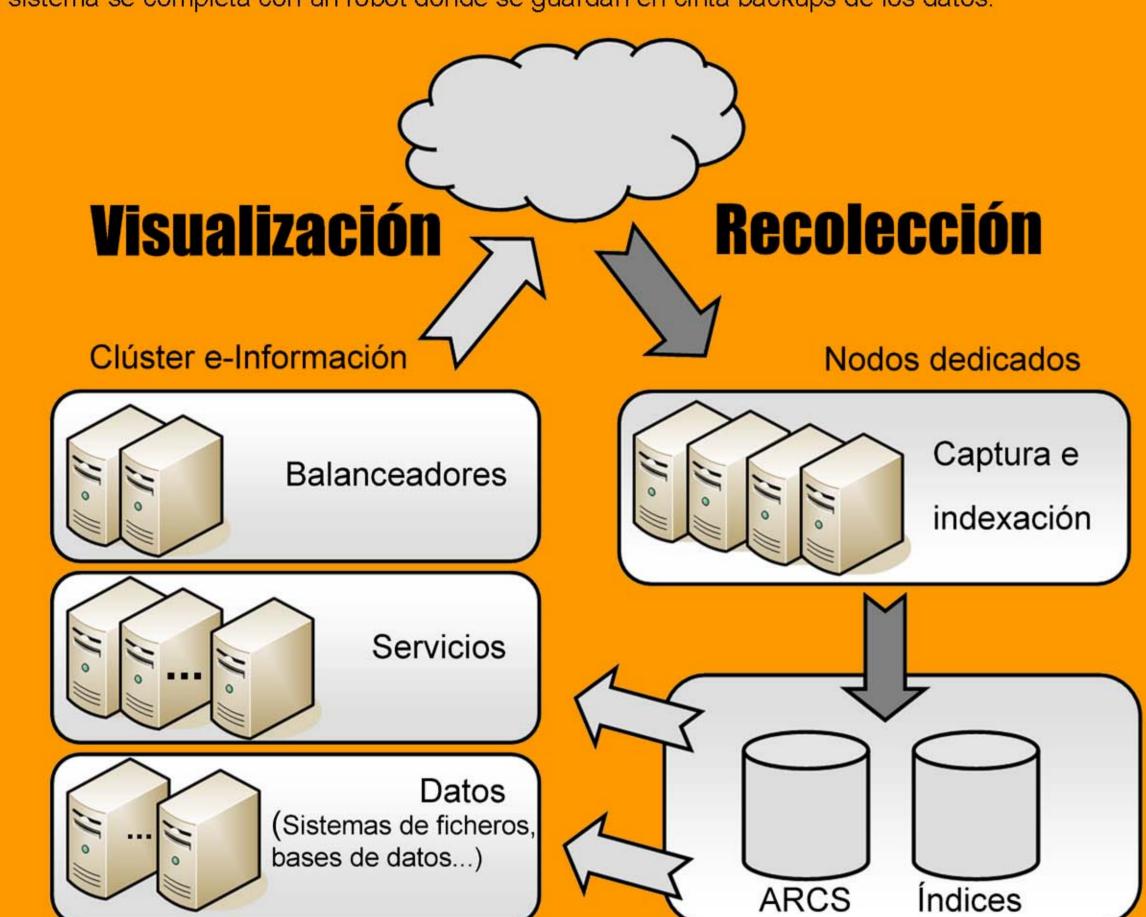
CENTRE DE SUPERCOMPUTACIO



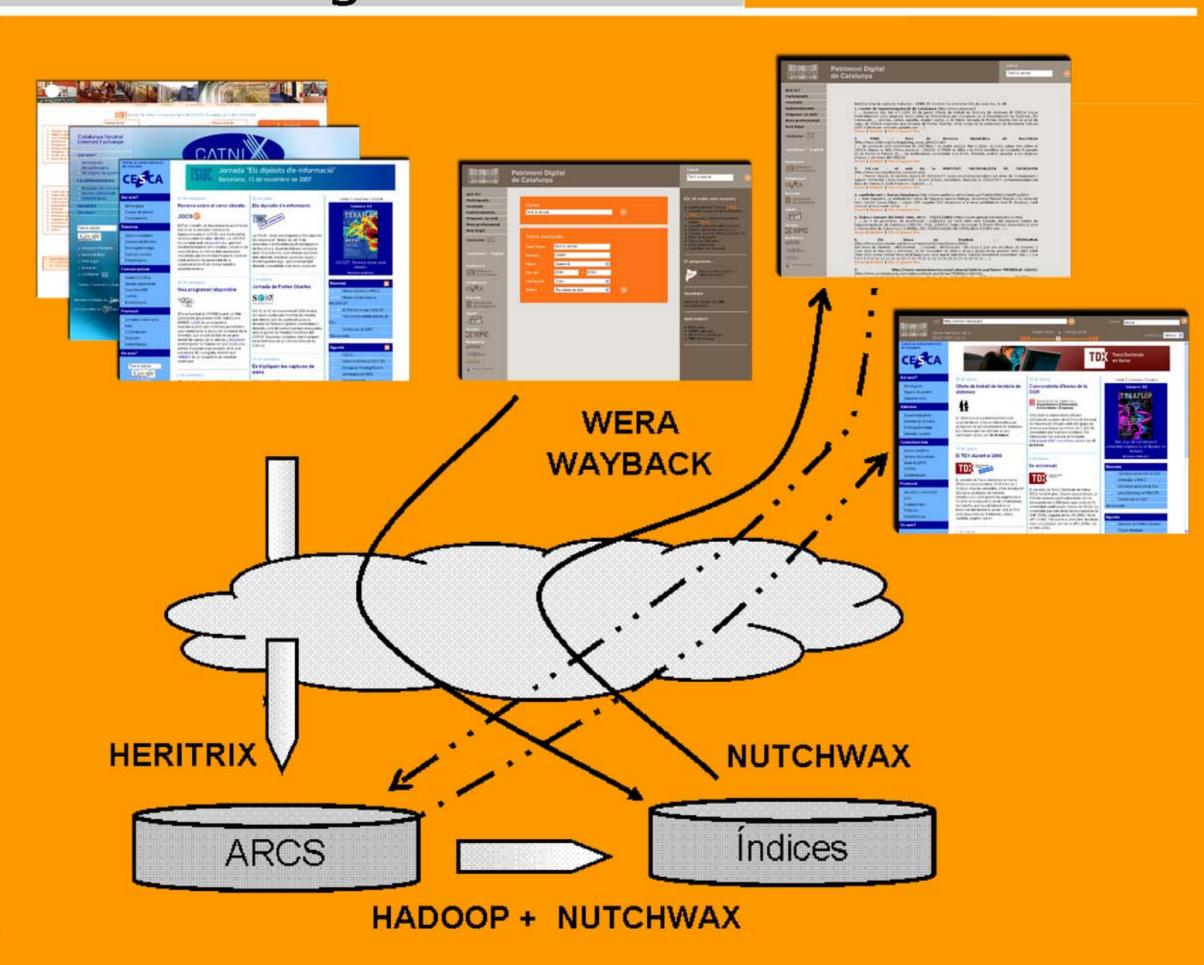
Arquitectura

Se dispone de cuatro nodos dedicados HP ProLiant DL360 G4p encargados de las funciones de recolección e indexación de las webs. Dichos nodos se encuentran virtualizados mediante Xen para, de este modo, adaptar mejor los recursos de acuerdo a las necesidades.

Por otro lado, de la búsqueda y visualización de resultados en la interfaz web se encarga un clúster Linux de alta disponibilidad compartido con otros repositorios cooperativos de e-información. Los nodos están conectados mediante fibra a una Storage Area Network (SAN) y el sistema se completa con un robot donde se guardan en cinta backups de los datos.



Tecnología



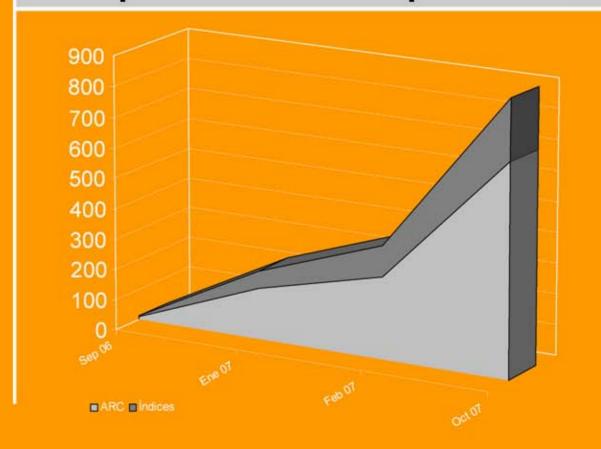
Estadísticas

813 webs 2.409 capturas 24 M de ficheros 900 GB de espacio 290 convenios 2 recopilaciones

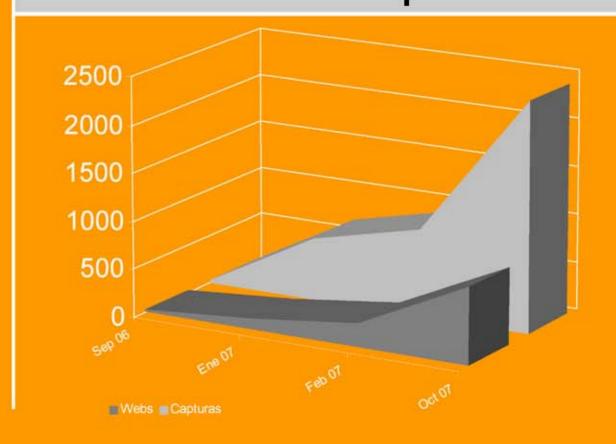
Primer año

En este primer año, una amplia selección de sitios web representativos del entramado que conforma la sociedad civil catalana, tales como ayuntamientos, universidades, asociaciones profesionales, culturales o deportivas, partidos políticos, empresas y medios de comunicación, han Jlegado ya a acuerdos de COOPERACION, unos 290 en total; asimismo, se han recibido más de 350 propuestas de otros recursos y se han focalizado los recursos digitales asociados a dos acontecimientos las elecciones al Parlament de Catalunya de 2006 y las elecciones municipales.

Espacio ocupado (GB)

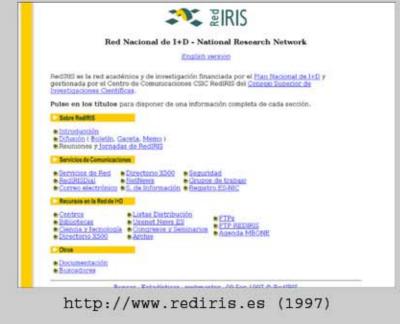


Recursos capturados



En 2009...

PADICAT es pionero en España, y para enero de 2009, con diversas capturas de 10.000 webs y 500 millones de ficheros en 30 TB, espera ser un réferente en Europa. También se pretende superar los 500 acuerdos de colaboracion con instituciones representativas de la sociedad civil catalana.









http://www.google.com (1998)



http://www.napster.com (2000)



http://www.apache.og (1996)





http://www.tve.es (1999)



http://www.nintendo.com (1996)



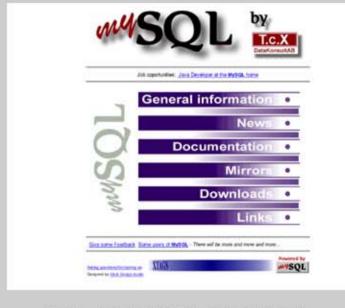


http://www.bnc.es (2002)

ePege RecentChanges Ereferences can edit this page right next lits a free, community project	
come to Wikipedia, a collaborative project to produce a complete erg 2001 and already have over 8,000 articles. We want to make any page-copyedit, write a little, write a lot. See the Wikipedia Por r questions. If you're vesting Wikipedia for the first time westome! Free Documentation License.	over 100,000, so let's get to work-anyone or of for information on how to edit pages and
Philosophy, Mathematics, and Natural Science Astronomy and Astrophysics - Budger - Chamstry - Earth Scienc Science - Statistics	os - Mathematics - Philosophy - Physics -
Social Sciences Anomalous Phenomena - Anthropology - Archaeology - Countries History - History of Science and Technology - Language - Linguis	
Applied Arts and Sciences Agriculture - Architecture - Business and Industry - Communicat Equals and Consumer Science - Health Sciences - Larr - Library s Technology - Transport	
Culture Classics - Critical Theory - Dance - Entertainmann - Film - Game Easting - Performing Arts - Recreation - Religion - Sculpture - E Arts and Design	
Other Category Schemes About Wikipedia category schemes - Library of Congress catalog	
arranged by topic - Year in Review - Historical acopersaries - Ba International Wikipedias	
	e (Hannu) - [German (Deutschi) - [Esperant anese (Whongo)] - [Portuguese (Portuguese)



http://www.wordperfect.com (1996)



http://www.mysql.com (1998)



http://www.archive.org (1997)



http://www.tvcatalunya.com (1998)



http://www.amazon.com (2000)

Las imágenes superiores pertenecen a Internet Archive, que tiene como objetivo archivar la web mundial. El alcance global de dicho proyecto imposibilita la captura detallada a nivel local. Las imágenes inferiores corresponden a PADICAT, preservando la realidad catalana.



http://www.macba.es (2007)



http://www.tricicle.com (2007)



http://www.parlament-cat.net (2006)



http://www.domini.cat (2007)



http://www.gencat.net (2007)



PATRIMONI DIGITAL DE CATALUNYA

http://www.padi.cat