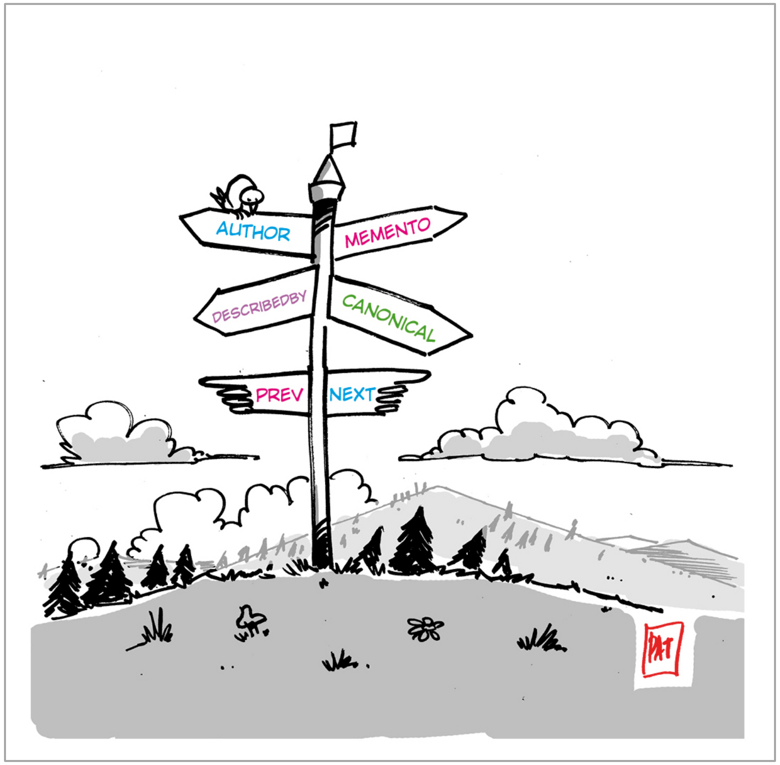
Establishing New Levels of Interoperability for Web-Based Scholarship



Cartoon by:
Patrick Hochstenbach

Herbert Van de Sompel
Los Alamos National Laboratory
@hvdsomp

Acknowledgments: Michael L. Nelson, David Rosenthal, Geoff Bilder, Simeon Warner, Harihar Shankar, Shawn Jones

D-Lib® Magazine
The Magazine of Digital Library Research
HOME | ABOUT D-LIB | CURRENT ISSUE | ARCHIVE | INDEXES | CALENDAR | AUTHOR GUIDELINES | SUBSCRIBE | CONTACT D-LIB

D-Lib Magazine

November/December 2015
Volume 21, Number 11/12
Table of Contents

Reminiscing About 15 Years of Interoperability Efforts

Herbert Van de Sompel
Los Alamos National Laboratory
herbertv@lanl.gov

Michael L. Nelson
Old Dominion University
mln@cs.odu.edu

DOI: 10.1045/november2015-vandesompel

https://dx.doi.org/10.1045/november2015-vandesompel

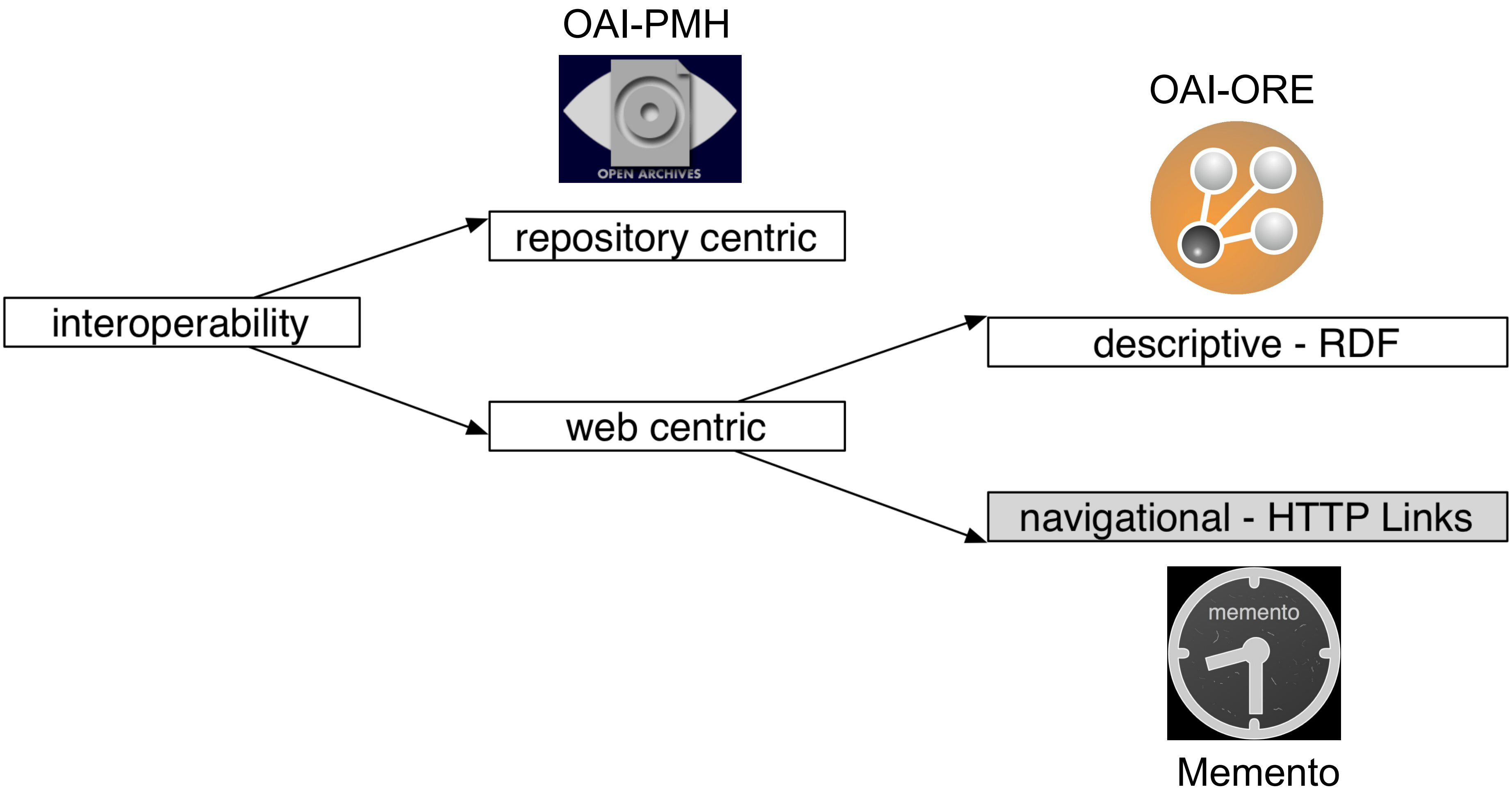
Research Communication & Research Process on the Web

- A highly distributed activity

- Turning this distributed activity from a gathering of silo-ed nodes into an ecology of collaborating nodes, requires establishing interoperability
    - In the web context, this seems like a rather unique challenge: Most web enterprises do not want interoperability they want dominance, monopoly

- To a large extent, interoperability <u>across</u> this distributed activity remains restricted to persistent identification of communicated objects and contributors
    - Which results in added-value – services can be created

Evolution of Our Thinking about Interoperability



OAI-PMH

OAI-ORE

interoperability

repository centric

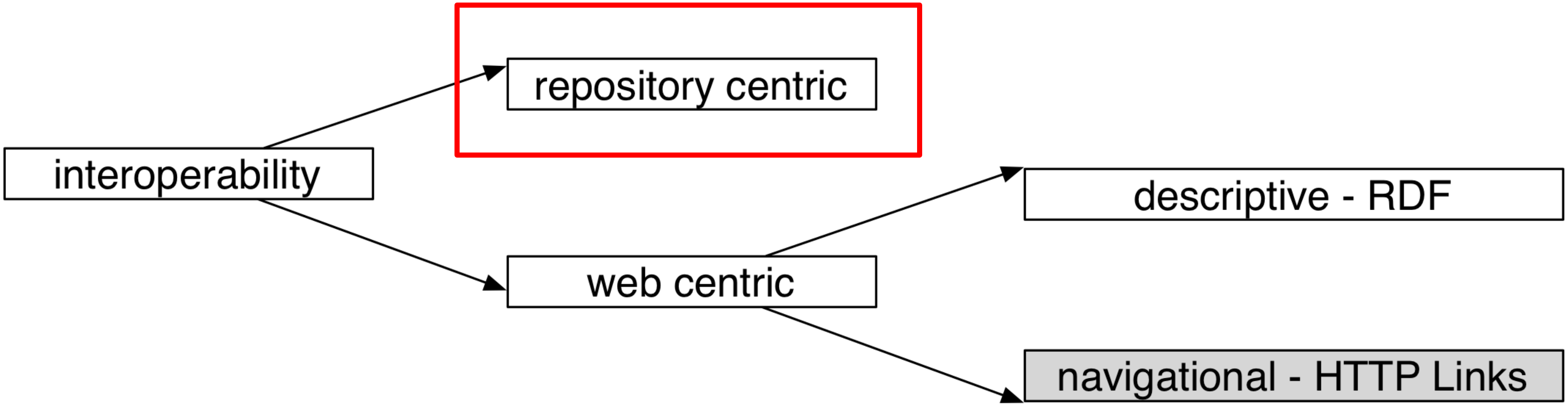
web centric

descriptive - RDF

navigational - HTTP Links

Memento

Evolution of Our Thinking about Interoperability



1999

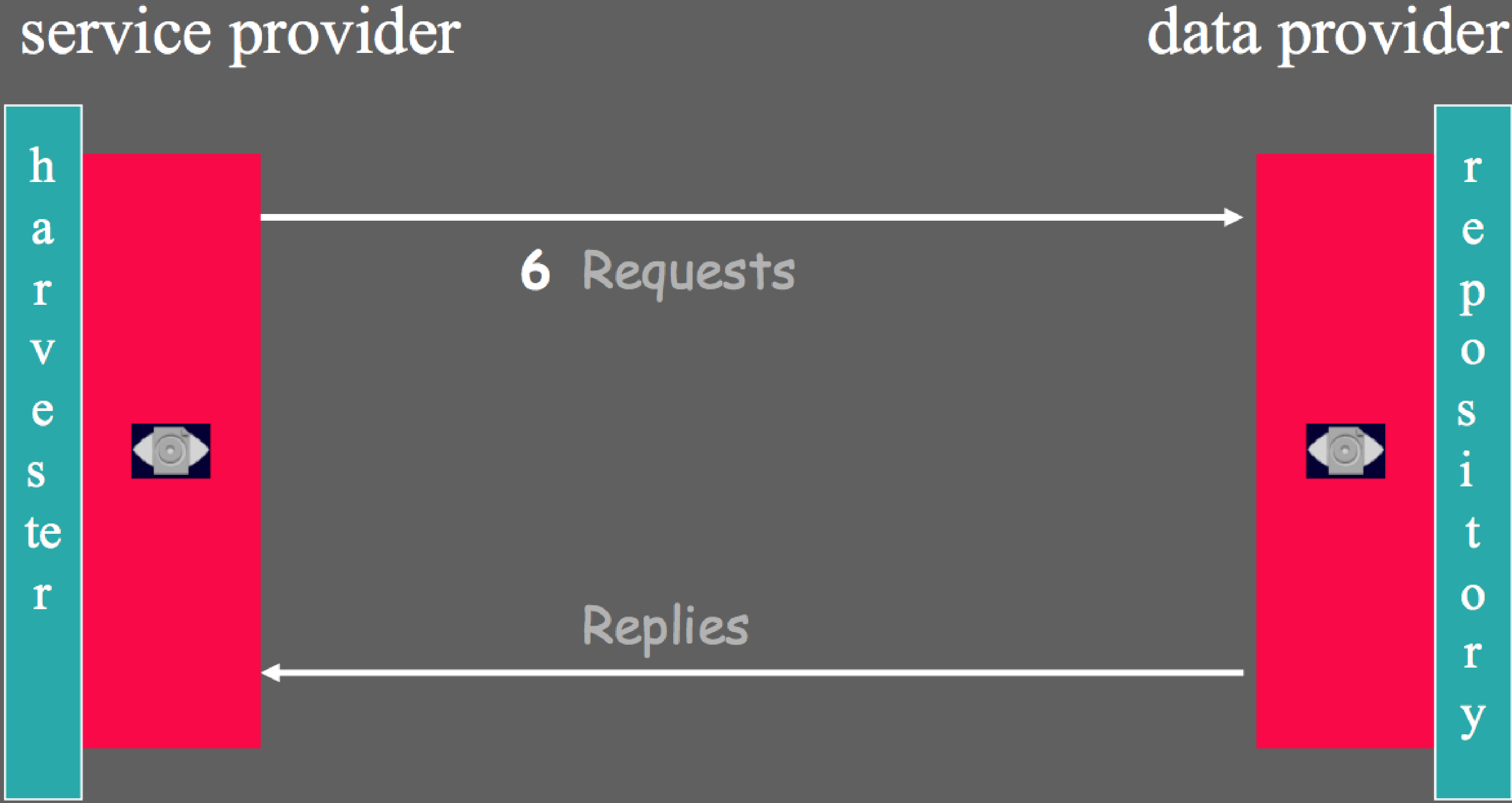
- OAI was a heroic effort to fundamentally transform scholarly communication
  - By promoting communication via preprints, non-peer-reviewed papers

- The OAI took a technical approach to achieve the goal
  - Make preprints easier to discover, access

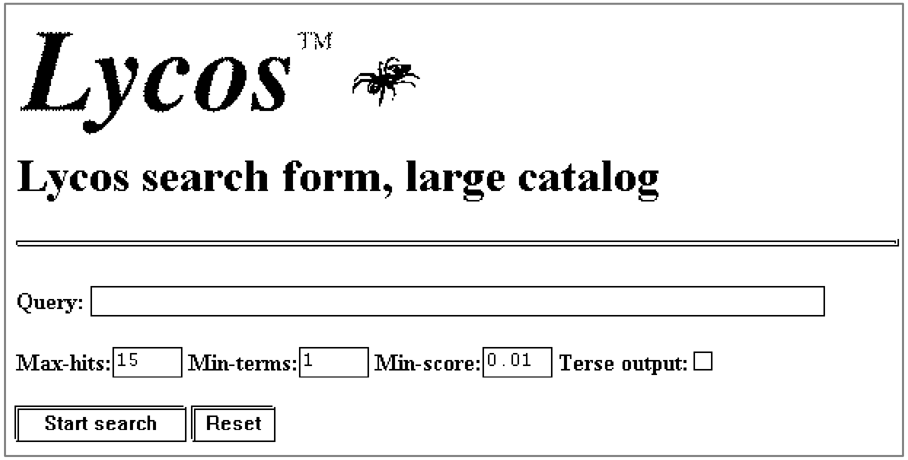
http://www.openarchives.org/OAI/openarchivesprotocol.html

The OAMH protocol is a low-barrier interoperability specification for the recurrent exchange of metadata between systems

the Metadata Harvesting protocol



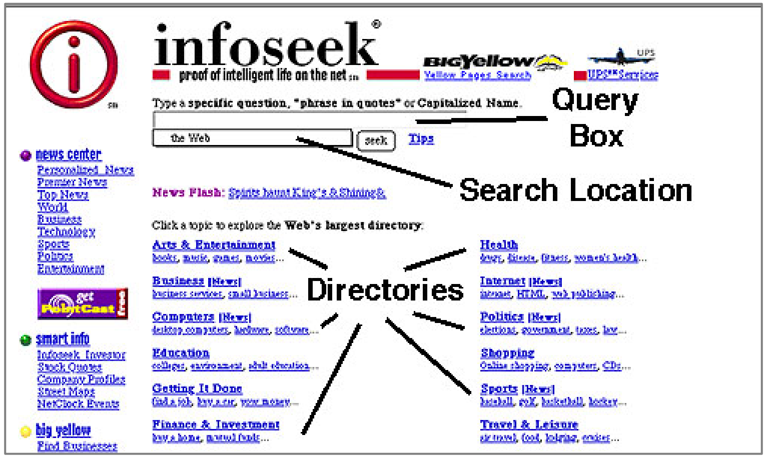
Those Were the Days


Lycos™

Lycos search form, large catalog

Query:

Max-hits: 15 Min-terms: 1 Min-score: 0.01 Terse output: □


infoseek®
proof of intelligent life on the net℠

Type a specific question, "phrase in quotes" or Capitalized Name.

the Web

Query Box

Search Location

News Flash: Spirits haunt King's & Shining&

Click a topic to explore the Web's largest directory:

Directories

- news center
  - Personalized News
  - Premier News
  - Top News
  - World
  - Business
  - Technology
  - Sports
  - Politics
  - Entertainment
- smart info
  - Infoseek Investor
  - Stock Quotes
  - Company Profiles
  - Street Maps
  - NetClock Events
- big yellow
  - Find Businesses

Arts & Entertainment
books, music, games, movies...

Business [News]
business services, small business...

Computers [News]
desktop computers, hardware, software...

Education
colleges, environment, adult education...

Getting It Done
find a job, buy a car, save money...

Finance & Investment
buy a home, mutual funds...

Health
drugs, fitness, illnesses, women's health...

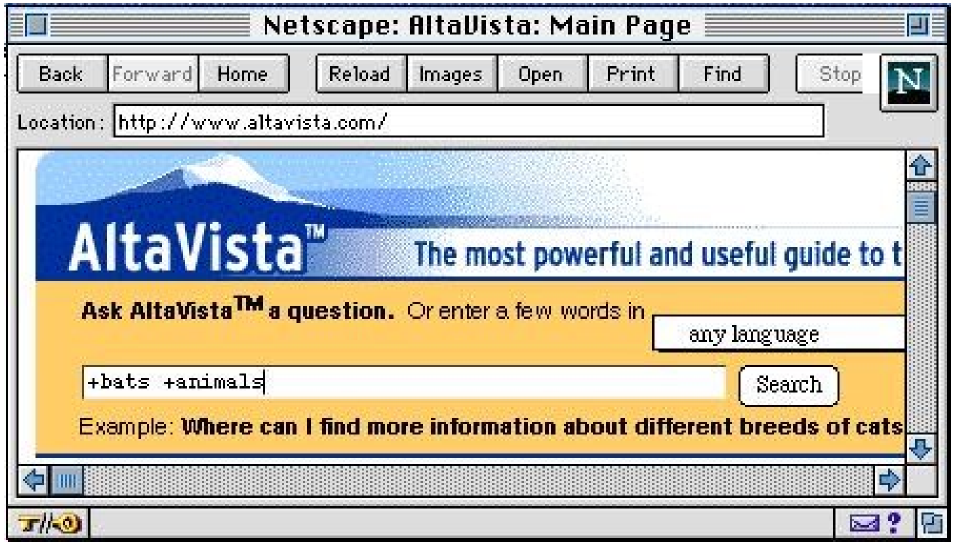
Internet [News]
internet, HTML, web publishing...

Politics [News]
elections, government, taxes, law...

Shopping
Online shopping, computers, CDs...

Sports [News]
baseball, golf, basketball, hockey...

Travel & Leisure
air travel, food, lodging, cruises...


Netscape: AltaVista: Main Page

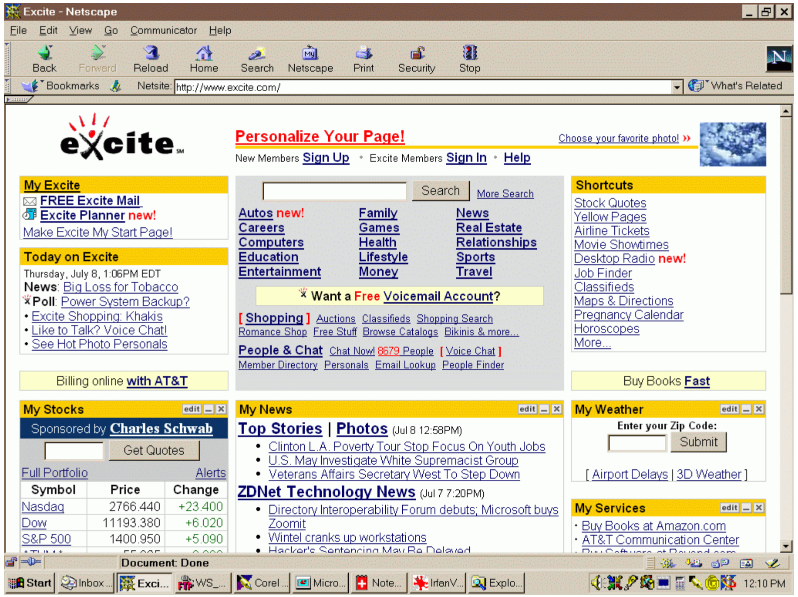
Location: http://www.altavista.com/

AltaVista™ The most powerful and useful guide to t

Ask AltaVista™ a question. Or enter a few words in any language

+bats +animals

Example: Where can I find more information about different breeds of cats


Excite - Netscape

Netsite: http://www.excite.com/

excite℠

Personalize Your Page!
New Members Sign Up • Excite Members Sign In • Help

Choose your favorite photo! »

My Excite
- FREE Excite Mail
- Excite Planner new!
- Make Excite My Start Page!

Today on Excite
Thursday, July 8, 1:06PM EDT
News: Big Loss for Tobacco
Poll: Power System Backup?
- Excite Shopping: Khakis
- Like to Talk? Voice Chat!
- See Hot Photo Personals

Billing online with AT&T

Autos new!
Careers
Computers
Education
Entertainment
Family
Games
Health
Lifestyle
Money
News
Real Estate
Relationships
Sports
Travel

Want a Free Voicemail Account?

[Shopping] Auctions Classifieds Shopping Search
Romance Shop Free Stuff Browse Catalogs Bikinis & more...

People & Chat Chat Now! 8679 People [Voice Chat]
Member Directory Personals Email Lookup People Finder

Shortcuts
- Stock Quotes
- Yellow Pages
- Airline Tickets
- Movie Showtimes
- Desktop Radio new!
- Job Finder
- Classifieds
- Maps & Directions
- Pregnancy Calendar
- Horoscopes
- More...

Buy Books Fast

My Stocks
Sponsored by Charles Schwab

Full Portfolio Alerts

Symbol	Price	Change
Nasdaq	2766.440	+23.400
Dow	11193.380	+6.020
S&P 500	1400.950	+5.090

My News
Top Stories | Photos (Jul 8 12:58PM)
- Clinton L.A. Poverty Tour Stop Focus On Youth Jobs
- U.S. May Investigate White Supremacist Group
- Veterans Affairs Secretary West To Step Down

ZDNet Technology News (Jul 7 7:20PM)
- Directory Interoperability Forum debuts; Microsoft buys Zoomit
- Wintel cranks up workstations
- Hacker's Sentencing May Be Delayed

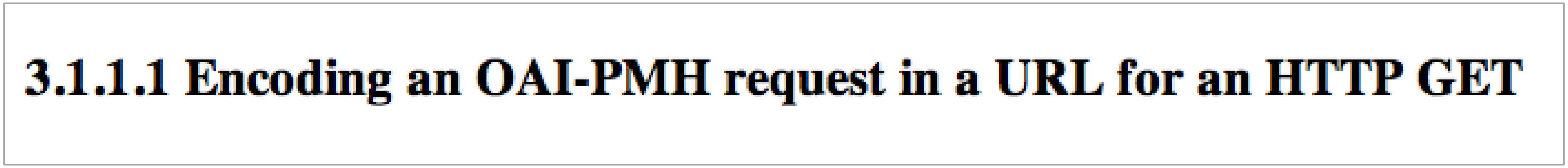
My Weather
Enter your Zip Code:

[Airport Delays | 3D Weather]

My Services
- Buy Books at Amazon.com
- AT&T Communication Center

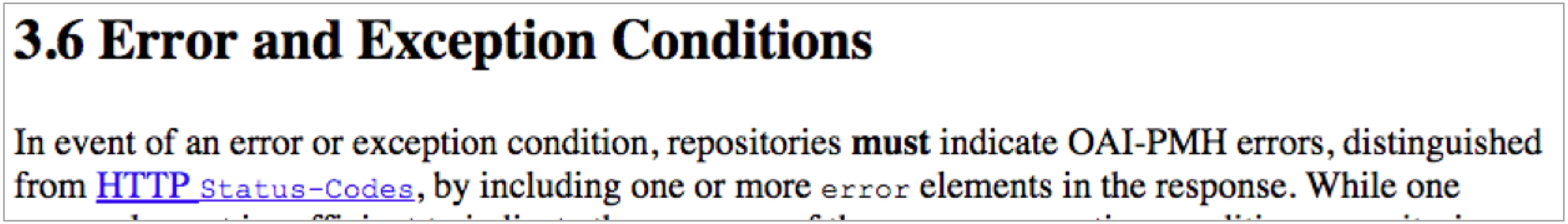
Document: Done

http://



3.1.1.1 Encoding an OAI-PMH request in a URL for an HTTP GET

Don't trust HTTP



3.6 Error and Exception Conditions

In event of an error or exception condition, repositories **must** indicate OAI-PMH errors, distinguished from HTTP `Status-Codes`, by including one or more `error` elements in the response. While one

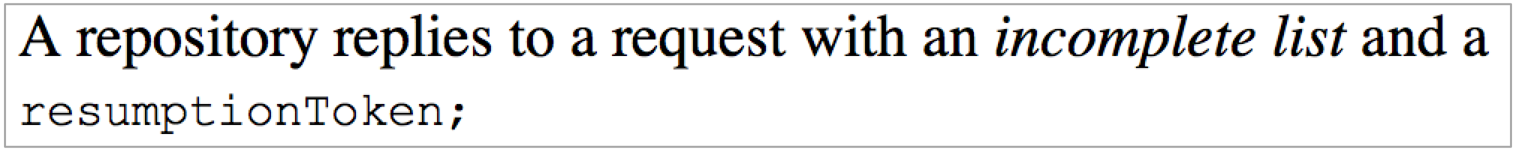
http://

OPEN ARCHIVES

```
http://an.oa.org/OAI-script?
verb=GetRecord&identifier=oai:arXiv.org:hep-
th/9901001&metadataPrefix=oai_dc
```

HTTP GET with GetRecord verb

http://



A repository replies to a request with an *incomplete list* and a `resumptionToken;`

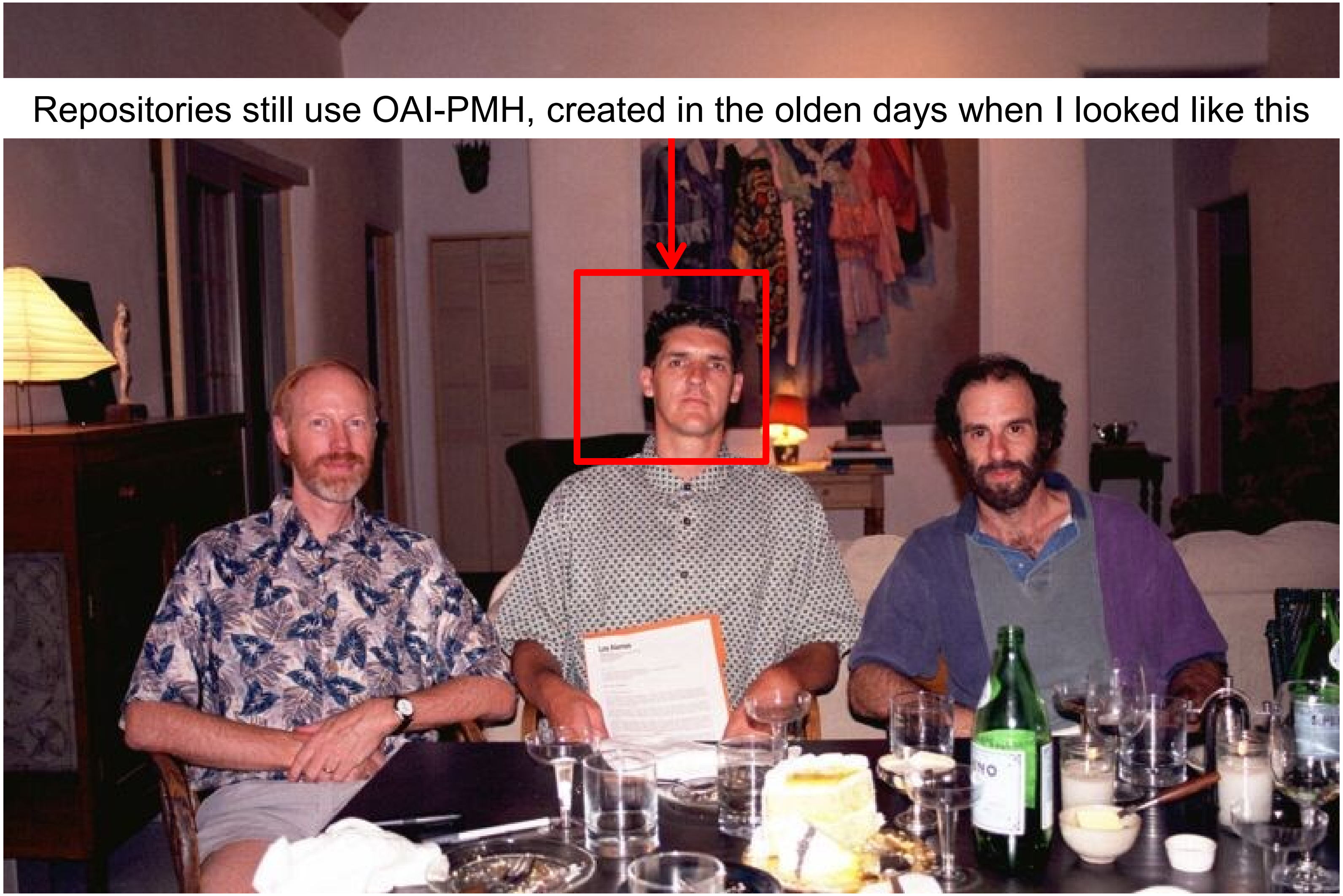
An HTTP link

Repository-Centric Interoperability Paradigm

Address interoperability challenges from the perspective of a node, e.g. an IR, a publisher, a web-based authoring portal, a software repository, …

- **The node at the center of the universe**
- Define a machine interface for your node, expect others to use it
- Piggybacking on the web without truly embracing its core technologies
- The node resembles a brick & mortar library that can be visited subject to well-intended yet idiosyncratic policies – the interface

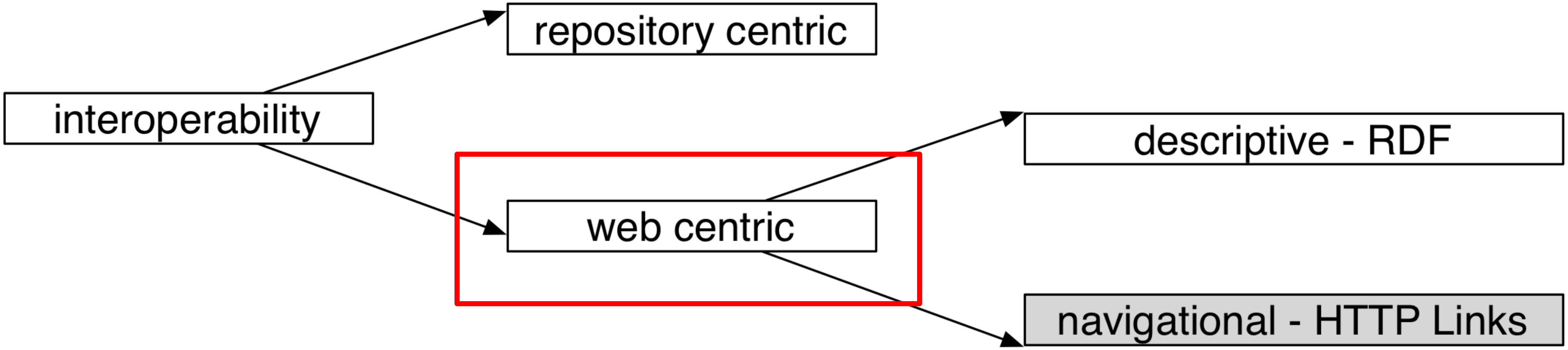
Launching the OAI - Luce, Van de Sompel, Ginsparg (1999)



Repositories still use OAI-PMH, created in the olden days when I looked like this

https://www.openarchives.org/meetings/SantaFe1999/rhp.htm

Evolution of Our Thinking about Interoperability



Web-Centric, Resource-Centric Interoperability Paradigm

Address interoperability challenges from the perspective of the web

- **The resource at the center of the universe**
  - The notion of a node, a repository, not even of a web server exists in the architecture of the web

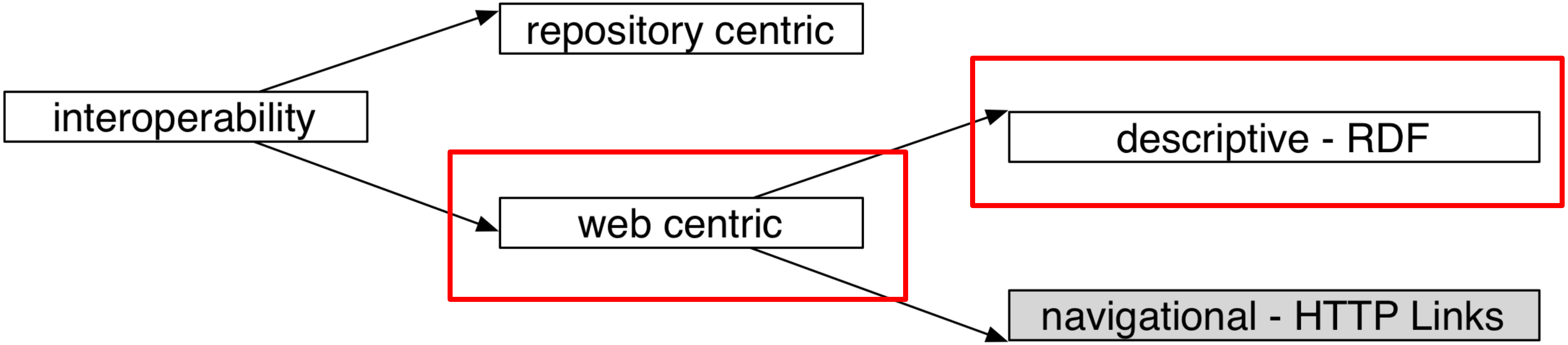
- The tools of the interoperability trade are the primitives of the web

Tools of the Web-Centric Interoperability Trade

- Resource
- URI
- HTTP as the API: HEAD/GET, POST, PUT, DELETE
- Representation
- Media Type
- Link
- Content Negotiation

W3C Architecture of the World Wide Web

Evolution of Our Thinking about Interoperability

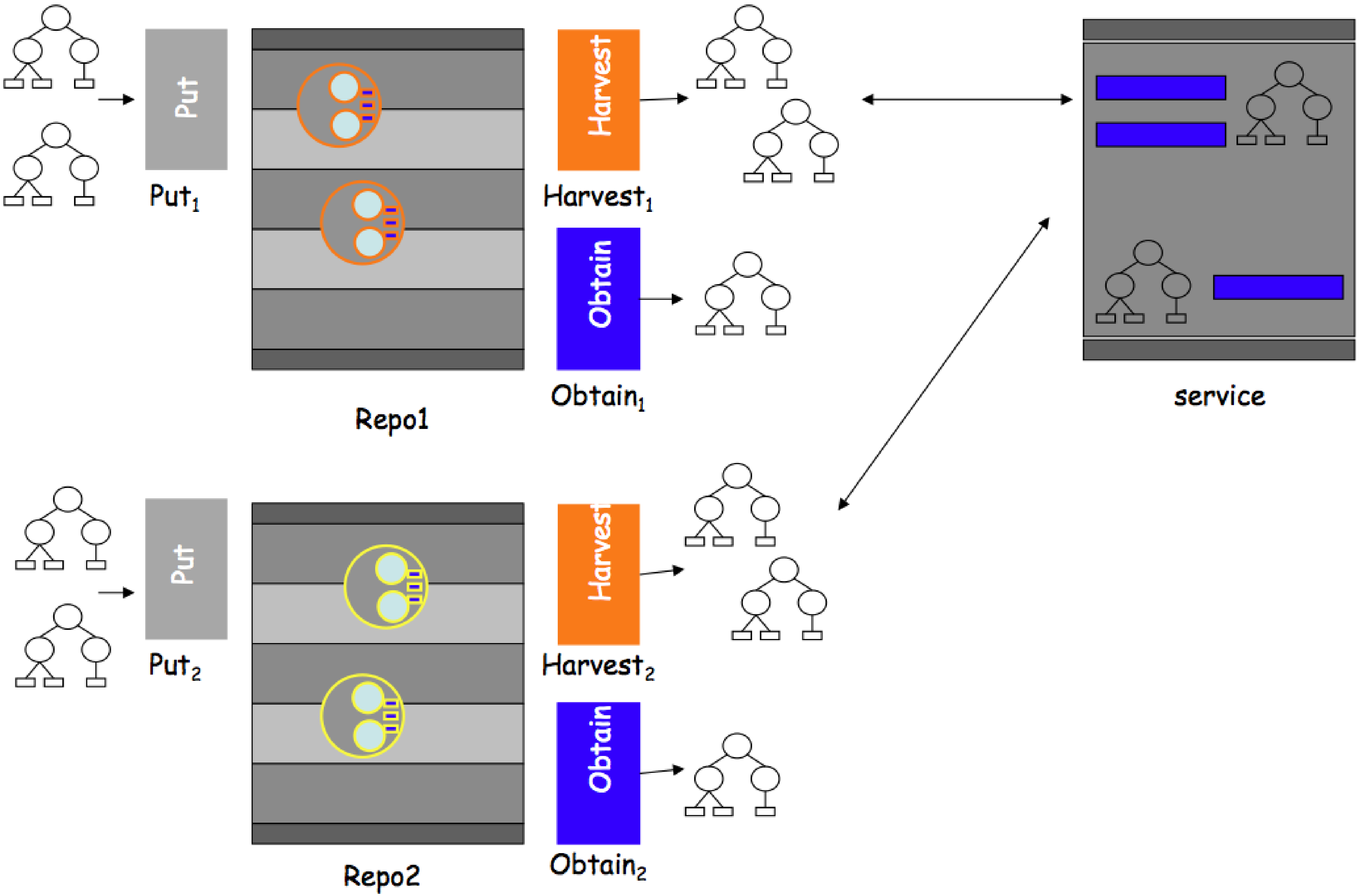


- OAI-ORE observation: Scholarly assets are rapidly becoming *compound*, consisting of multiple resources with various:
  - Relationships
  - Interdependencies

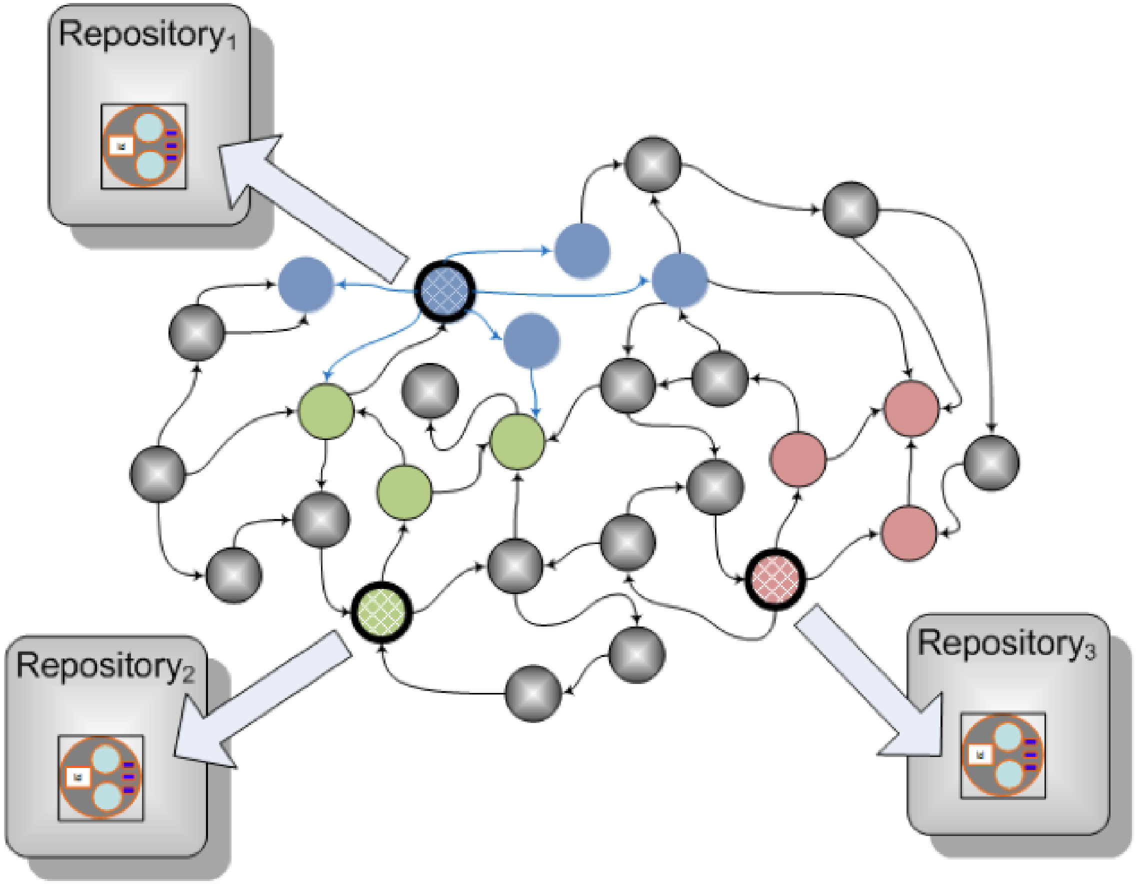
- How to convey this compound-ness in an interoperable manner so that applications can access, consume such assets?

2006

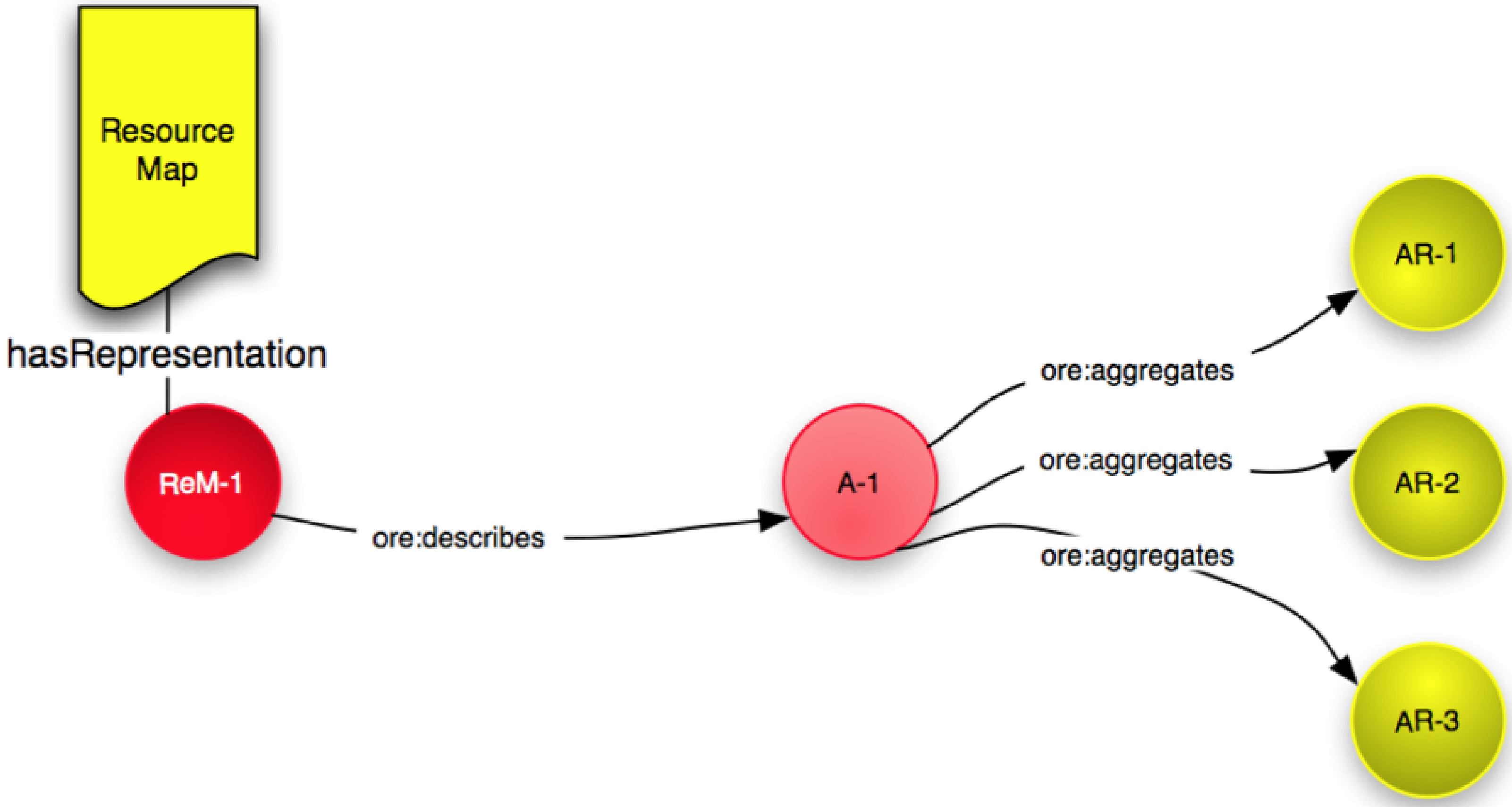
http://www.openarchives.org/ore/1.0/toc


Put
Put₁
Harvest
Harvest₁
Obtain
Obtain₁
Repo1
service
Put
Put₂
Harvest
Harvest₂
Obtain
Obtain₂
Repo2

ORE Will Allow Web Crawlers to Unambiguously Recover CDO Structure from the Web Graph



Express the `ore:describes` relationship



Tools of the Web-Centric Interoperability Trade – RDF Stack

- Resource
- URI
- HTTP as the API
- Representation
- Media Type
- Link
- Content Negotiation, e.g. for preferred Media Type

W3C Architecture of the World Wide Web

- Typed Link
- Controlled Vocabularies for Typed Links

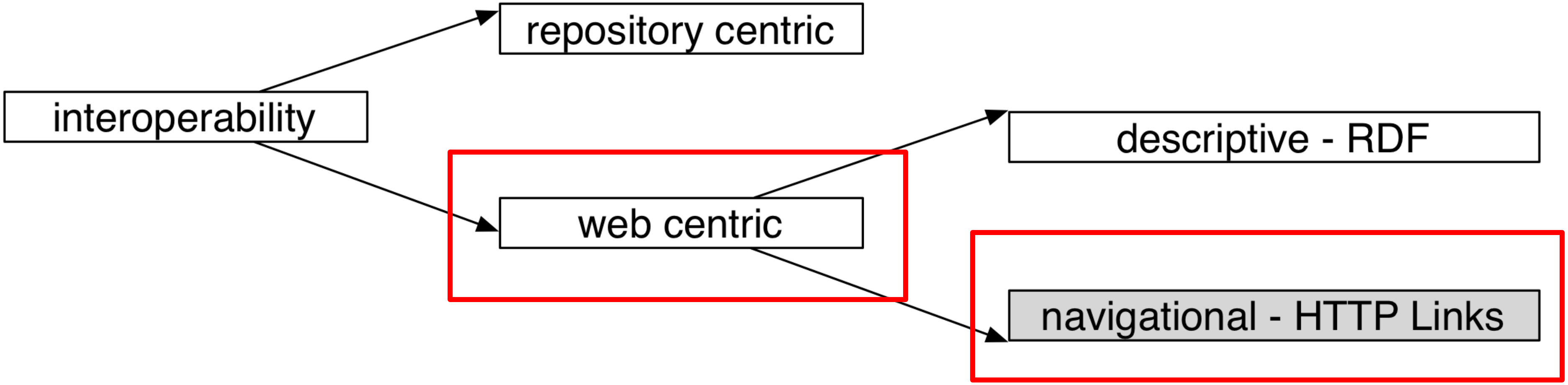
RDF, RDFS, OWL

Interoperability via RDF, RDFS, OWL Stack

Used by various interoperability efforts, e.g. OAI-ORE, Open Annotation, W3C PROV, Research Objects, …

- Provides extensive expressiveness for description
- Typically based on publishing documents that adhere to a certain “profile” and reveal relations, properties, …
- Non-Trivial barrier to entry as illustrated by slow adoption, likely related to unfamiliar technology stack

Evolution of Our Thinking about Interoperability



- Memento is about the Web and time:
  - Resources evolve over time
  - Only the current resource version is available from a resource’s URI
  - How to seamlessly access prior versions, if they exist, using the resource’s URI and a version datetime

- Memento looks at this problem for the Web, in general:
  - Time-Based access to resource versions across web archives, resource versioning systems

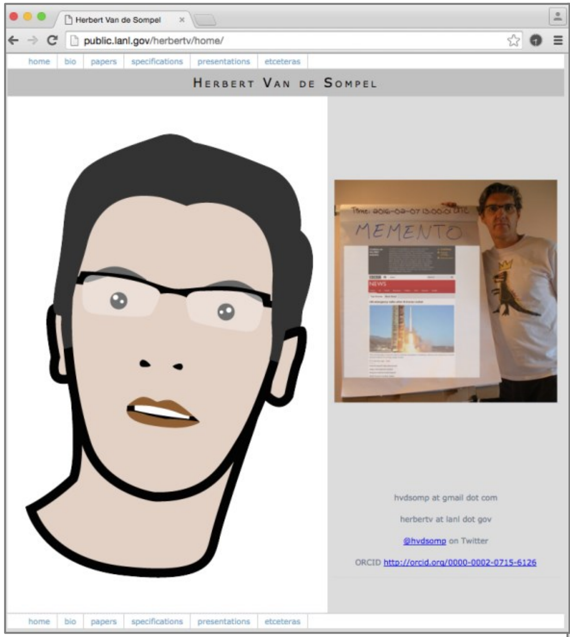
2009

RFC7089 - http://mementoweb.org/guide/rfc/
Memento and Resource Versioning - http://mementoweb.org/guide/howto/

Today

Select Date Mar 20 2007

Apr 03 2007

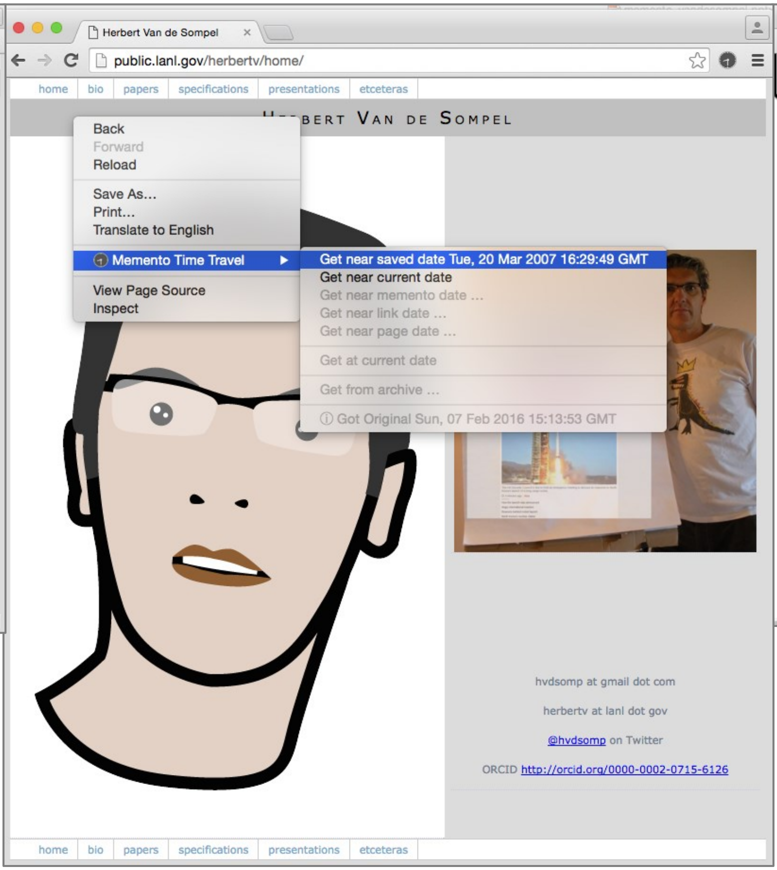

Herbert Van de Sompel
public.lanl.gov/herbertv/home/
home | bio | papers | specifications | presentations | etceteras
Herbert Van de Sompel
home | bio | papers | specifications | presentations | etceteras

hvdsomp at gmail dot com

herbertv at lanl dot gov

@hvdsomp on Twitter

ORCID http://orcid.org/0000-0002-0715-6126

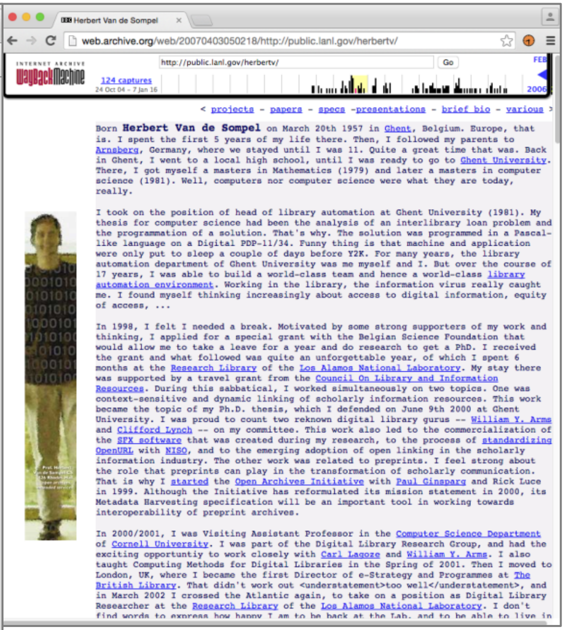

Herbert Van de Sompel
public.lanl.gov/herbertv/home/
home | bio | papers | specifications | presentations | etceteras
Herbert Van de Sompel
Back
Forward
Reload
Save As...
Print...
Translate to English
Memento Time Travel
View Page Source
Inspect
Get near saved date Tue, 20 Mar 2007 16:29:49 GMT
Get near current date
Get near memento date ...
Get near link date ...
Get near page date ...
Get at current date
Get from archive ...
Got Original Sun, 07 Feb 2016 15:13:53 GMT
home | bio | papers | specifications | presentations | etceteras

hvdsomp at gmail dot com

herbertv at lanl dot gov

@hvdsomp on Twitter

ORCID http://orcid.org/0000-0002-0715-6126


Herbert Van de Sompel
web.archive.org/web/20070403050218/http://public.lanl.gov/herbertv/
INTERNET ARCHIVE
WaybackMachine
http://public.lanl.gov/herbertv/
Go
124 captures
24 Oct 04 - 7 Jan 16
FEB
2006

< projects - papers - specs -presentations - brief bio - various >

Born **Herbert Van de Sompel** on March 20th 1957 in Ghent, Belgium. Europe, that is. I spent the first 5 years of my life there. Then, I followed my parents to Arnsberg, Germany, where we stayed until I was 11. Quite a great time that was. Back in Ghent, I went to a local high school, until I was ready to go to Ghent University. There, I got myself a masters in Mathematics (1979) and later a masters in computer science (1981). Well, computers nor computer science were what they are today, really.

I took on the position of head of library automation at Ghent University (1981). My thesis for computer science had been the analysis of an interlibrary loan problem and the programmation of a solution. That's why. The solution was programmed in a Pascal-like language on a Digital PDP-11/34. Funny thing is that machine and application were only put to sleep a couple of days before Y2K. For many years, the library automation department of Ghent University was me myself and I. But over the course of 17 years, I was able to build a world-class team and hence a world-class library automation environment. Working in the library, the information virus really caught me. I found myself thinking increasingly about access to digital information, equity of access, ...

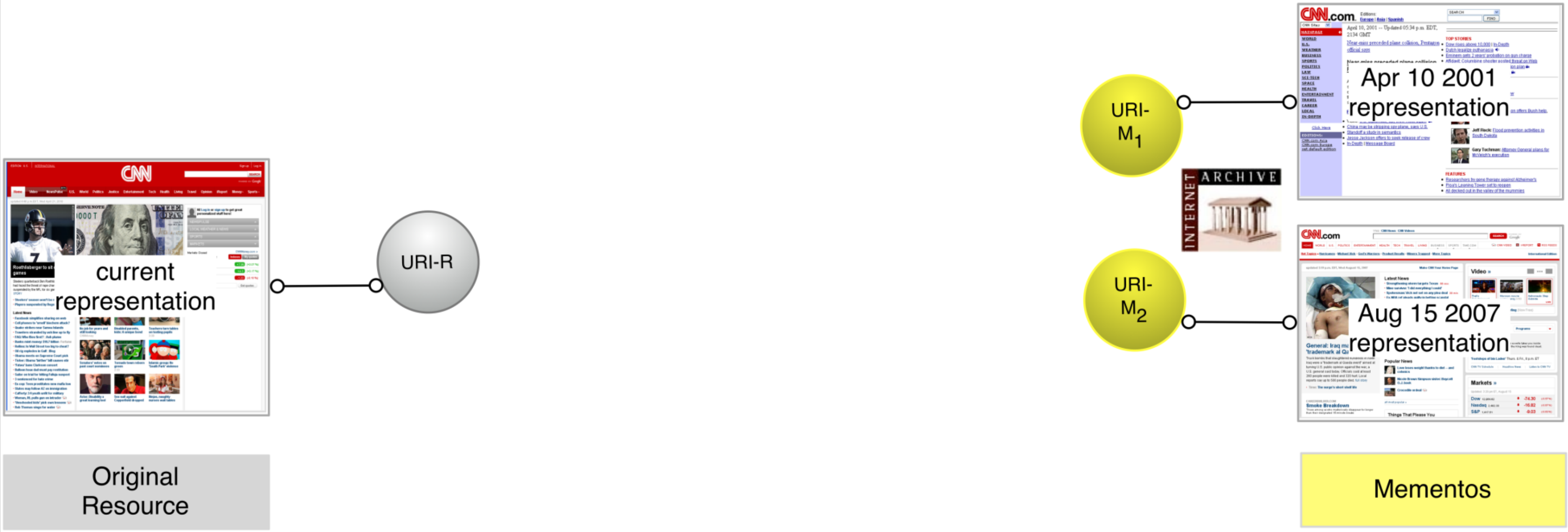
In 1998, I felt I needed a break. Motivated by some strong supporters of my work and thinking, I applied for a special grant with the Belgian Science Foundation that would allow me to take a leave for a year and do research to get a PhD. I received the grant and what followed was quite an unforgettable year, of which I spent 6 months at the Research Library of the Los Alamos National Laboratory. My stay there was supported by a travel grant from the Council On Library and Information Resources. During this sabbatical, I worked simultaneously on two topics. One was context-sensitive and dynamic linking of scholarly information resources. This work became the topic of my Ph.D. thesis, which I defended on June 9th 2000 at Ghent University. I was proud to count two reknown digital library gurus -- William Y. Arms and Clifford Lynch -- on my committee. This work also led to the commercialization of the SFX software that was created during my research, to the process of standardizing OpenURL with NISO, and to the emerging adoption of open linking in the scholarly information industry. The other work was related to preprints. I feel strong about the role that preprints can play in the transformation of scholarly communication. That is why I started the Open Archives Initiative with Paul Ginsparg and Rick Luce in 1999. Although the Initiative has reformulated its mission statement in 2000, its Metadata Harvesting specification will be an important tool in working towards interoperability of preprint archives.

In 2000/2001, I was Visiting Assistant Professor in the Computer Science Department of Cornell University. I was part of the Digital Library Research Group, and had the exciting opportuntiy to work closely with Carl Lagoze and William Y. Arms. I also taught Computing Methods for Digital Libraries in the Spring of 2001. Then I moved to London, UK, where I became the first Director of e-Strategy and Programmes at The British Library. That didn't work out <understatement>too well</understatement>, and in March 2002 I crossed the Atlantic again, to take on a position as Digital Library Researcher at the Research Library of the Los Alamos National Laboratory. I don't find words to express how happy I am to be back at the Lab, and to be able to live in

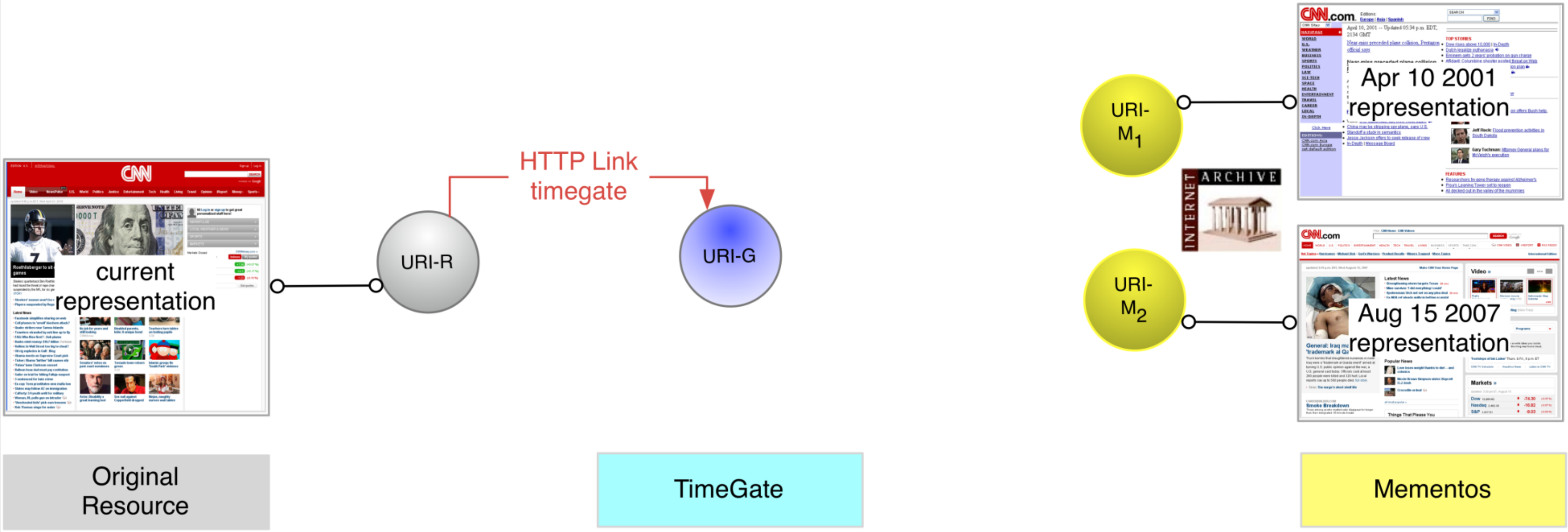
From
Internet Archive

Memento for Chrome at http://bit.ly/memento-for-chrome

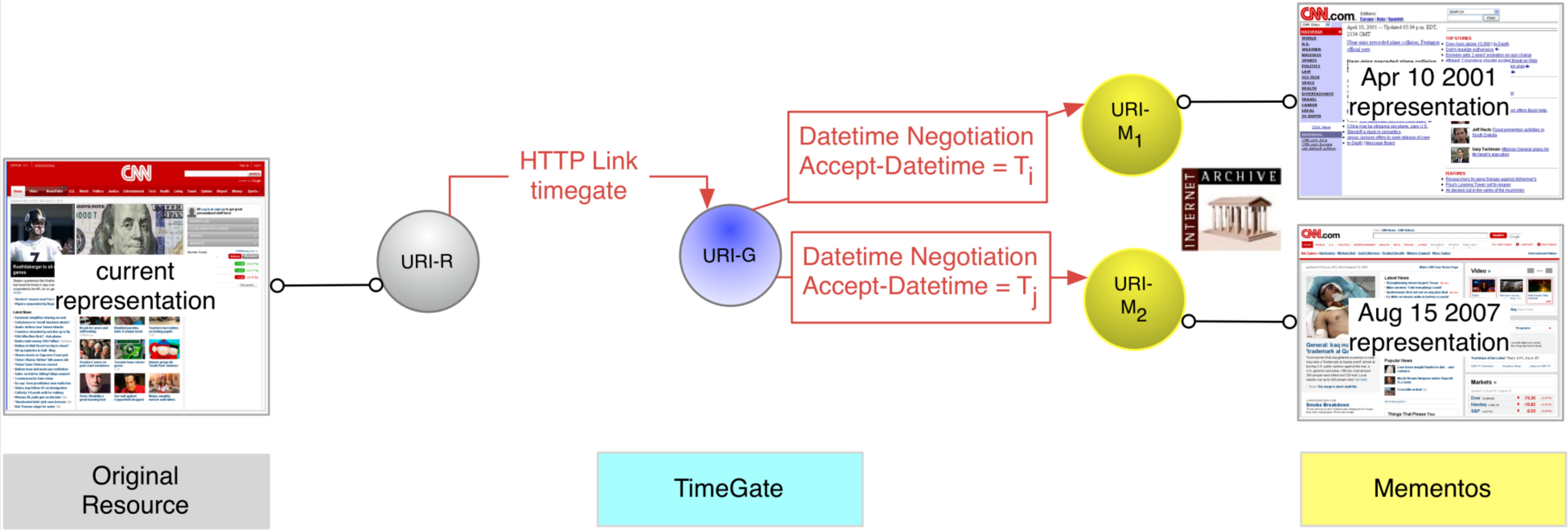
Original Resource and Mementos



Bridge from Present to Past



Bridge from Present to Past



current representation

URI-R

HTTP Link timegate

URI-G

Datetime Negotiation Accept-Datetime = T_i

Datetime Negotiation Accept-Datetime = T_j

URI-M_1

URI-M_2

INTERNET ARCHIVE

Apr 10 2001 representation

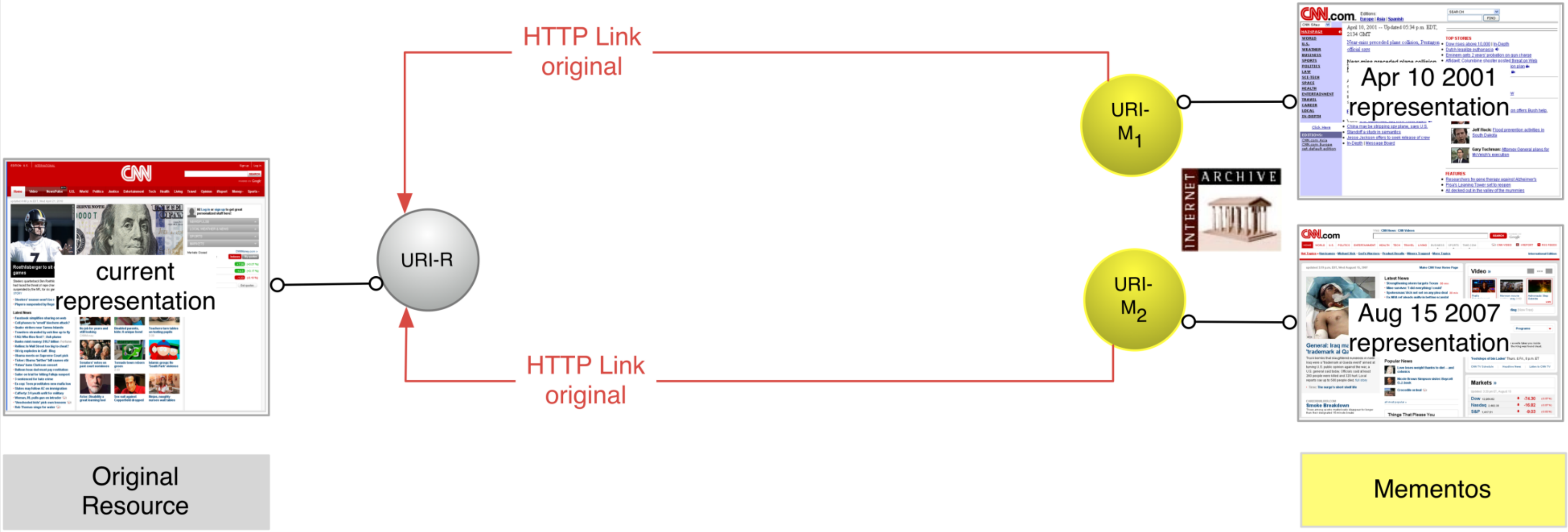
Aug 15 2007 representation

Original Resource

TimeGate

Mementos

Bridge from Past to Present



HTTP Link original

HTTP Link original

URI-R

URI-M_1

URI-M_2

current representation

Apr 10 2001 representation

Aug 15 2007 representation

INTERNET ARCHIVE

Original Resource

Mementos

Tools of the Web-Centric Interoperability Trade – HTTP Stack

- Resource
- URI
- HTTP as the API
- Representation
- Media Types
- Link
- Content Negotiation, e.g. for Media Type, Time

W3C Architecture of the World Wide Web

- Typed Link
- Controlled Vocabularies for Typed Links

HTTP Links, IANA link relation registry, community link relation types

HATEOAS – Hypermedia As The Engine Of Application State

http://en.wikipedia.org/wiki/HATEOAS

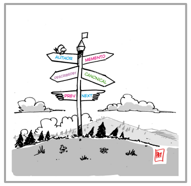
Interoperability via HTTP Links, IANA Link Relation Types

Used by Memento, ResourceSync, Signposting the Scholarly Web:

- Provides coarse expressiveness for navigation via IANA registered relation types (expressed as reserved terms)
    - Finer grained expressiveness via community-defined relation types (expressed as HTTP URIs)
- Typically based on publishing typed links that support a client to navigate among resources in an informed manner
- Low implementation barrier because of familiar technology stack

Establishing New Levels of Interoperability: Examples

ResourceSync



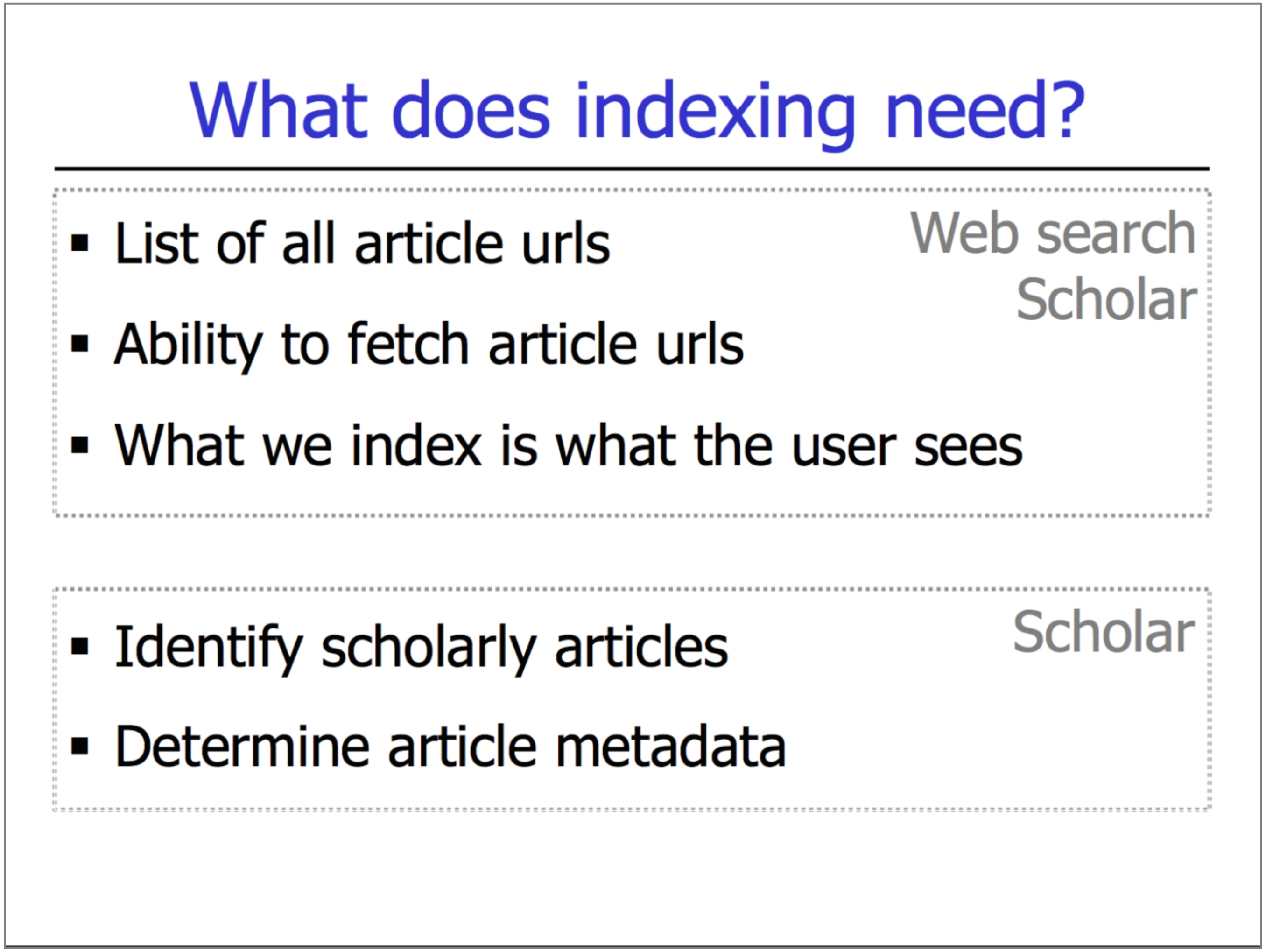
Signposting the Scholarly Web

Robust Links

ResourceSync

Herbert Van de Sompel
COAR Annual Meeting, Vienna, Austria, 12/04/2016

Anurag Acharya Told Us Why We Need ResourceSync



What does indexing need?

Web search
Scholar

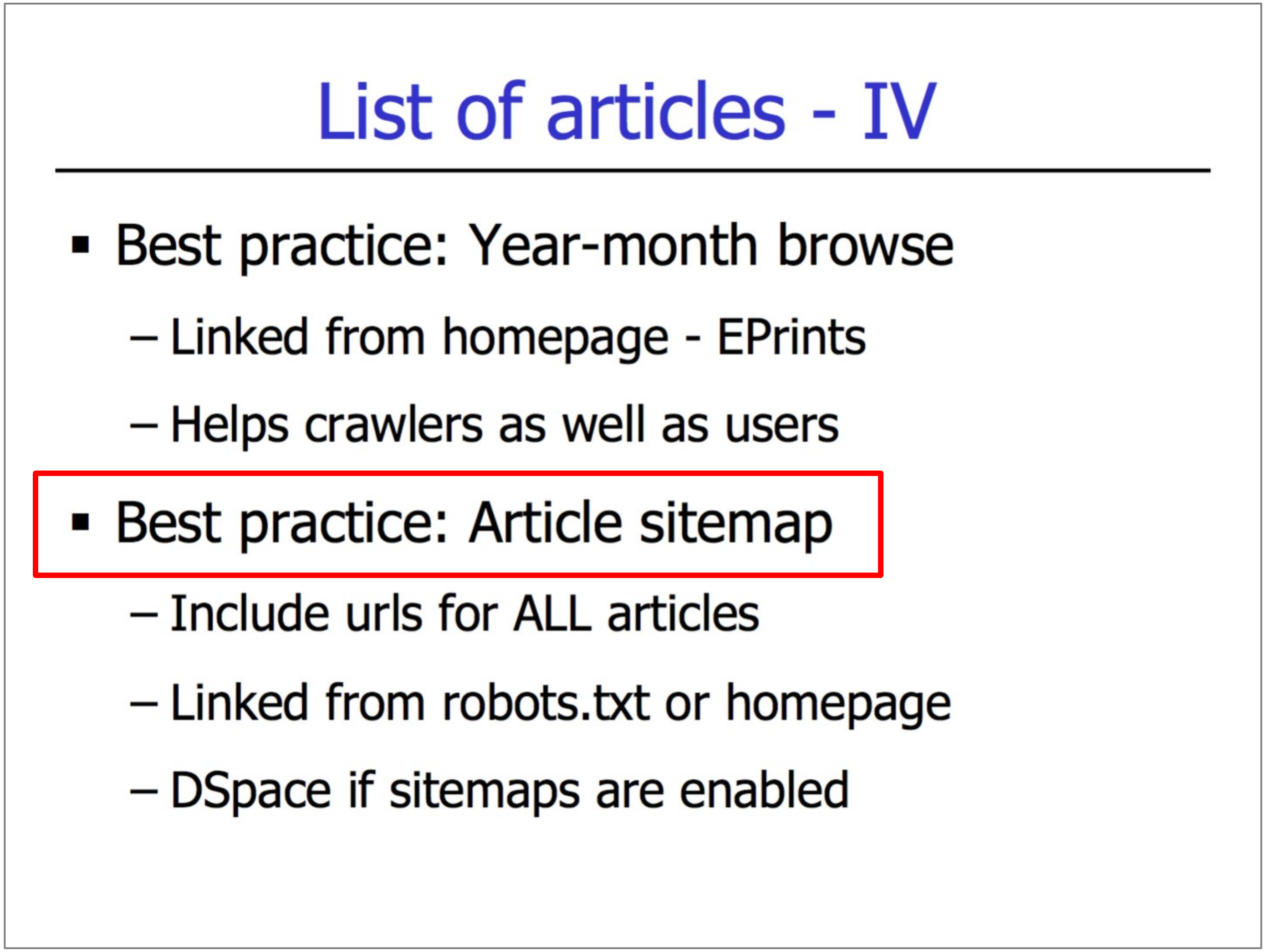
- List of all article urls
- Ability to fetch article urls
- What we index is what the user sees

Scholar

- Identify scholarly articles
- Determine article metadata

Anurag Acharya. Indexing Repositories: Pitfalls & Best Practices. Open Repositories 2015 Keynote.
http://www.or2015.net/wp-content/uploads/2015/06/or-2015-anurag-google-scholar.pdf

Anurag Acharya Told Us Why We Need ResourceSync



List of articles - IV

- Best practice: Year-month browse
  - Linked from homepage - EPrints
  - Helps crawlers as well as users
- Best practice: Article sitemap
  - Include urls for ALL articles
  - Linked from robots.txt or homepage
  - DSpace if sitemaps are enabled

Anurag Acharya. Indexing Repositories: Pitfalls & Best Practices. Open Repositories 2015 Keynote.
http://www.or2015.net/wp-content/uploads/2015/06/or-2015-anurag-google-scholar.pdf

ResourceSync is Based on Sitemaps

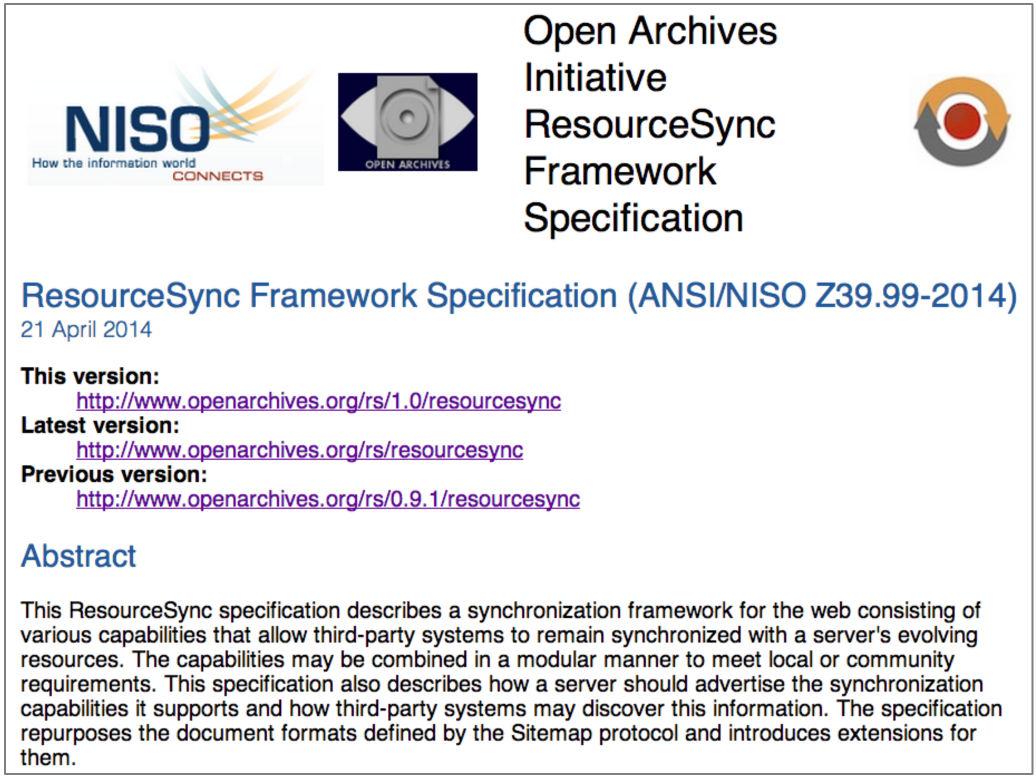
- Sitemap is the document format used throughout the framework
  - Used widely by web servers to advertise their resources to search engines

```
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9”>

  <url>
    <loc>http://example.com/res1</loc>
    <lastmod>2013-01-02T13:00:00Z</lastmod>
  </url>

  <url>
    <loc>http://example.com/res2</loc>
    <lastmod>2013-01-02T14:00:00Z</lastmod>
  </url>
  …
</urlset>
```

ResourceSync, ANSI/NISO Z39.99-2014



Open Archives Initiative ResourceSync Framework Specification

ResourceSync Framework Specification (ANSI/NISO Z39.99-2014)

21 April 2014

This version:
http://www.openarchives.org/rs/1.0/resourcesync
Latest version:
http://www.openarchives.org/rs/resourcesync
Previous version:
http://www.openarchives.org/rs/0.9.1/resourcesync

Abstract

This ResourceSync specification describes a synchronization framework for the web consisting of various capabilities that allow third-party systems to remain synchronized with a server's evolving resources. The capabilities may be combined in a modular manner to meet local or community requirements. This specification also describes how a server should advertise the synchronization capabilities it supports and how third-party systems may discover this information. The specification repurposes the document formats defined by the Sitemap protocol and introduces extensions for them.

- Synchronization of resources from a Source to Destinations
  - Includes exposing repository content to aggregators, search engines
- Applies to any resource with an HTTP URI
- Leverages key ingredients of web interoperability, follow your nose, existing Search Engine Optimization practice

http://www.openarchives.org/rs/toc

Publish Inventory, Changes, Notifications

- Repository communicates about the state of its resources:
    - Publish inventory: snapshot of the state of resources at a moment in time
    - Publish changes: enumeration of resource changes that occurred during a temporal interval
    - Notify about changes: send notifications as changes occur

Payload for Inventory, Changes, Notifications

- A repository may communicate additional information pertaining to each resource:
  - Technical metadata about a resource: content encoding, content length, mime type, content-based hash
  - Links to related resources: mirror copies, alternate representations, resource versions, diff between current and previous version, metadata-to-content link, content-to-metadata link, collection membership, persistent identifier, etc.

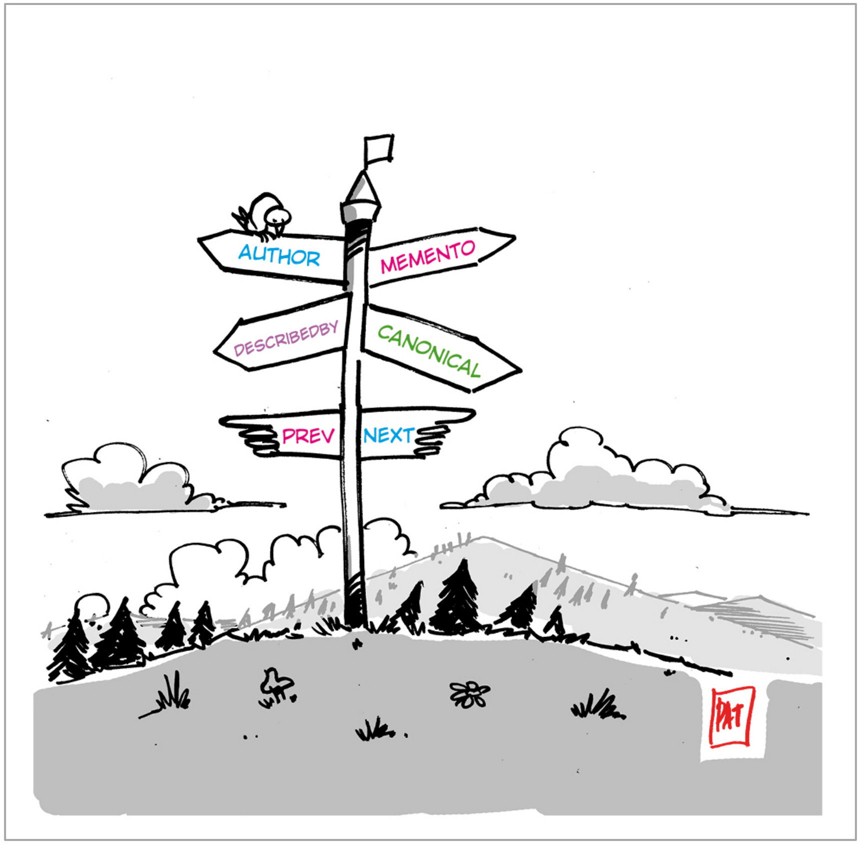
ResourceSync is Based on Sitemaps

- Extensions to Sitemaps:
  - <rs:ln> for links
  - <rs:md> for metadata

```
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9”>
        xmlns:rs="http://www.openarchives.org/rs/terms/”>
  <rs:ln …/>
  <rs:md …/>

  <url>
    <loc>http://example.com/res1</loc>
    <lastmod>2013-01-02T13:00:00Z</lastmod>
    <rs:ln …/>
    <rs:md …/>
  </url>
  …
</urlset>
```

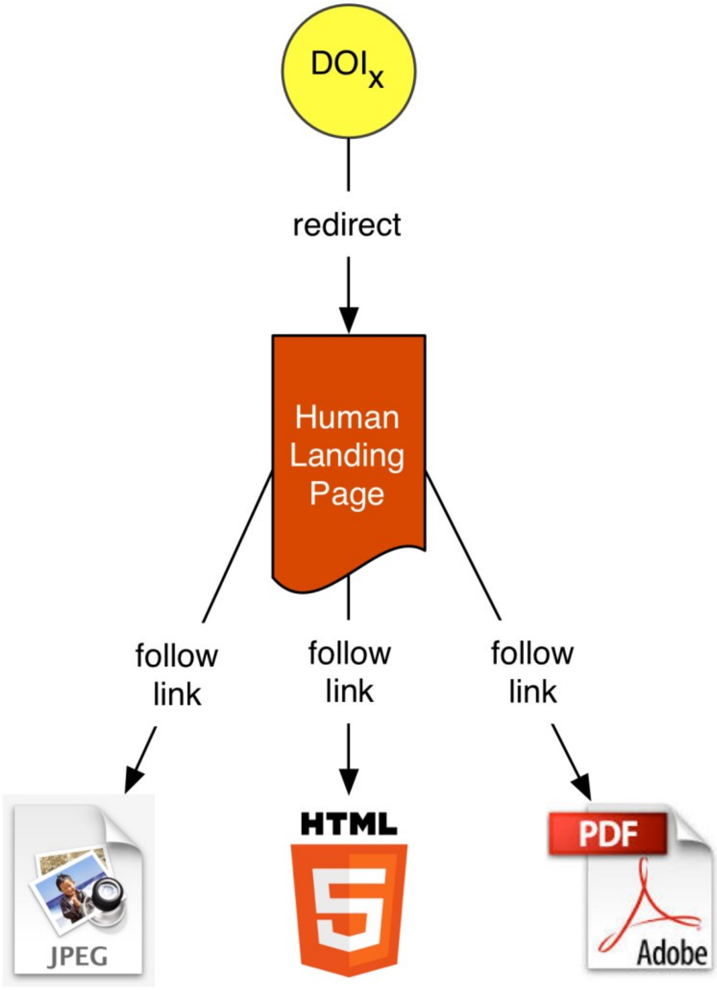
Signposting the Scholarly Web

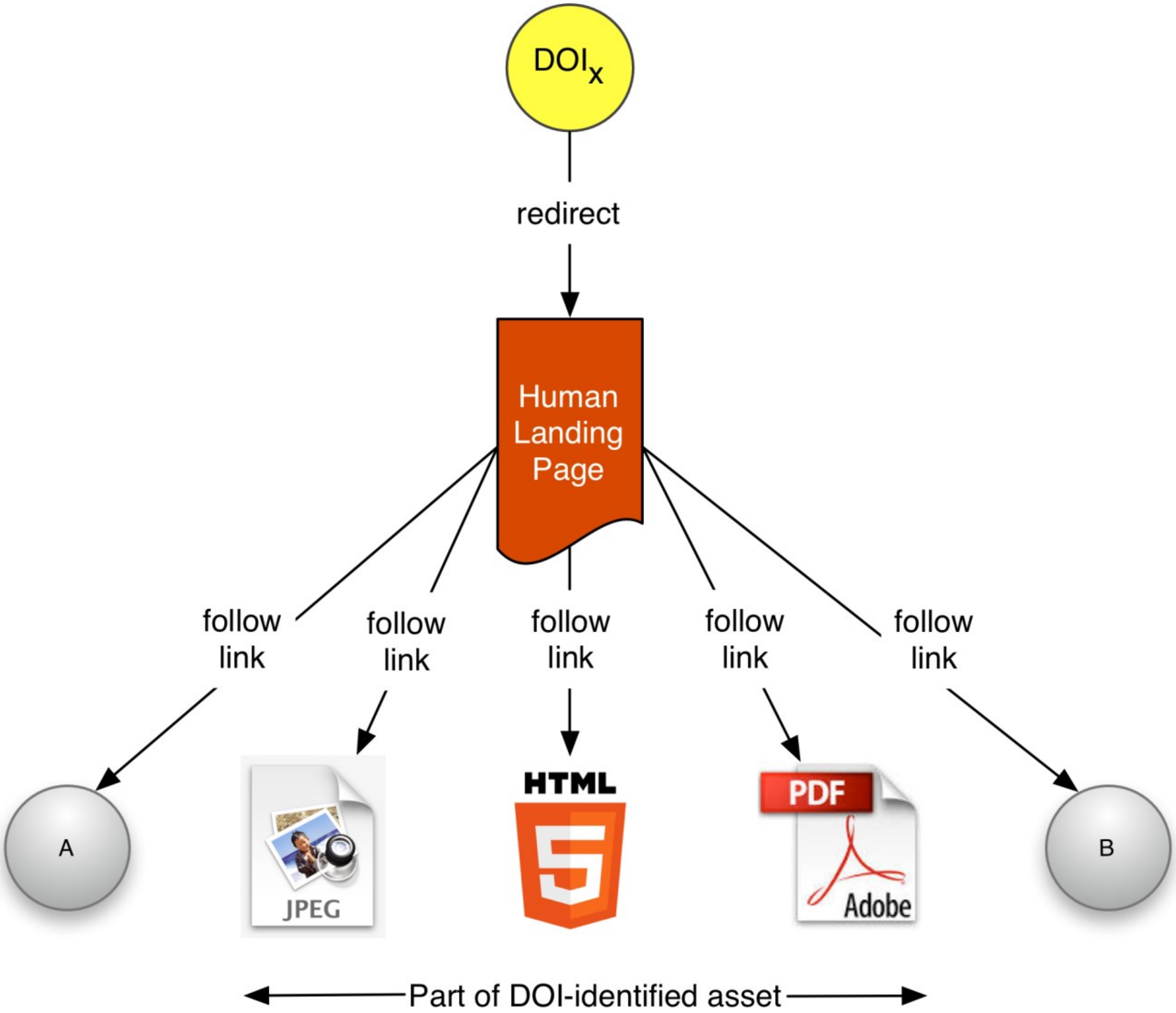


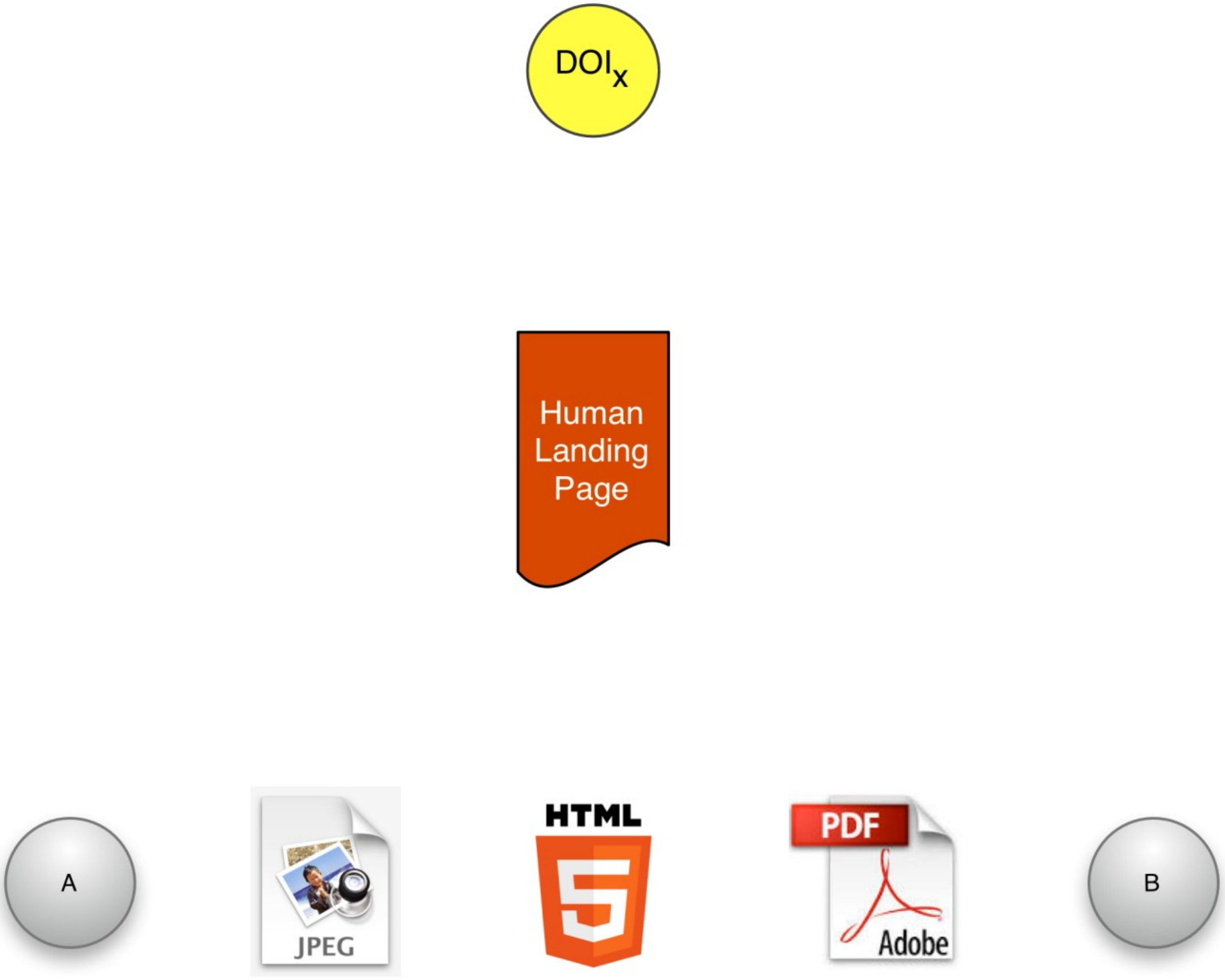
Example pattern: The PID, the Landing Page, the Stuff

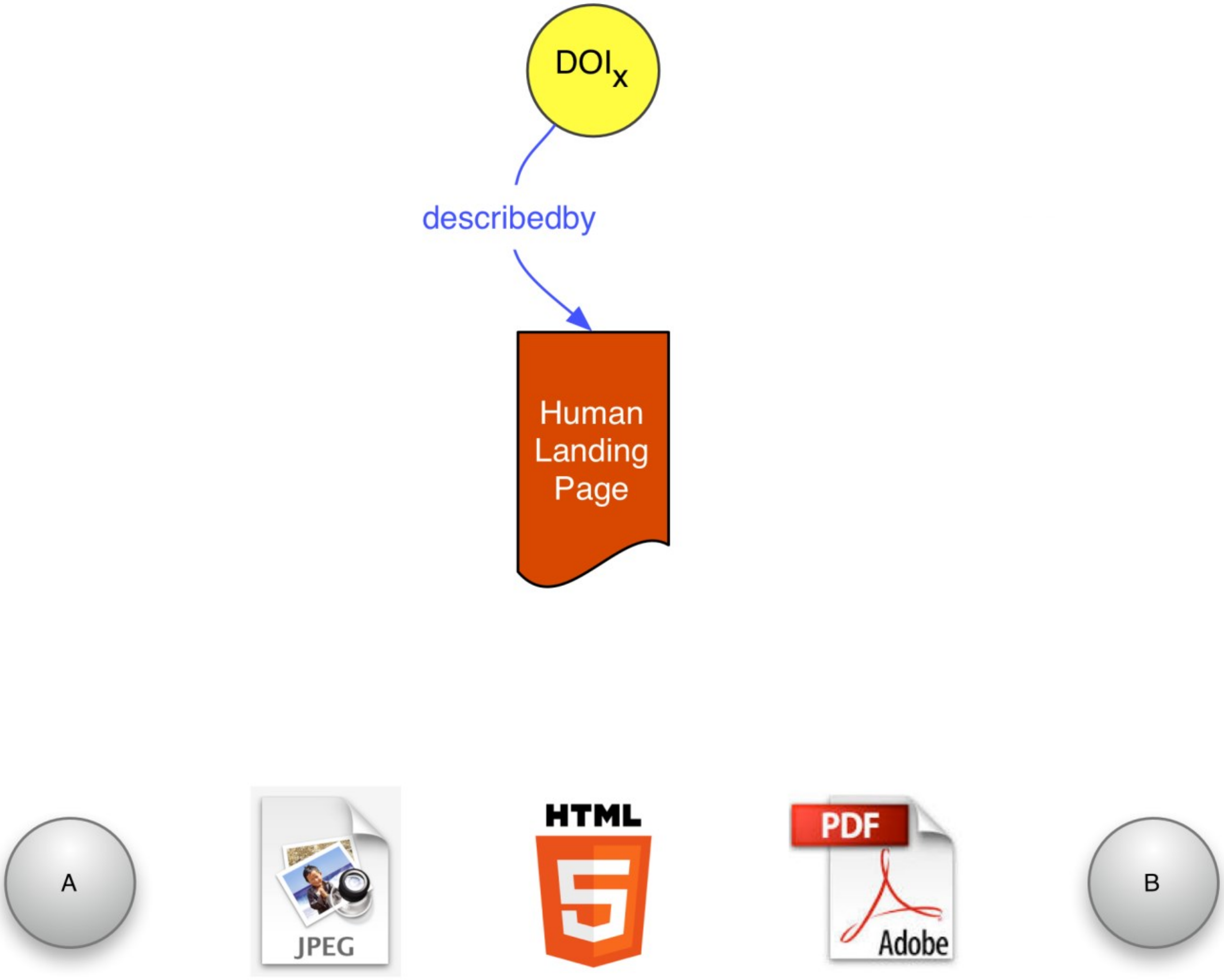
Herbert Van de Sompel
COAR Annual Meeting, Vienna, Austria, 12/04/2016

Los Alamos
NATIONAL LABORATORY


DOI$_x$
redirect
Human Landing Page
follow link
follow link
follow link
JPEG
HTML
PDF
Adobe

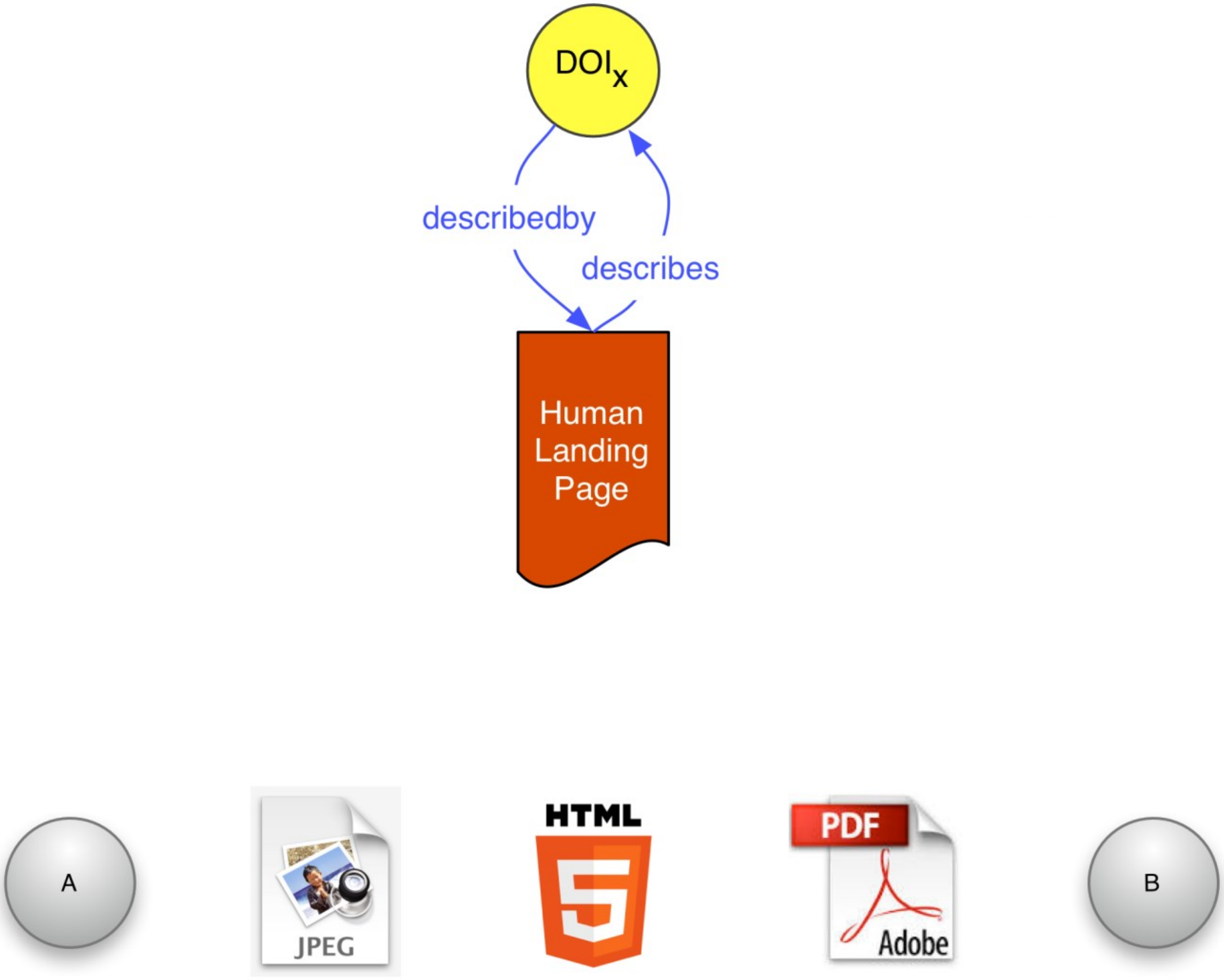

DOI_x
redirect
Human Landing Page
follow link
follow link
follow link
follow link
follow link
A
JPEG
HTML
PDF
Adobe
B
Part of DOI-identified asset

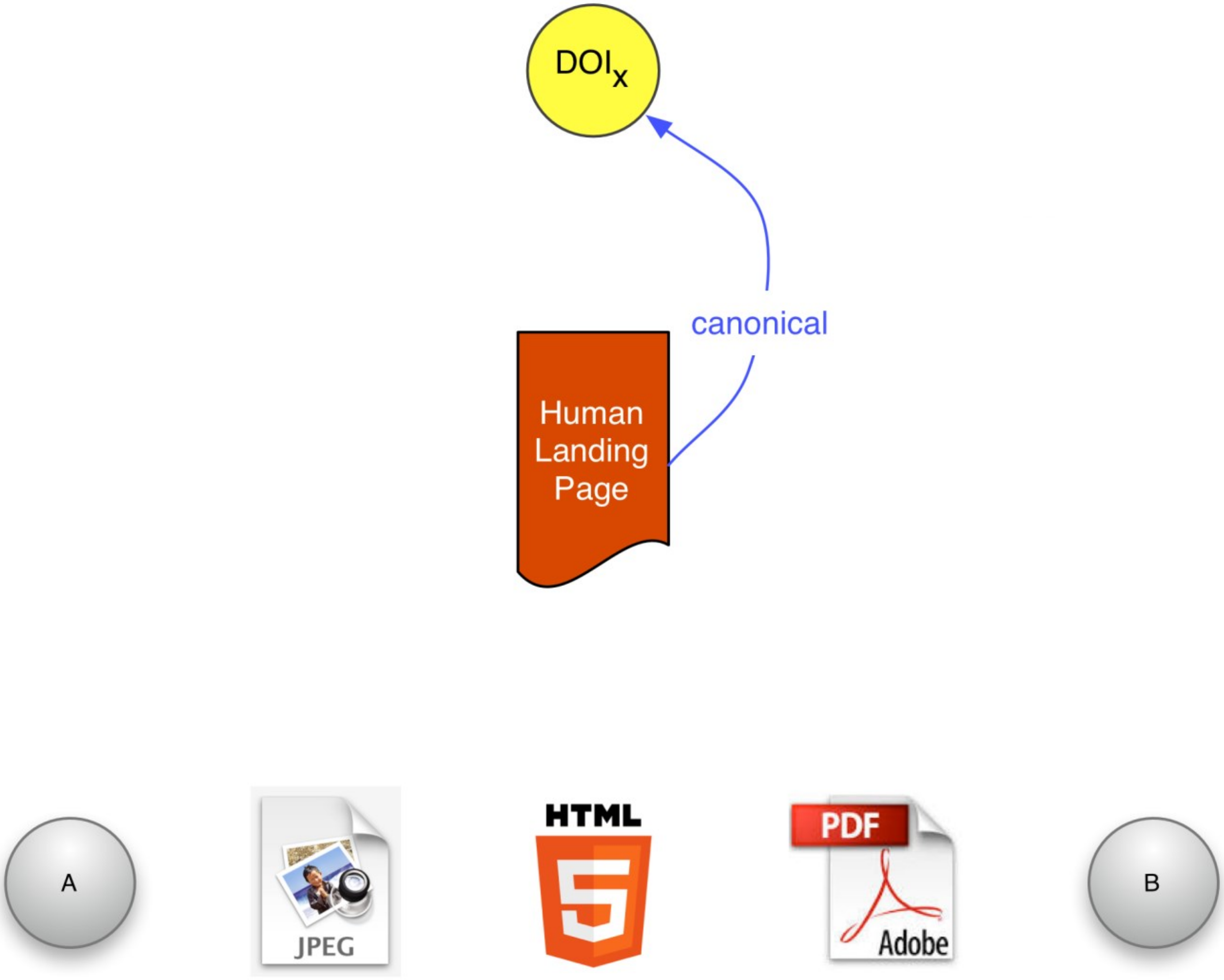

DOI$_x$
Human Landing Page
A
JPEG
HTML
PDF
Adobe
B

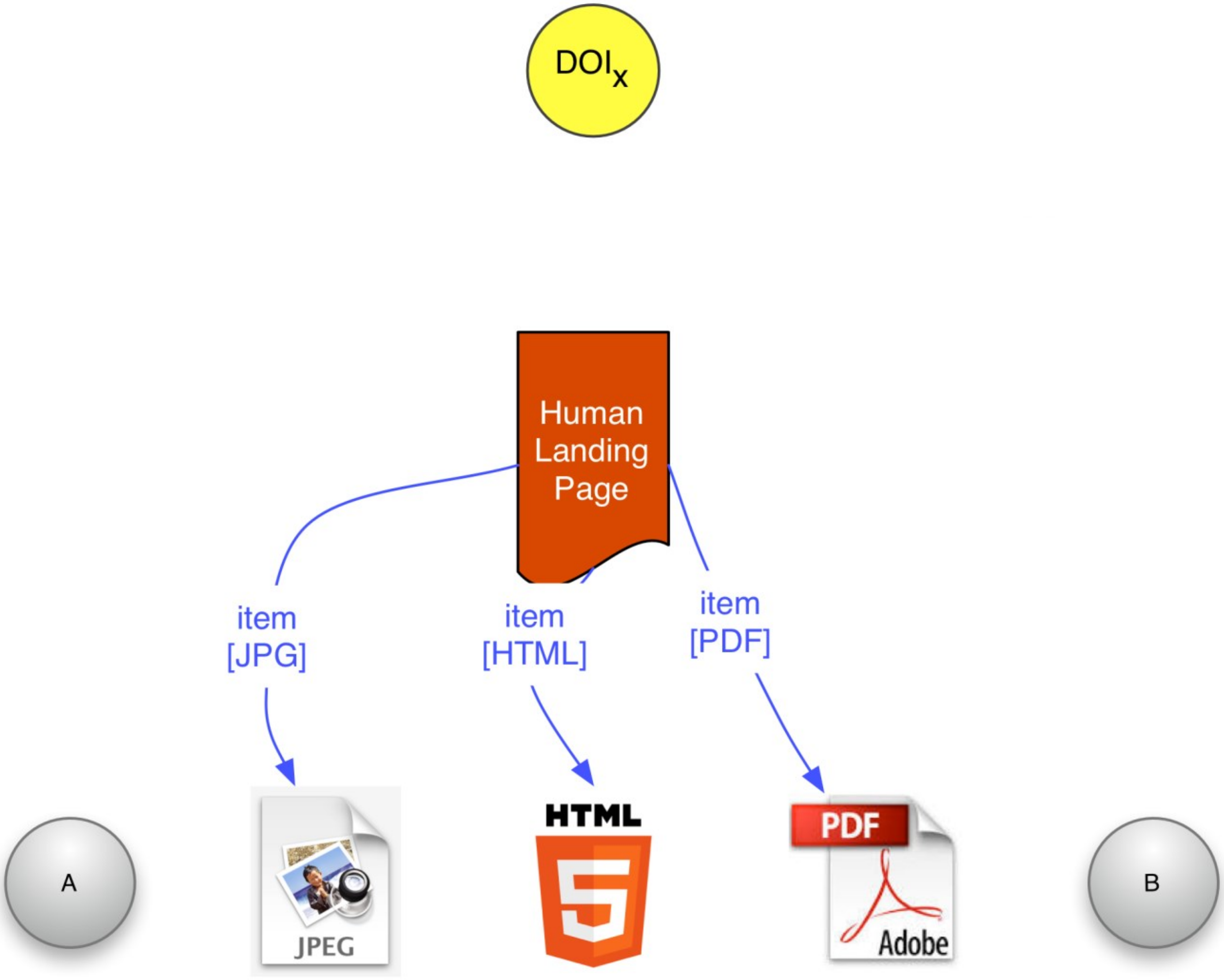

DOIx
describedby
Human Landing Page
A
JPEG
HTML
PDF
Adobe
B

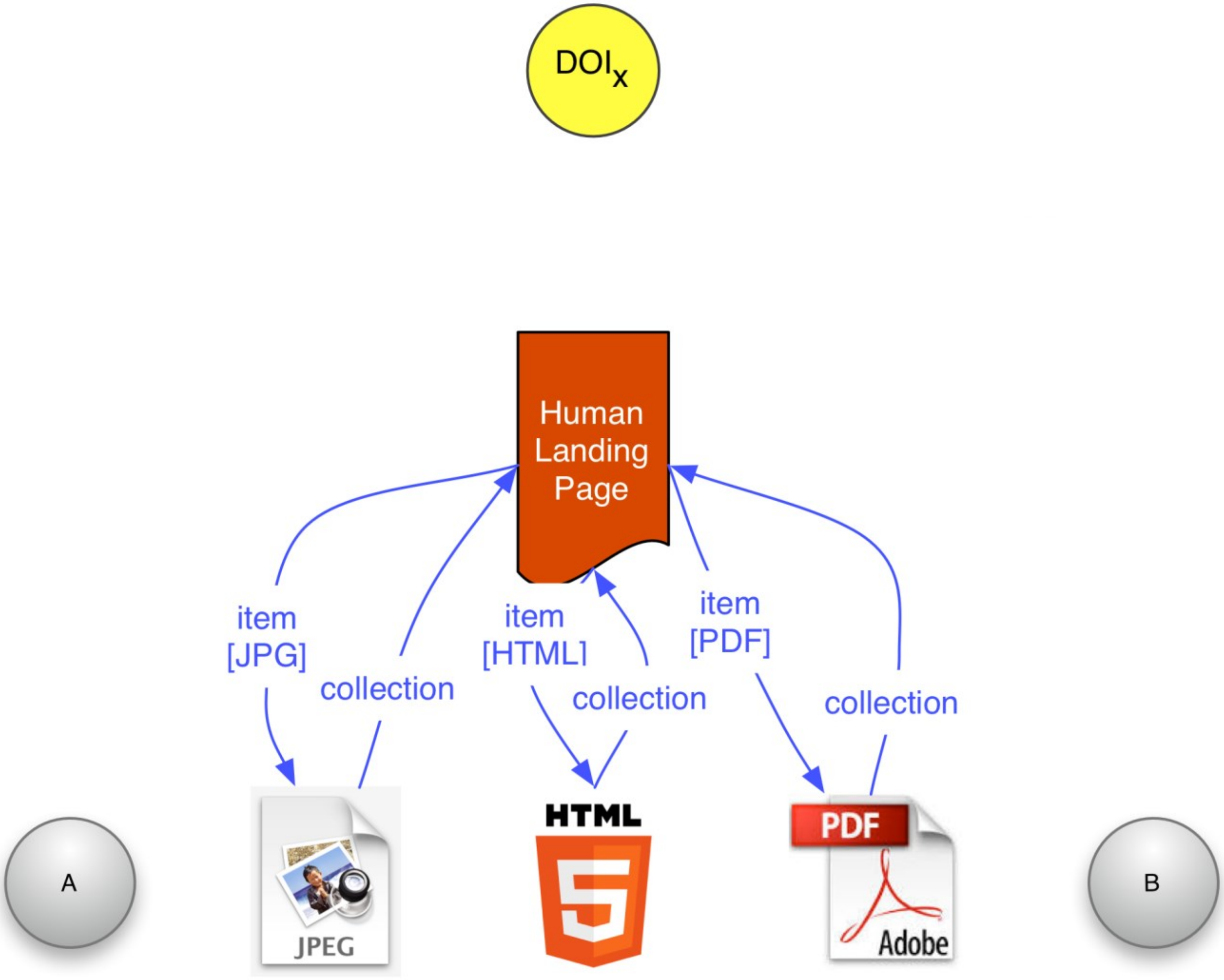
Response to HTTP HEAD on http://dx.doi.org/10.2218/ijdc.v9i1.320

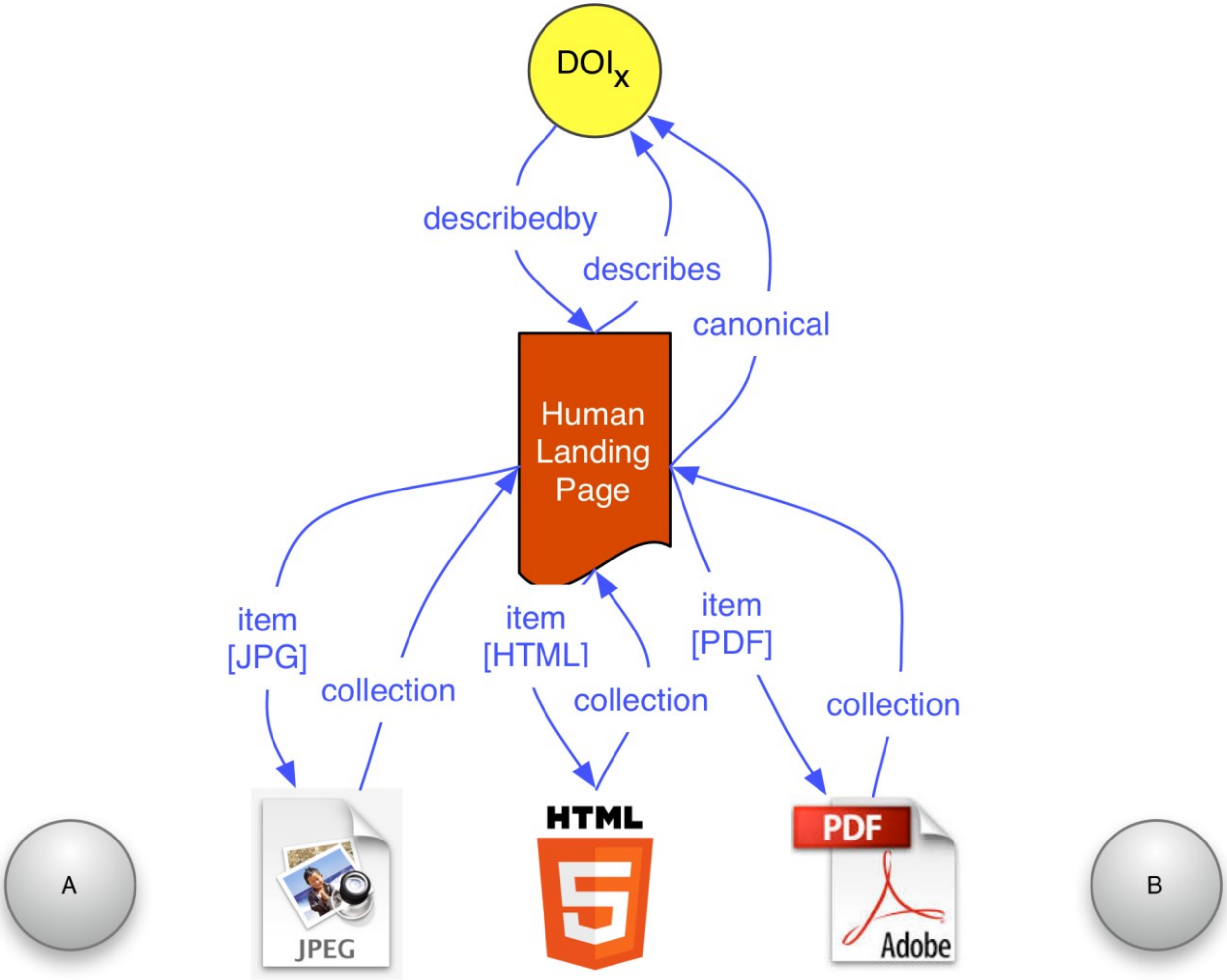
```
HTTP/1.1 303 See Other
Server: Apache-Coyote/1.1
Date: Fri, 9 Jan 2015 16:31:46 GMT
Vary: Accept
Location: http://www.ijdc.net/index.php/ijdc/article/view/320
Link: <http://www.ijdc.net/index.php/ijdc/article/view/320>
 ; rel=” describedby”
 ; type=“text/html”
Content-Length: 188
```


DOI$_x$
describedby
describes
Human Landing Page
A
JPEG
HTML
PDF
Adobe
B


DOI_x
canonical
Human Landing Page
A
JPEG
HTML
PDF
Adobe
B


DOI$_x$
Human Landing Page
item [JPG]
item [HTML]
item [PDF]
A
JPEG
HTML
PDF
Adobe
B


DOI$_x$
Human Landing Page
item [JPG]
collection
item [HTML]
collection
item [PDF]
collection
A
JPEG
HTML
PDF
Adobe
B


DOI$_x$
describedby
describes
canonical
Human Landing Page
item [JPG]
collection
item [HTML]
collection
item [PDF]
collection
JPEG
HTML
PDF
Adobe
A
B

This Allows a Machine Agent …

- To understand that the splash page describes the DOI-identified asset

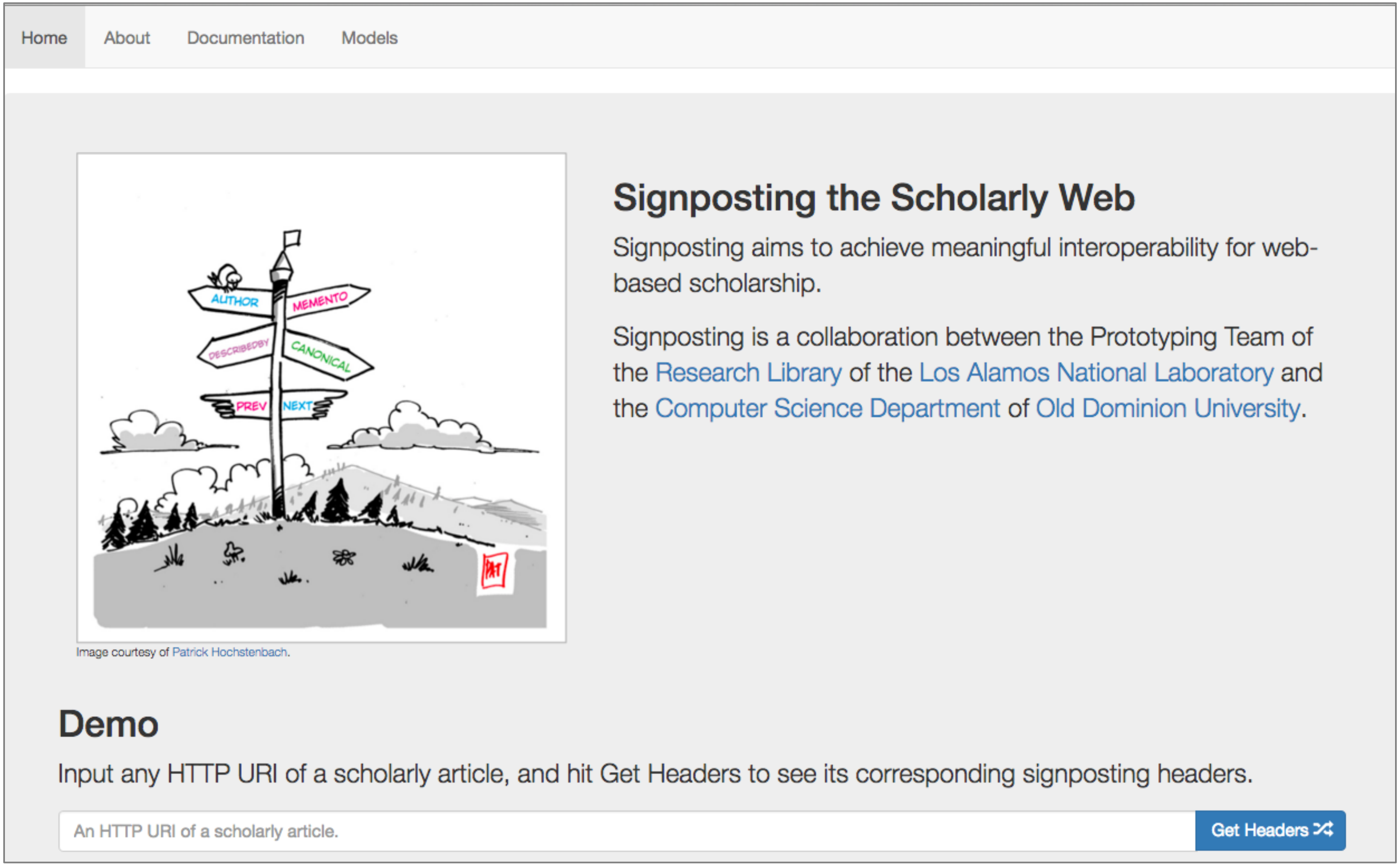
- To determine that resource A is not part of the DOI-identified asset

- To navigate towards the profile of the authors of the asset when landing on any of the constituent resources of the DOI-identified asset

- To understand that a DOI is associated with the PDF, HTML, and JPEG resources and that this DOI should preferably be used to refer to those resources

- To associate annotations made to the HTML page with the DOI

Signposting: Work in Progress



Signposting the Scholarly Web

Signposting aims to achieve meaningful interoperability for web-based scholarship.

Signposting is a collaboration between the Prototyping Team of the Research Library of the Los Alamos National Laboratory and the Computer Science Department of Old Dominion University.

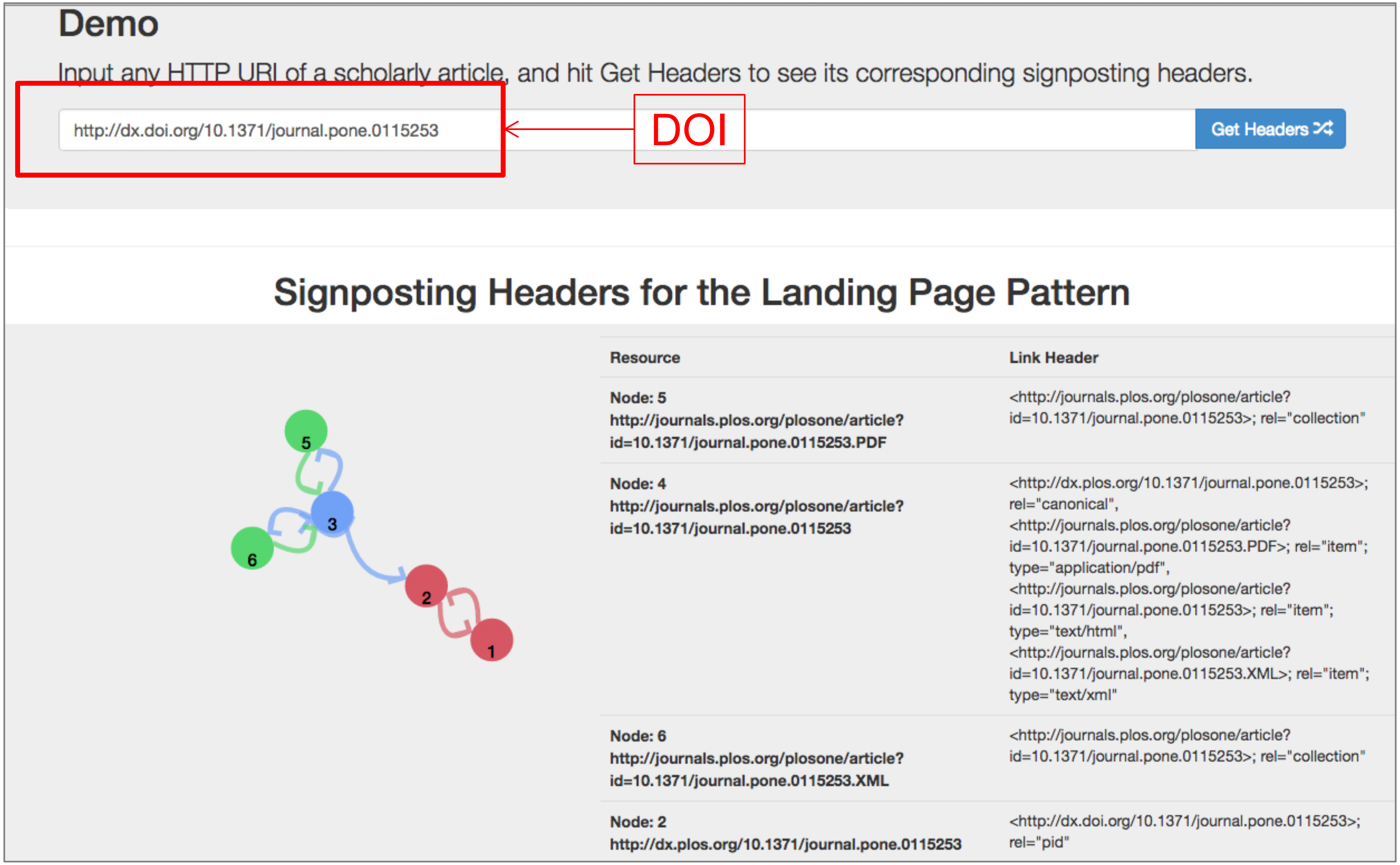
Image courtesy of Patrick Hochstenbach.

Demo

Input any HTTP URI of a scholarly article, and hit Get Headers to see its corresponding signposting headers.

An HTTP URI of a scholarly article.

Signposting: Work in Progress



Demo

Input any HTTP URI of a scholarly article, and hit Get Headers to see its corresponding signposting headers.

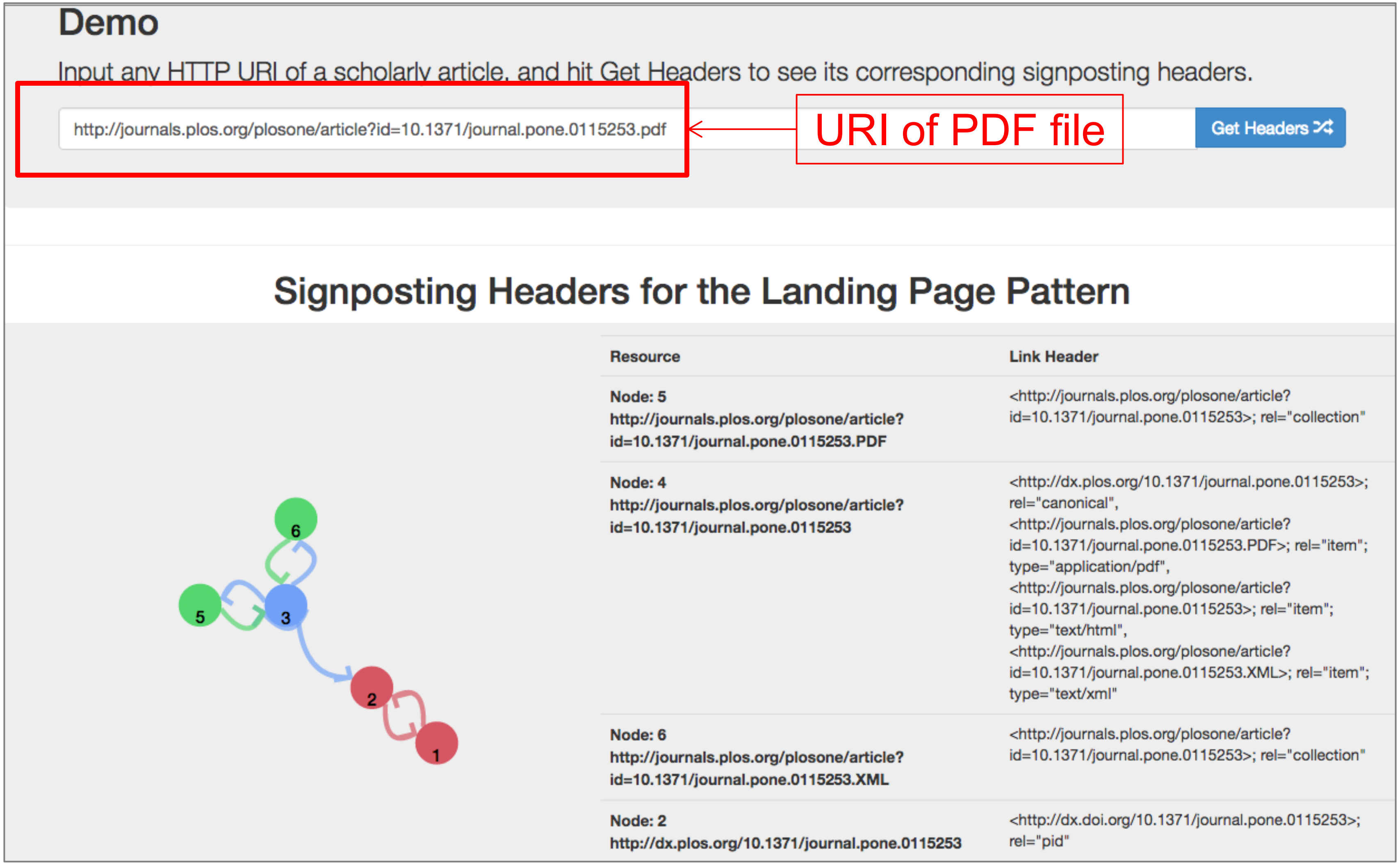
http://dx.doi.org/10.1371/journal.pone.0115253

DOI

Signposting Headers for the Landing Page Pattern

Resource	Link Header
Node: 5 **http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115253.PDF**	<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115253>; rel="collection"
Node: 4 **http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115253**	<http://dx.plos.org/10.1371/journal.pone.0115253>; rel="canonical", <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115253.PDF>; rel="item"; type="application/pdf", <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115253>; rel="item"; type="text/html", <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115253.XML>; rel="item"; type="text/xml"
Node: 6 **http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115253.XML**	<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115253>; rel="collection"
Node: 2 **http://dx.plos.org/10.1371/journal.pone.0115253**	<http://dx.doi.org/10.1371/journal.pone.0115253>; rel="pid"

Signposting: Work in Progress



Demo

Input any HTTP URI of a scholarly article, and hit Get Headers to see its corresponding signposting headers.

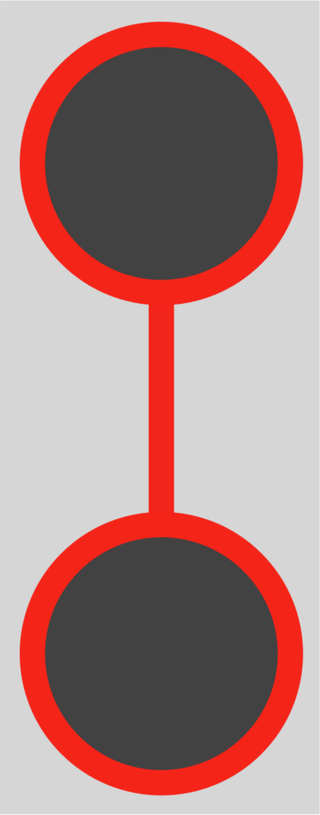
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115253.pdf

URI of PDF file

Signposting Headers for the Landing Page Pattern

Resource	Link Header
Node: 5 **http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115253.PDF**	<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115253>; rel="collection"
Node: 4 **http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115253**	<http://dx.plos.org/10.1371/journal.pone.0115253>; rel="canonical", <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115253.PDF>; rel="item"; type="application/pdf", <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115253>; rel="item"; type="text/html", <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115253.XML>; rel="item"; type="text/xml"
Node: 6 **http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115253.XML**	<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115253>; rel="collection"
Node: 2 **http://dx.plos.org/10.1371/journal.pone.0115253**	<http://dx.doi.org/10.1371/journal.pone.0115253>; rel="pid"

Robust Links

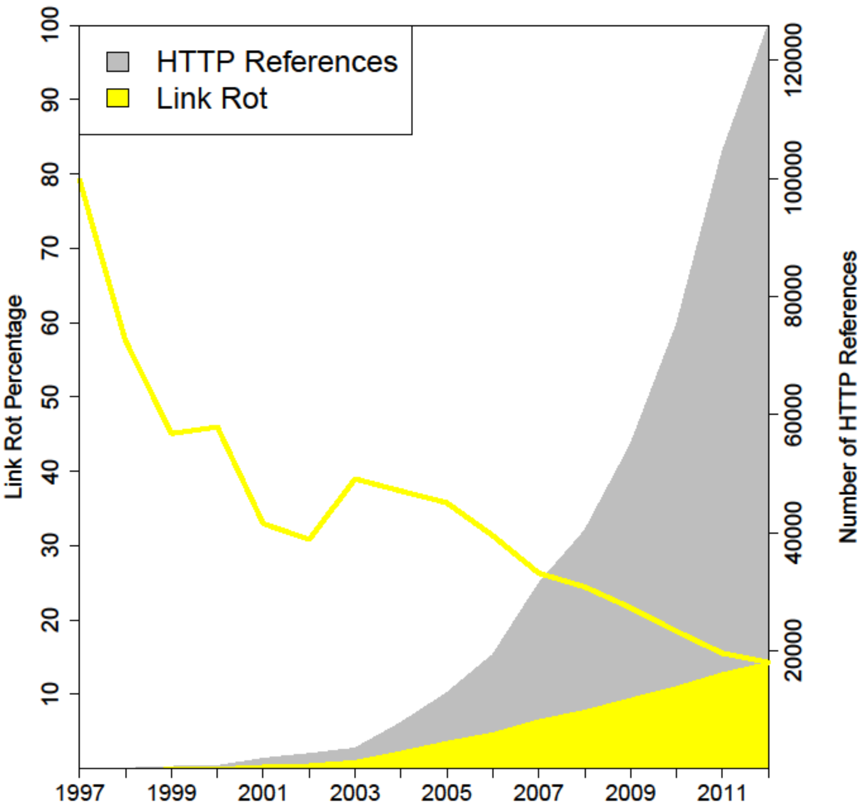


Reference Rot

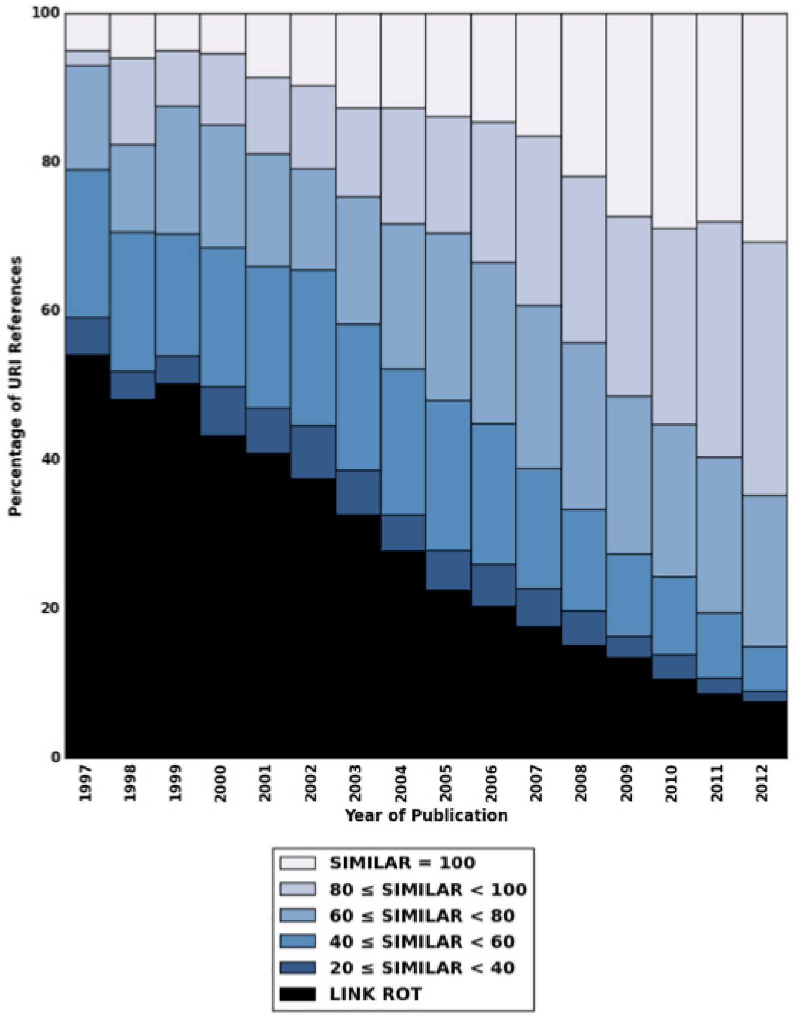
- Links to Web at Large resources are subject to Reference Rot:
  - Link Rot: Link stops working, e.g. HTTP 404 “Not Found”
  - Content Drift: Linked content changes over time
    - Possibly to the extent that it becomes no longer representative of the content that was initially referenced

hiberlink

Link Rot



Content Drift



Martin Klein et al. (2014) Scholarly context not found. In: PLOS ONE
http://dx.doi.org/10.1371/journal.pone.0115253

Combating Reference Rot

① Create a snapshot of the referenced resource in one of many web archives that support on-demand archiving (manual, API):

- archive.today
- Internet Archive
- perma.cc
- webcitation.org

② Reference snapshots actionably by using:

- Original URI
- Snapshot URI
- Date/Time of snapshot

in order to maximize link robustness

http://robustlinks.mementoweb.org/about/

Reference Resources Actionably

- When referencing resources, use Link Decorations to convey Original URI, Snapshot URI, Date/Time

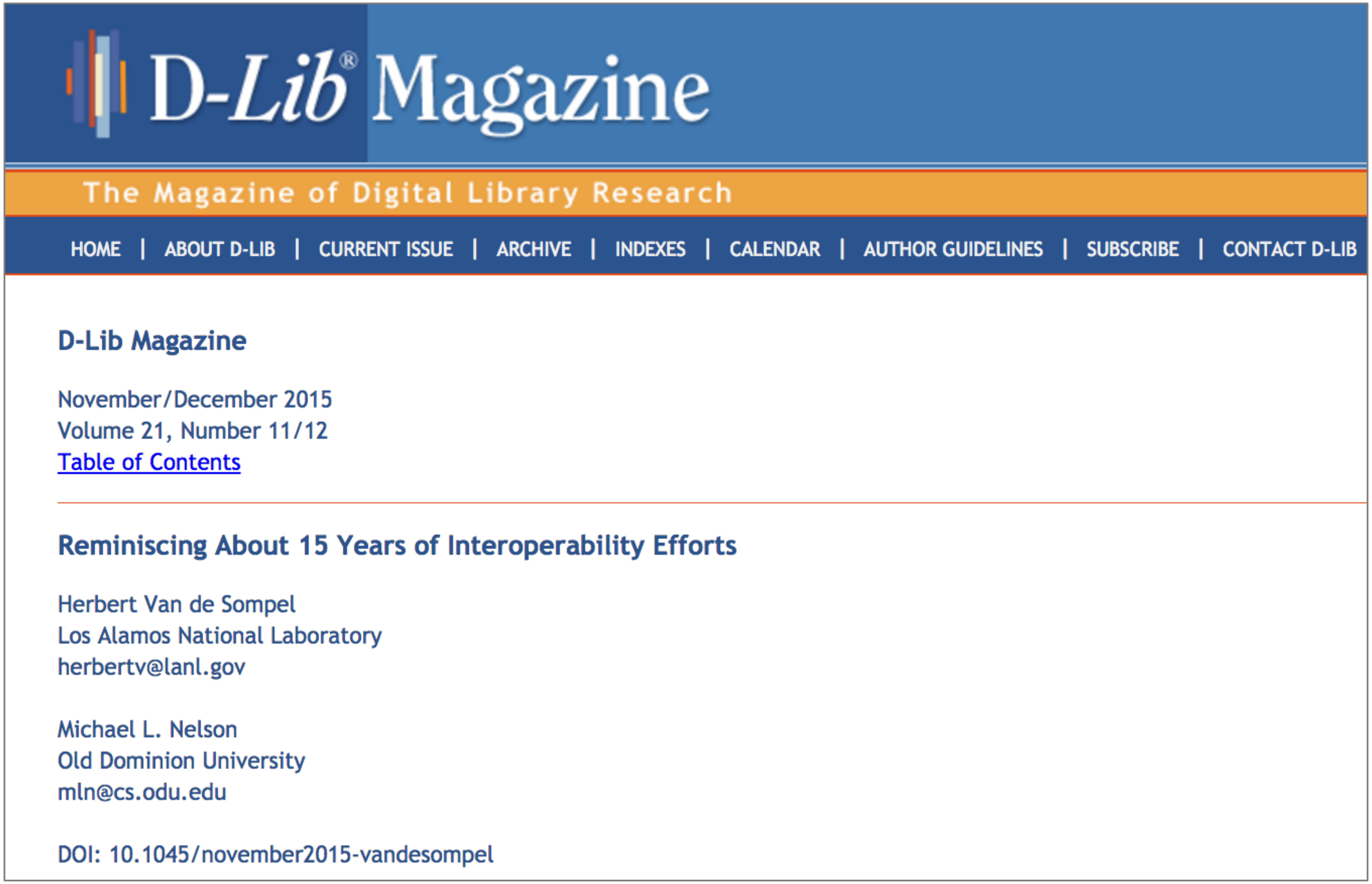
```
<a href=“http://hiberlink.org”
    data-versionurl=“https://archive.is/drFFu”
    data-versiondate=“2015-11-16” >
```

```
<a href=“https://archive.is/drFFu”
    data-originalurl=“http://hiberlink.org”
    data-versiondate=“2015-11-16” >
```

- Legitimate in HTML5
- Can be made actionable with JavaScript, e.g. robustlinks.js

Herbert Van de Sompel et al. (2015) Robust Links - Link Decorations
http://robustlinks.mementoweb.org/spec/

See Robust Links at Work



D-Lib® Magazine
The Magazine of Digital Library Research
HOME | ABOUT D-LIB | CURRENT ISSUE | ARCHIVE | INDEXES | CALENDAR | AUTHOR GUIDELINES | SUBSCRIBE | CONTACT D-LIB

D-Lib Magazine

November/December 2015
Volume 21, Number 11/12
Table of Contents

Reminiscing About 15 Years of Interoperability Efforts

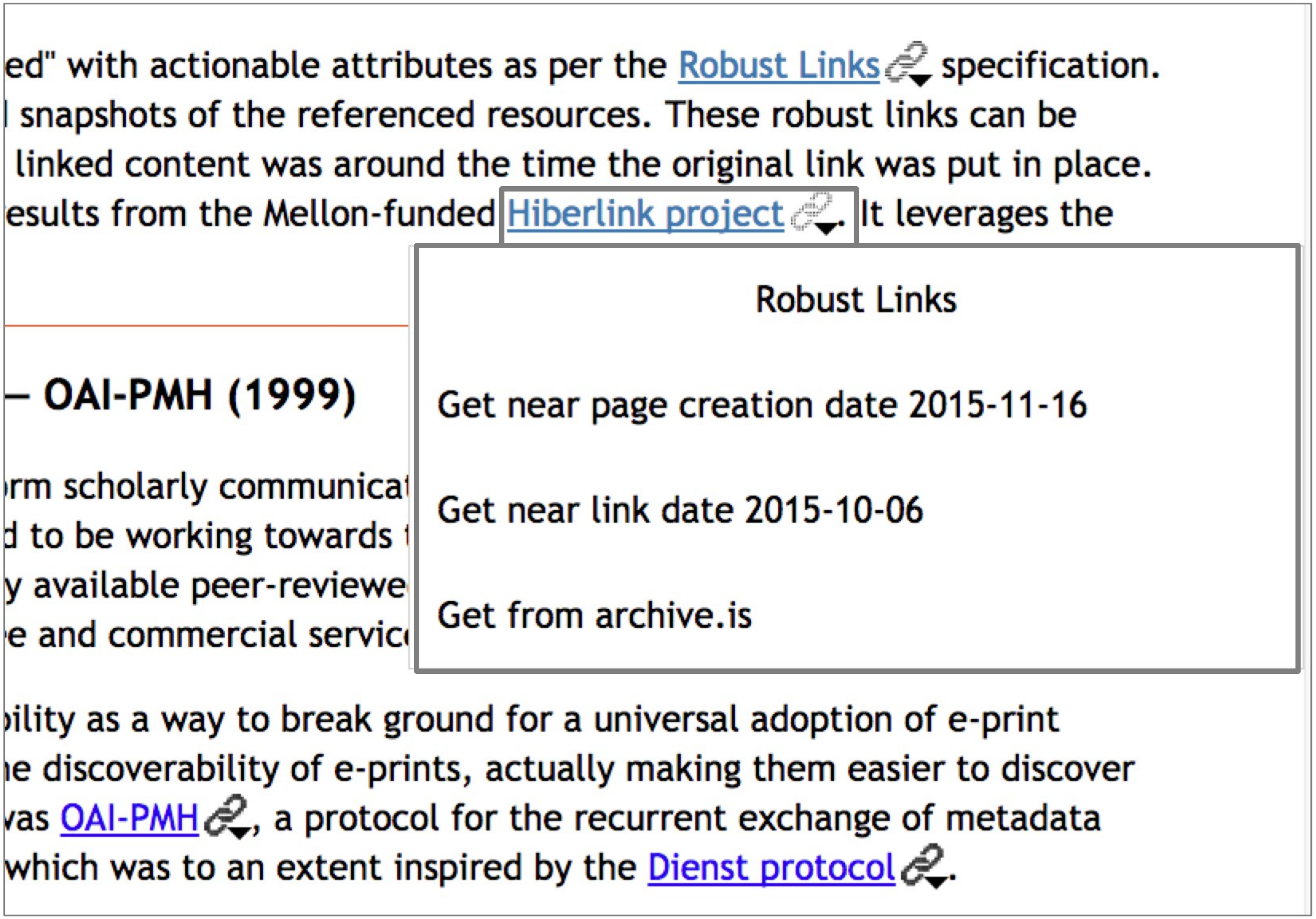
Herbert Van de Sompel
Los Alamos National Laboratory
herbertv@lanl.gov

Michael L. Nelson
Old Dominion University
mln@cs.odu.edu

DOI: 10.1045/november2015-vandesompel

https://dx.doi.org/10.1045/november2015-vandesompel

See Robust Links at Work



ed" with actionable attributes as per the Robust Links specification.
l snapshots of the referenced resources. These robust links can be
linked content was around the time the original link was put in place.
esults from the Mellon-funded Hiberlink project. It leverages the

– OAI-PMH (1999)

rm scholarly communicat
d to be working towards t
y available peer-reviewe
e and commercial servic

Robust Links

Get near page creation date 2015-11-16

Get near link date 2015-10-06

Get from archive.is

ility as a way to break ground for a universal adoption of e-print
e discoverability of e-prints, actually making them easier to discover
vas OAI-PMH, a protocol for the recurrent exchange of metadata
which was to an extent inspired by the Dienst protocol.

https://dx.doi.org/10.1045/november2015-vandesompel

Conclusion

There is no real conclusion. There are insights:

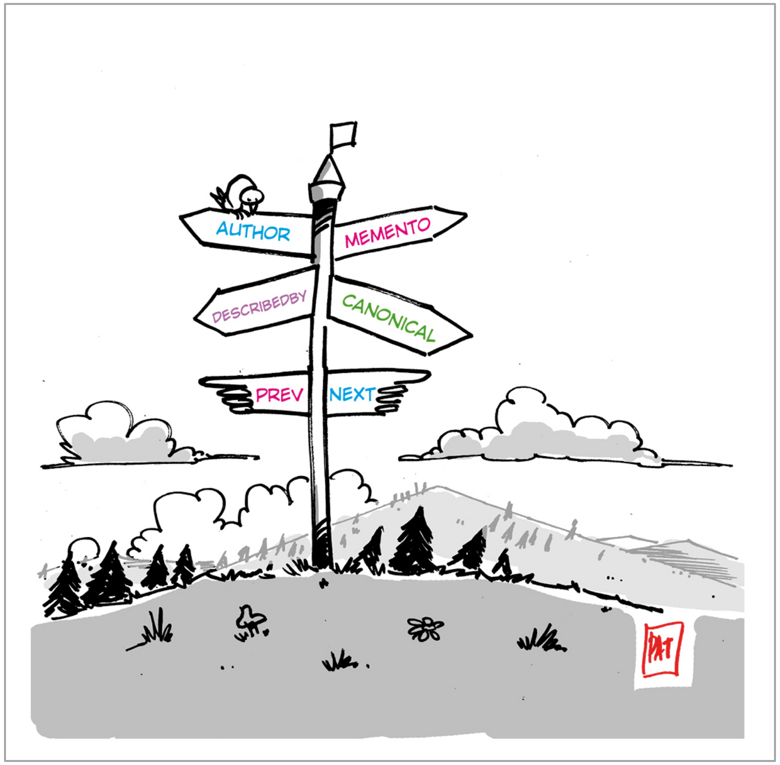
- One doesn't do interoperability because of interoperability but to enable cross-node applications that add value

- Establishing interoperability across a vast amount of nodes is a huge challenge. But meaningful levels of interoperability can be achieved via really basic approaches.

- Unfortunately, not even discovery is a solved problem (although the solution is available)
  - Anurag's keynote is a real embarrassment for our community

Leading organizations and projects should promote web-centric cross-repository interoperability

Establishing New Levels of Interoperability for Web-Based Scholarship



Cartoon by:
Patrick Hochstenbach

Herbert Van de Sompel
Los Alamos National Laboratory
@hvdsomp