



# HiberActive

## Pro-Active Archiving of Web References from Scholarly Articles

**Martin Klein**, Harihar Shankar, Herbert Van de Sompel  
Los Alamos National Laboratory  
**@martinklein**

Richard Wincewicz  
Edina, University of Edinburgh

<http://hiberlink.org> #hiberlink  
<http://mementoweb.org> #memento  
<http://openarchives.org/rs/> #resourcesync



# Link Rot

OR | Open Repositories

Jun 09, 2014

Home Steering Committee Contact Us

## Open Repositories

### Welcome to Open Repositories

Repositories increasingly play a pivotal role in the emerging information landscape. Through a format that blends open user group meetings for DSpace, Fedora, and Eprints, followed by general conference sessions that cover cross-cutting and overarching issues, Open Repositories attempts to create opportunities to explore the challenges faced by user communities and others in today's world.

The many repository platforms available today are changing the nature of scholarly communication. Institutions such as universities, research laboratories, publishers, libraries, and commercial organizations are creating innovative repository-based systems that address the entire lifecycle of information — from supporting the creation and management of digital content, to enabling use, re-use, and interconnection of information, to ultimately ensuring long-term preservation and archiving.

### OR13: Charlottetown, Prince Edward Island July 8-12, 2013

The 2013 International Conference on Open Repositories (OR13) was held in Charlottetown, Prince Edward Island, Canada on July 8-12, 2013.

### Future Conferences

**Open Repositories 2014:: Helsinki, Finland**  
9th International Conference on Open Repositories  
June 9th -- 13th, 2014

### Past Conferences

**Open Repositories 2013 :: Charlottetown, Prince Edward Island, Canada**  
8th International Conference on Open Repositories  
July 8th -- 12th, 2013

**Open Repositories 2012 :: Edinburgh, Scotland**  
7th International Conference on Open Repositories  
July 9th -- 13th, 2012

**Open Repositories 2011 :: Austin, Texas USA**  
6th International Conference on Open Repositories  
June 6th -- 11th, 2011



**HiberActive** - Martin Klein et al.  
Open Repositories, Helsinki, Finland, June 10th 2014



**Los Alamos**  
NATIONAL LABORATORY  
THE UNIVERSITY of EDINBURGH

# Link Rot

conference sessions that cover cross-cutting and overarching issues, Open Repositories attempts to create opportunities to explore the challenges faced by user communities and others in today's world.

The many repository platforms available today are changing the nature of scholarly communication. Institutions such as universities, research laboratories, publishers, libraries, and commercial organizations are creating innovative repository-based systems that address the entire lifecycle of information — from supporting the creation and management of digital content, to enabling use, re-use, and interconnection of information, to ultimately ensuring long-term preservation and archiving.

## OR13: Charlottetown, Prince Edward Island July 8-12, 2013

The [2013 International Conference on Open Repositories \(OR13\)](#) was held in Charlottetown, Prince Edward Island, Canada on July 8-12, 2013.

The OR2013 conference theme was **Use, Reuse, Reproduce**. One of the most important roles of repositories is to enable greater use and reuse of their contents— whether those contents are library collections, scholarly articles, research data, or software—and metadata. The notion of use and reuse can be extended to repository infrastructure as well. Many repositories are based on open source software that can be freely reused and adapted to serve local needs; other efforts are also emerging both in conjunction with and outside traditional repository platforms to encourage discipline or community specific reuse and sharing of software, services, and infrastructure. In addition there is a growing interest and need to document and share the code and workflows used to produce research results — particularly in computationally intensive fields — in order to promote reproducible research.

Visit the [conference website](#) to learn more.

To get announcements about this and future Open Repositories events, you may join the OR announcements mailing list here: <http://groups.google.com/group/open-repositories>.

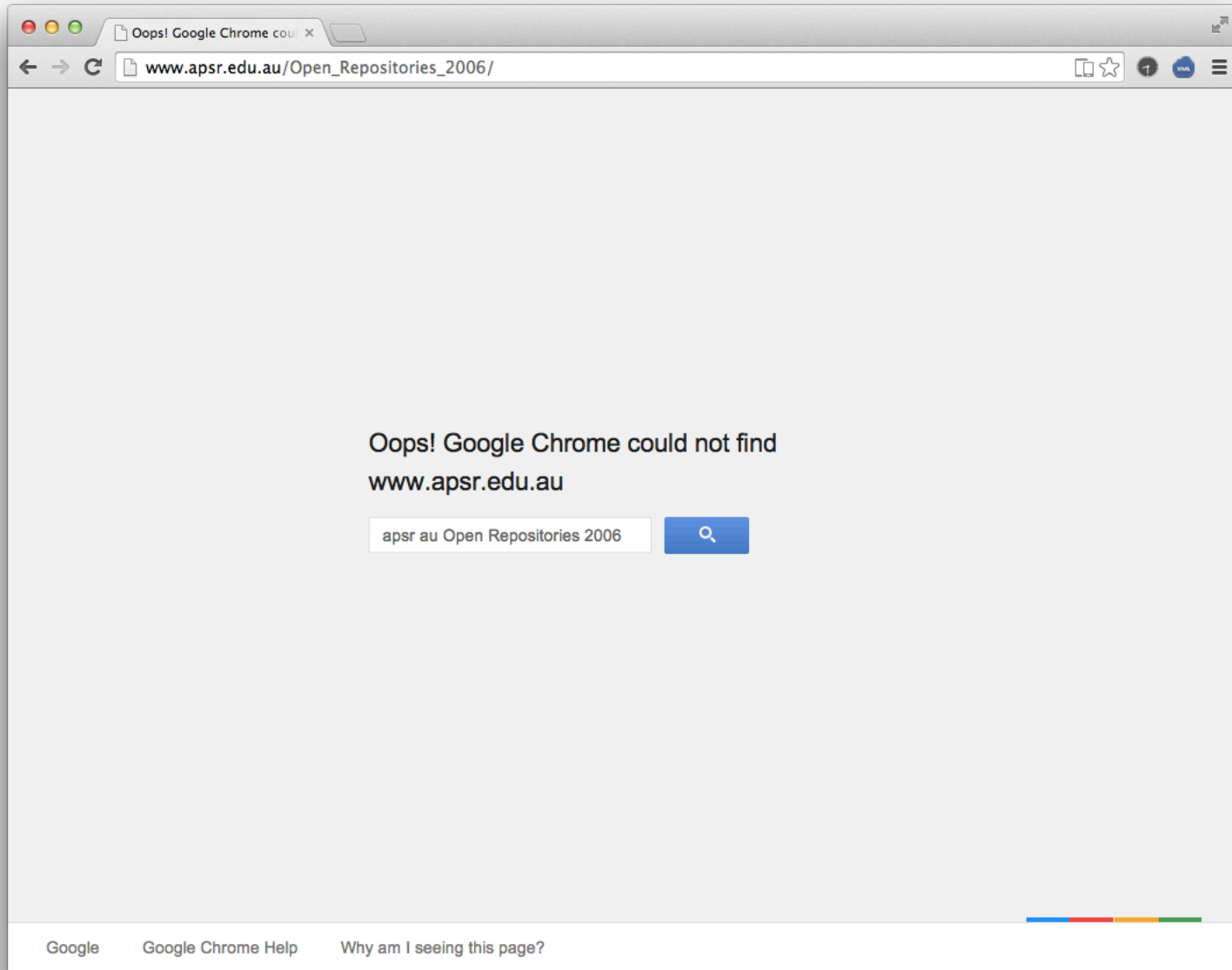
[f](#) [t](#) [g+](#) [+](#)

### Past Conferences

- Open Repositories 2013 :: Charlottetown, Prince Edward Island, Canada**  
8th International Conference on Open Repositories  
July 8th – 12th, 2013
- Open Repositories 2012 :: Edinburgh, Scotland**  
7th International Conference on Open Repositories  
July 9th – 13th, 2012
- Open Repositories 2011 :: Austin, Texas USA**  
6th International Conference on Open Repositories  
June 6th – 11th, 2011
- Open Repositories 2010 :: Madrid, Spain**  
5th International Conference on Open Repositories  
July 6th – 9th, 2010
- Open Repositories 2009 :: Atlanta, Georgia USA**  
4th International Conference on Open Repositories  
May 18th – 21st, 2009
- Open Repositories 2008 :: Southampton, United Kingdom**  
3rd International Conference on Open Repositories  
April 1st – 4th, 2008
- Open Repositories 2007 :: San Antonio, Texas USA**  
2nd International Conference on Open Repositories  
January 23-26, 2007
- Open Repositories 2006 :: Sydney, Australia**  
1st International Conference on Open Repositories  
January 31 – February, 3, 2006



# Link Rot



**HiberActive** - Martin Klein et al.  
Open Repositories, Helsinki, Finland, June 10th 2014



**Los Alamos**  
NATIONAL LABORATORY  
THE UNIVERSITY of EDINBURGH

# Archived Version

INTERNET ARCHIVE  
WayBackMachine

http://www.aprs.edu.au/Open\_Repositories\_2006/

37 captures  
23 Nov 05 - 18 Jun 12

NOV FEB MAR Close X  
2005 2006 2007 Help ?

Australian Partnership for Sustainable Repositories

## Open Repositories 2006

Partner's Area: Login | Help

### APSR

- Home
- Program
- Speakers
- Events:
  - Forum: Well-Integrated Repository
  - Symposium: Managing Openness
- User Groups
  - DSpace
  - EPrints
  - Fedora
- Registration
- Costs
- Social events
- Accommodation
- Venue

### Open Repositories 2006, 31 January - 3 February, 2006

Registrations for Open Repositories 2006 will close at 5pm AEST on Monday January 23. After that time, contact **APSR**.

#### Program now available

- Forum: The Well-Integrated Repository
- Symposium: Managing Openness in Digital Repositories
- DSpace User Group Meeting: Building Communities with DSpace
- Fedora User Group Meeting
- EPrints User Group Meeting

Guest speakers will include Sandy Payette of Cornell Computing and Information Science, Mackenzie Smith of MIT Libraries Digital Library Research Group, Robert Tansley of HP Labs, John Townsend of HarvestRoad, Les Carr and Tim Miles Board from EPrints and other international experts in digital repository architecture and broader e-research frameworks.

APSR is delighted to congratulate the recipients of two Hewlett Packard awards, **Dr. M.G. Sreekumar** from India and **Ms Edna Nyika** from Tanzania. [More](#)

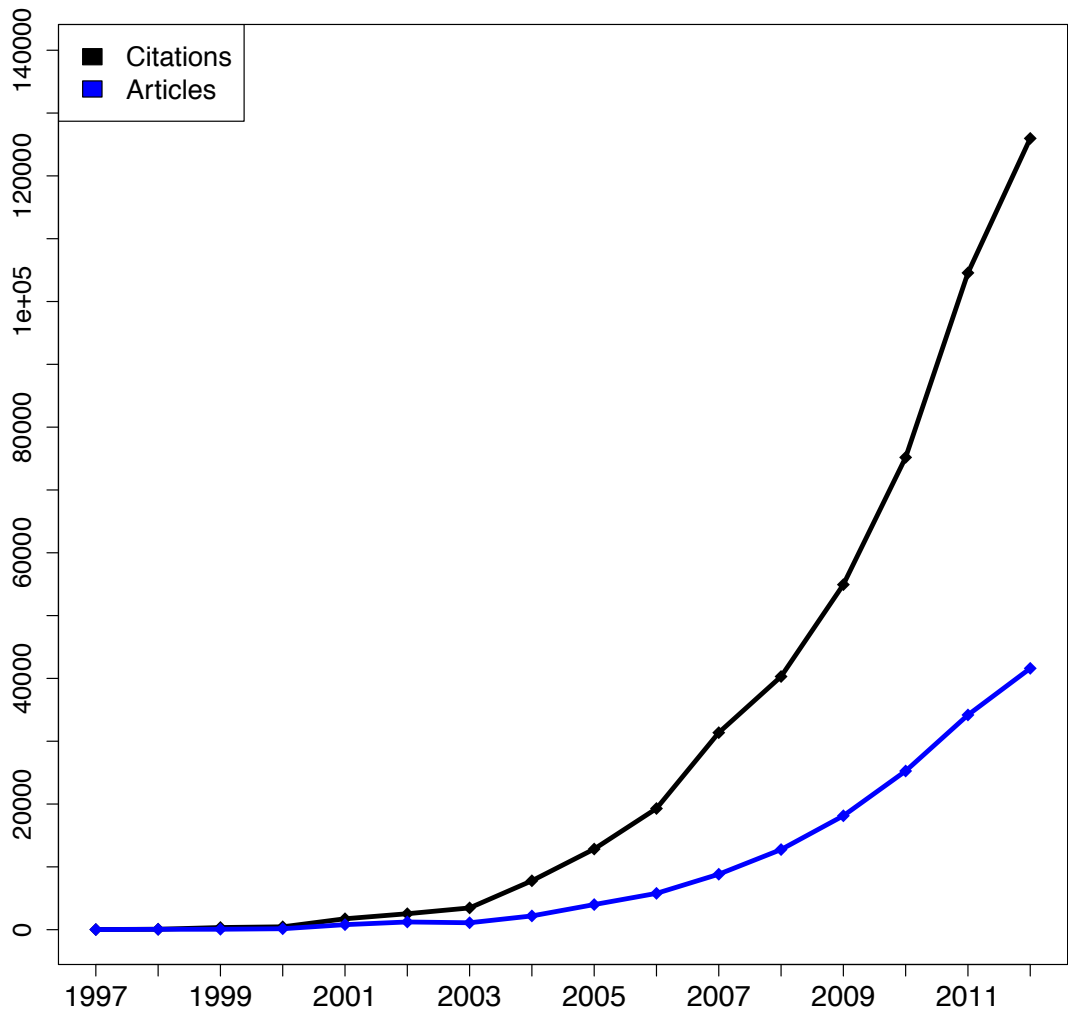
For an overview of the order of events, take a look at the program [schedule](#).

**Open Repositories 2006** will be held at the University of Sydney. This is an interlinked set of events collaboratively convened by APSR with the ARROW project, MAMS, Macquarie University, and the University of Technology Sydney.

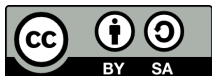
Those considering coming "down under" from far off places might consider extending their stay to attend the



# Scholarly Articles Increasingly Link to Web Resources



URIs extracted from PubMed Central articles



## **D-Lib Magazine September 2004**

Volume 10 Number 9

ISSN 1082-9873

### **Rethinking Scholarly Communication**

#### **Building the System that Scholars Deserve**

[Herbert Van de Sompel](#)

Los Alamos National Laboratory, Research Library  
<herbertv@lanl.gov>

[Sandy Payette](#)

Cornell University, Computing and Information Science  
<payette@cs.cornell.edu>

[John Erickson](#)

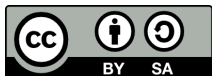
Hewlett-Packard Laboratories, Digital Media Systems Lab  
<john.erickson@hp.com>

[Carl Lagoze](#)

Cornell University, Computing and Information Science  
<lagoze@cscornell.edu>

[Simeon Warner](#)

Cornell University, Computing and Information Science  
<simeon@cs.cornell.edu>



## References

**!Exist**

Atkins, D. et al.. 2003. National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, *Revolutionizing Science and Engineering through Cyber-infrastructure*, <[http://www.communitytechnology.org/nsf\\_ci\\_report/](http://www.communitytechnology.org/nsf_ci_report/)>.

**Archived**

**Exist**

Brody, T., Kampa, S., Harnad, S., Carr, L. and Hitchcock, S. 2003. Digitometric Services for Open Archives Environments. In *Proceedings of European Conference on Digital Libraries 2003*, pages pp. 207-220, Trondheim, Norway. <<http://eprints.ecs.soton.ac.uk/archive/00007503/>>.

**Archived**

Frey, J., De Roure, D. and Carr, L. 2002. *Publication at Source: Scientific Communication from a Publication Web to a Data Grid*. <<http://eprints.ecs.soton.ac.uk/archive/00007852/>>.

Henry, G. 2003. On-line publishing in the 21-st Century: Challenges and Opportunities. *D-Lib Magazine*, Volume 9, Issue 10. <[doi:10.1045/october2003-henry](https://doi.org/10.1045/october2003-henry)>.

**!Exist**

Lynch, C. 2003. Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL Bimonthly Report* 226. February 2003, <<http://www.arl.org/newsltr/226/ir.html>>.

**Archived**

**!Exist**

Payette, S., and Staples, T. 2002. The Mellon Fedora Project: Digital Library Architecture Meets XML and Web Services. *European Conference on Research and Advanced Technology for Digital Libraries*, Rome, Italy, September 2002. <<http://www.fedora.info/documents/ecdl2002final.pdf>>.

**!Archived**

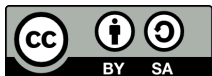
Pöschl, U. 2004. Interactive Journal Concept for Improved Scientific Publishing and Quality Assurance. *Learned Information*, Volume 17, Number 2, pp 105-113. <[doi:10.1087/095315104322958481](https://doi.org/10.1087/095315104322958481)>.

Reich, V. and Rosenthal, D. 2001. LOCKSS: A Permanent Web Publishing and Access System. *D-Lib Magazine*, Volume 7, Issue 6. <[doi:10.1045/june2001-reich](https://doi.org/10.1045/june2001-reich)>.

**Exist**

Roosendaal, H., and Geurts, P. 1997. Forces and functions in scientific communication: an analysis of their interplay. *Cooperative Research Information Systems in Physics*, August 31 — September 4 1997, Oldenburg, Germany. <<http://www.physik.uni-oldenburg.de/conferences/crisp97/roosendaal.html>>.

**Archived**



**HiberActive** - Martin Klein et al.

Open Repositories, Helsinki, Finland, June 10th 2014



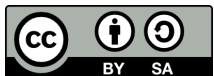
**Los Alamos**  
NATIONAL LABORATORY  
THE UNIVERSITY of EDINBURGH



# Ubiquity of the problem

- Harvard Law Review
  - 75% of links are dead
- top 1% Impact Factor Journals
  - 10% of links dead just 15 months after publication
- US Supreme Court decisions
  - 29% of links dead
  - **49% of links do not point to its original target**

<http://www.nytimes.com/2013/09/24/us/politics/in-supreme-court-opinions-clicks-that-lead-nowhere.html>



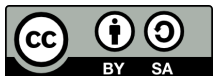
**HiberActive** - Martin Klein et al.  
Open Repositories, Helsinki, Finland, June 10th 2014



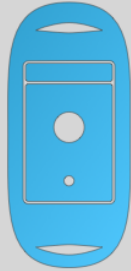
**Los Alamos**  
NATIONAL LABORATORY  
THE UNIVERSITY of EDINBURGH

# Entrance HiberActive

- Perfect scenario:
  - All authors archive all their cited URIs e.g., with Hiberlink's Zotero extension
- Current scenario:
  - Mostly accidental archiving of cited URIs
- HiberActive scenario:
  - Repositories actively support archiving
  - Utilize existing pro-active web archiving infrastructure
  - Provide value-added service based on archived data



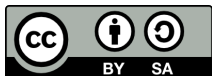
REPOSITORY



HIBERACTIVE



WEB ARCHIVE



**HiberActive** - Martin Klein et al.  
Open Repositories, Helsinki, Finland, June 10th 2014





# ResourceSync Notification

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset
  xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
  xmlns:rs="http://www.openarchives.org/rs/terms/">
  <url>
    <loc>http://core.kmi.open.ac.uk/download/pdf/287957.pdf</loc>
    <lastmod>2014-06-08T01:13:07Z</lastmod>
    <rs:md change="created"/>
  </url>
</urlset>
```



# ResourceSync Notification

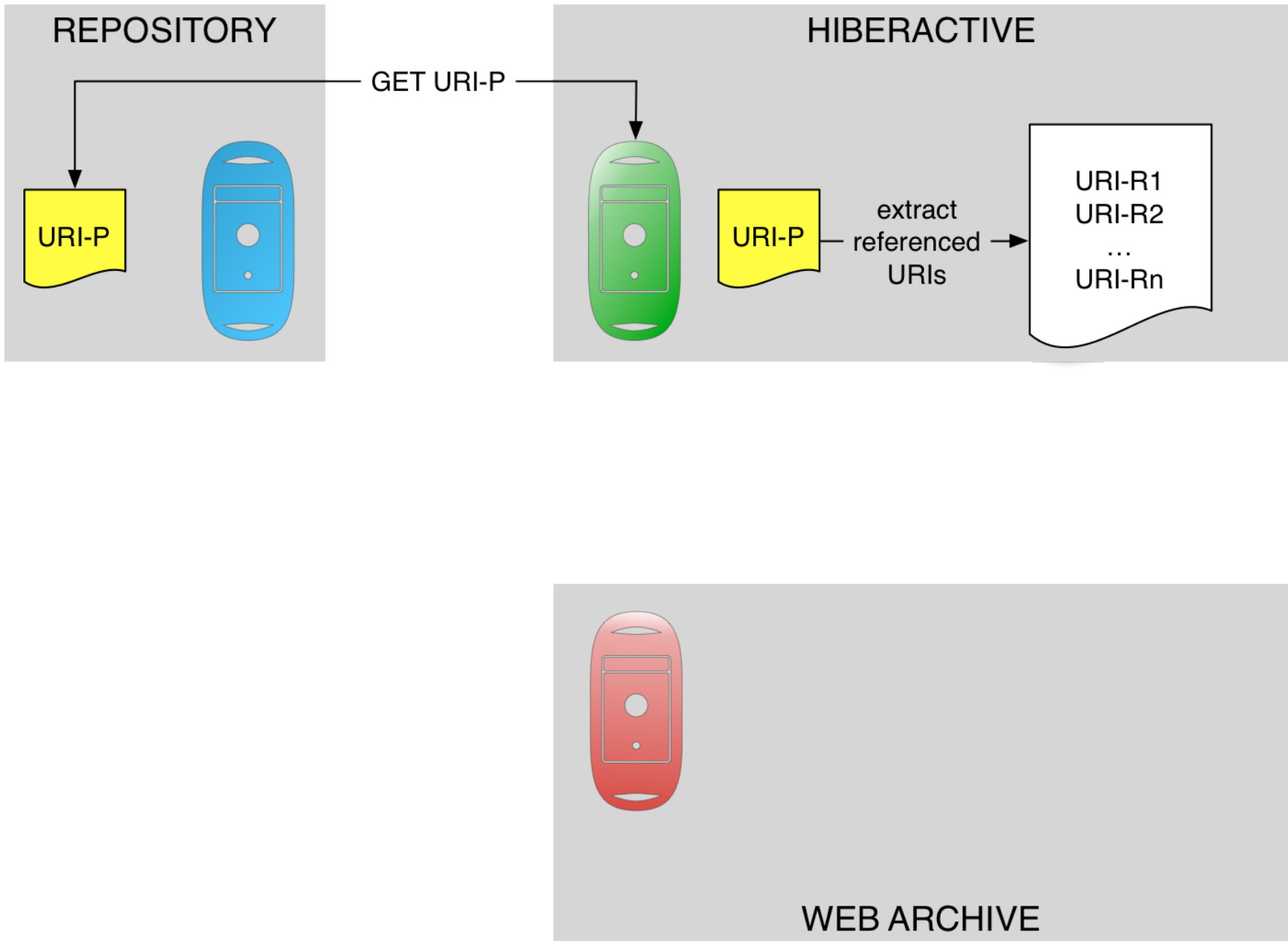
```
<?xml version="1.0" encoding="UTF-8"?>
<urlset
  xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
  xmlns:rs="http://www.openarchives.org/rs/terms/">
  <url>
    <loc>http://core.kmi.open.ac.uk/download/pdf/287957.pdf</loc>
    <lastmod>2014-06-08T01:13:07Z</lastmod>
    <rs:md change="created"/>
  </url>
</urlset>
```



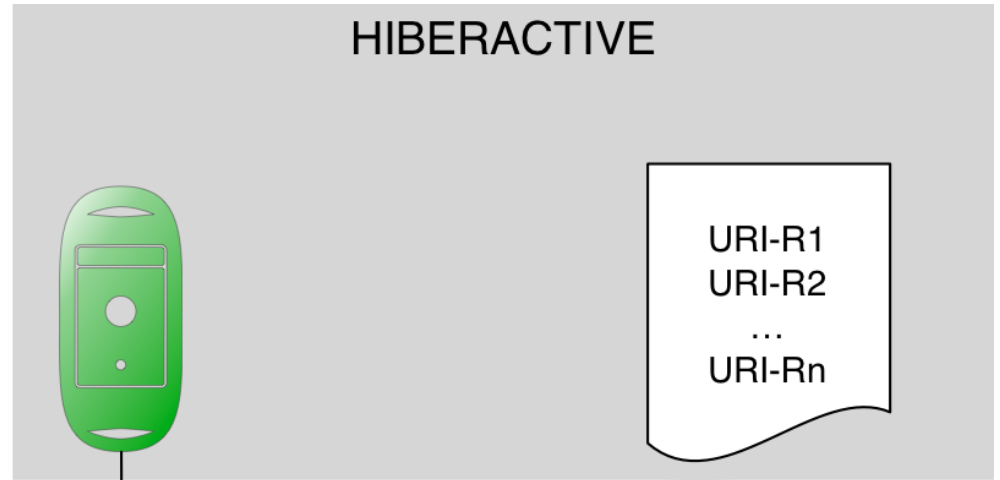
# ResourceSync Notification

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset
  xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
  xmlns:rs="http://www.openarchives.org/rs/terms/">
  <url>
    <loc>http://core.kmi.open.ac.uk/download/pdf/287957.pdf</loc>
    <lastmod>2014-06-08T01:13:07Z</lastmod>
    <rs:md change="created"/>
  </url>
</urlset>
```



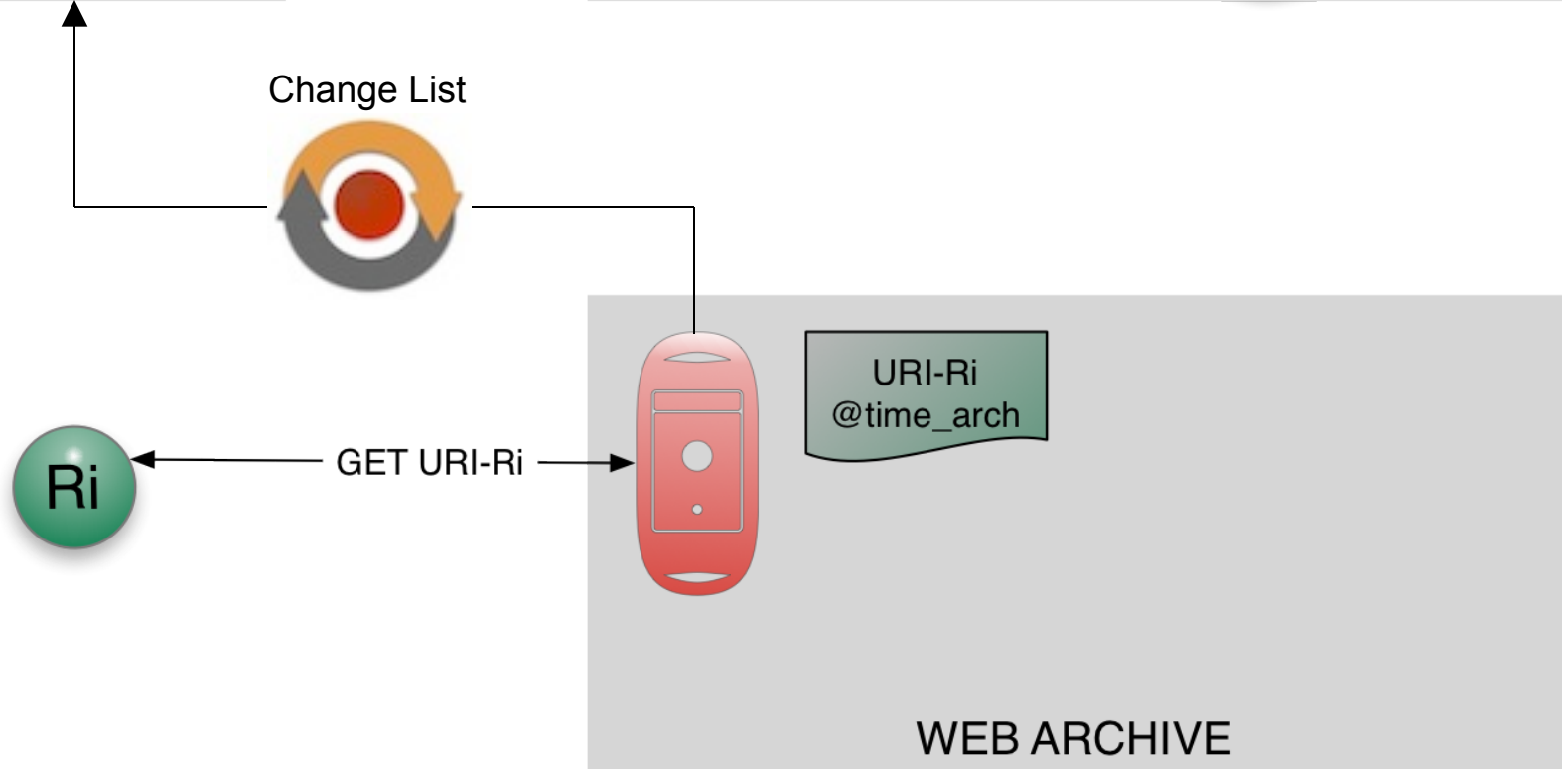
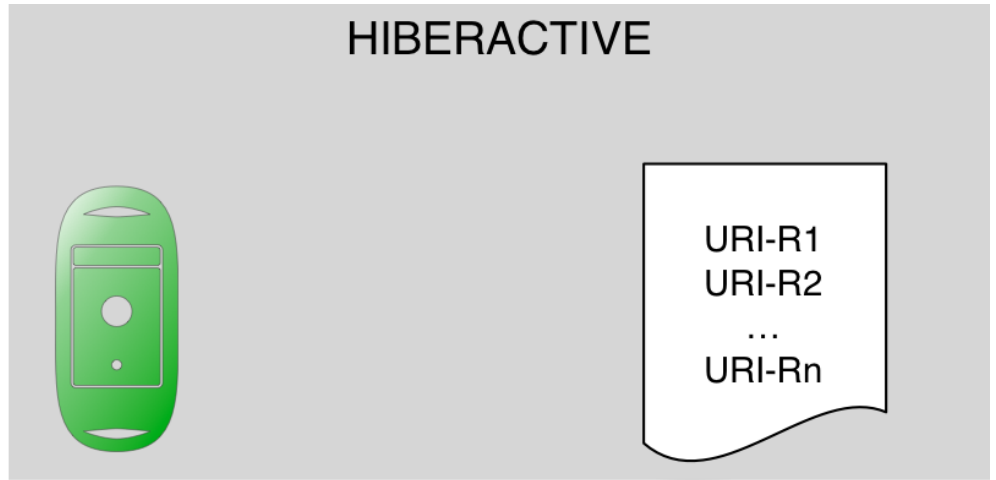






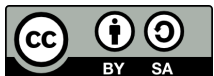
archive  
URI-Ri





# ResourceSync Change List

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset
  xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
  xmlns:rs="http://www.openarchives.org/rs/terms/">
  <rs:ln rel="up" href="http://hiberactive.lanl.gov/changelist.xml"/>
  <rs:md capability="changelist" from="2014-05-01T00:00:00Z"/>
  <url>
    <loc>http://archive.today/FL2as</loc>
    <lastmod>2014-06-08T03:31:19Z</lastmod>
    <rs:md change="created" />
    <rs:ln rel="original" href="http://obofoundry.org/ro/" />
    <rs:ln rel="collection" href="http://core.kmi.open.ac.uk/download/pdf/288127.pdf"/>
  </url>
</urlset
```



# ResourceSync Change List

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset
  xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
  xmlns:rs="http://www.openarchives.org/rs/terms/">
  <rs:ln rel="up" href="http://hiberactive.lanl.gov/changelist.xml"/>
  <rs:md capability="changelist" from="2014-05-01T00:00:00Z"/>
  <url>
    <loc>http://archive.today/FL2as</loc>
    <lastmod>2014-06-08T03:31:19Z</lastmod>
    <rs:md change="created" />
    <rs:ln rel="original" href="http://obofoundry.org/ro/" />
    <rs:ln rel="collection" href="http://core.kmi.open.ac.uk/download/pdf/288127.pdf"/>
  </url>
</urlset
```



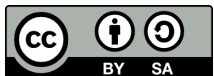
# ResourceSync Change List

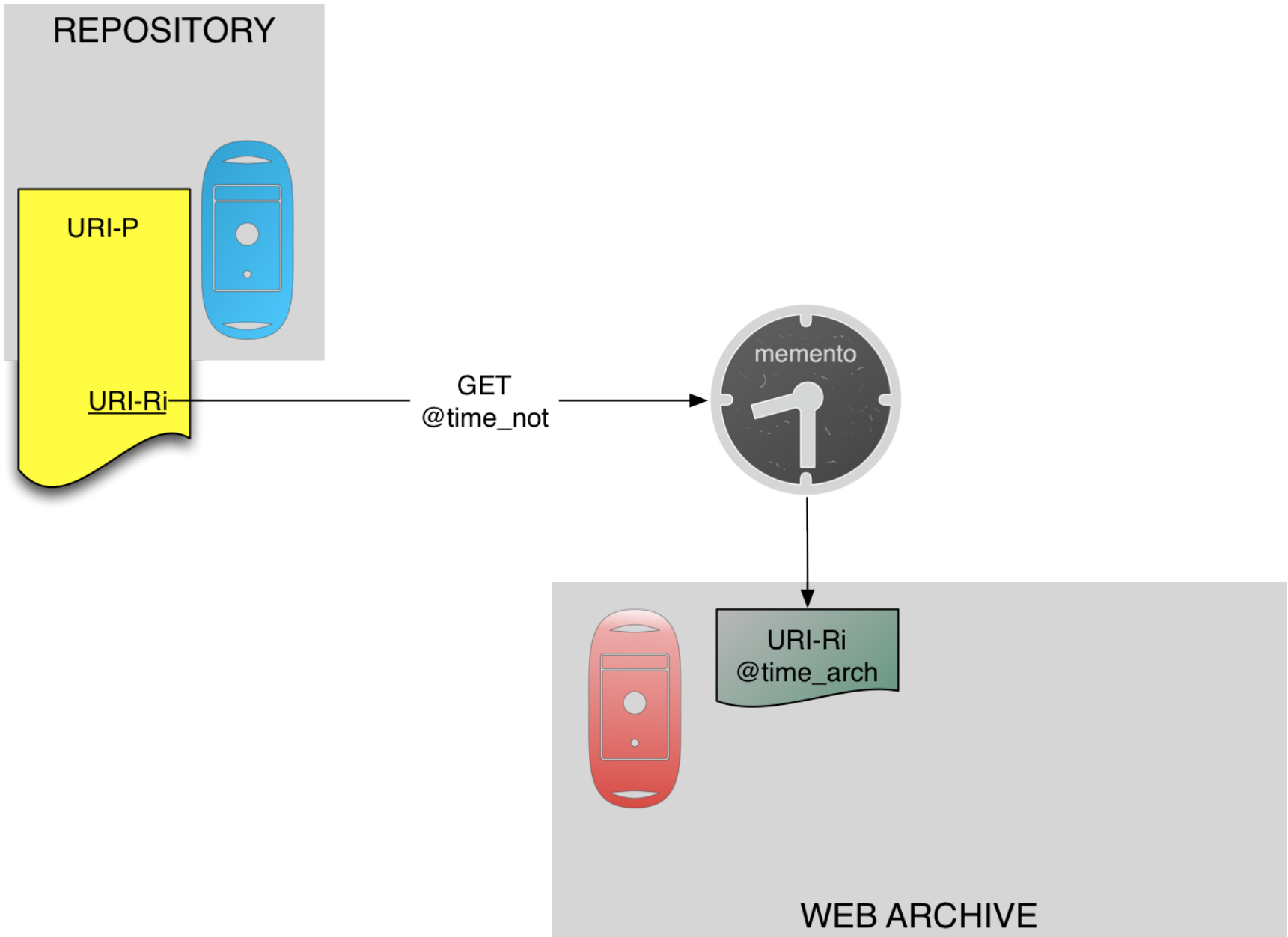
```
<?xml version="1.0" encoding="UTF-8"?>
<urlset
  xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
  xmlns:rs="http://www.openarchives.org/rs/terms/">
  <rs:ln rel="up" href="http://hiberactive.lanl.gov/changelist.xml"/>
  <rs:md capability="changelist" from="2014-05-01T00:00:00Z"/>
  <url>
    <loc>http://archive.today/FL2as</loc>
    <lastmod>2014-06-08T03:31:19Z</lastmod>
    <rs:md change="created" />
    <rs:ln rel="original" href="http://obofoundry.org/ro/" />
    <rs:ln rel="collection" href="http://core.kmi.open.ac.uk/download/pdf/288127.pdf"/>
  </url>
</urlset
```



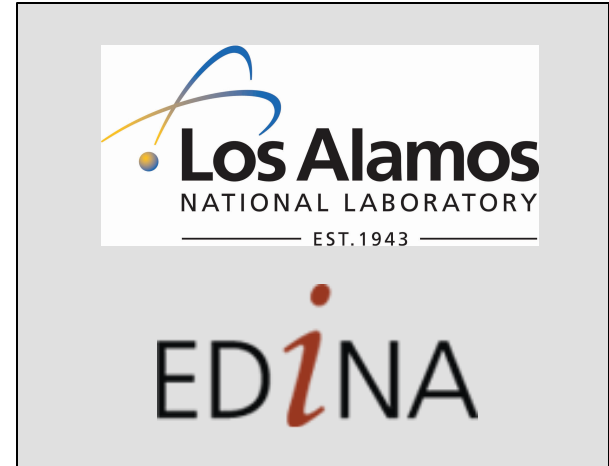
# ResourceSync Change List

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset
  xmlns="http://www.sitemaps.org/schemas/sitemap/0.9"
  xmlns:rs="http://www.openarchives.org/rs/terms/">
  <rs:ln rel="up" href="http://hiberactive.lanl.gov/changelist.xml"/>
  <rs:md capability="changelist" from="2014-05-01T00:00:00Z"/>
  <url>
    <loc>http://archive.today/FL2as</loc>
    <lastmod>2014-06-08T03:31:19Z</lastmod>
    <rs:md change="created" />
    <rs:ln rel="original" href="http://obofoundry.org/ro/" />
    <rs:ln rel="collection" href="http://core.kmi.open.ac.uk/download/pdf/288127.pdf"/>
  </url>
</urlset>
```





# Demo



[https://www.dropbox.com/s/rysi5ekmozbiq/4min\\_video.mp4](https://www.dropbox.com/s/rysi5ekmozbiq/4min_video.mp4)



**HiberActive** - Martin Klein et al.  
Open Repositories, Helsinki, Finland, June 10th 2014



**Los Alamos**  
NATIONAL LABORATORY  
THE UNIVERSITY of EDINBURGH





# HiberActive

## Pro-Active Archiving of Web References from Scholarly Articles

**Martin Klein**, Harihar Shankar, Herbert Van de Sompel  
Los Alamos National Laboratory  
**@martinklein**

Richard Wincewicz  
Edina, University of Edinburgh

<http://hiberlink.org>      #hiberlink  
<http://mementoweb.org>      #memento  
<http://openarchives.org/rs/>      #resourcesync

