Persistent Identifiers for the "Real Web"

David Wood

Eric Miller

The "Real Web"
Moby Dick is a depiction with author Herman Melville.
Moby Dick

1851

Herman Melville

Films

1998 1956 1930 1926

The Sea Beast

We (almost) do this on the World Wide Web
Page Not Found

I'm terribly sorry, but I can't find the page you have requested.

Believe me, I looked for it! Really. You can't even imagine the millions of calculations I had to go thru to look in all the places my owner could have left it. But I couldn't find it :-(

No, it's not your fault, it's all my own. I'm a bad server. I know that. Terrible. I should get a new job one day or another, but what else can I do?

A History of Silos

1970s  A neat little package

1980s  Client-Server

1990s  The Early Web
The Next Great Leap

Extending the Universal Client

Expanding the Universal Connection

Ubiquitous, reusable applications

Web of Data
The Next Great Leap

Extending the Universal Client

Expanding the Universal Connection

Web of Data

Ubiquitous, reusable applications

Providing the Universal Database
The Next Great Leap

Extending the Universal Client

Expanding the Universal Connection

Providing the Universal Database

Ubiquitous, reusable applications

Explaining the Logic

Web of Data

URL Curation
• PURLs provide permanent URLs for information resources on the World Wide Web.

• PURLs work by simple HTTP redirection.

• The Online Computer Library Center (OCLC) has run a public PURL service since 1995 at http://purl.org.

• Used by others for private PURL services (libraries, Internet namespaces, government, some corporations).

• Open Source, open standards.
Evolution

One PURL

Change in Host

Change in Hosting Organization

PURLs

• You might already be using PURLs:

  Mediawiki has a shorthand syntax for PURLs.

  [[[purlnet:scape]]] ➔ http://purl.org/NET/$1
Example:
Find an information resource

PURL server

I want to find Dave's blog

No problem, go to http://prototypo.blogspot.com
Simple Redirection

Client

GET PURL

HTTP 302

“Found” URL

GET URL

Resource

Resource Server

PURL Server

Other Persistent Identifier Schemes

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Community</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOI</td>
<td>Publishers</td>
</tr>
<tr>
<td>LSID</td>
<td>Life sciences</td>
</tr>
<tr>
<td>INFO URIs</td>
<td>Libraries, publishers</td>
</tr>
<tr>
<td>PURLs</td>
<td>Libraries, government</td>
</tr>
</tbody>
</table>
Scaling

• Use of URIs as identifiers grounds your scheme into the Web.
  - PURLs do this
  - DOIs and INFOs partially do
  - LSIDs don’t.

Problems and Solutions

• “PURLs provide one level of indirection, just like a single value DOI name”
  - Not any more.
• “PURL servers don’t know about each other”
  - Federation coming soon.
PURLs for Real-World Resources

• Zepheira and OCLC rewrote the PURL server from scratch to:
  - Modernize the code
  - Create a wider Open Source community
  - Add some new features

• We have a particular interest in supporting permanent Semantic Web URIs.

PURLs for Real-World Resources

• The new PURL server allows the creation of PURLs with these HTTP responses:
  301 Moved Permanently
  302 Found
  303 See Other
  307 Temporary Redirect
  404 Not Found
  410 Gone
• The new PURL server allows the creation of PURLs with these HTTP responses:

  301 Moved Permanently
  302 Found
  **303 See Other**
  307 Temporary Redirect
  404 Not Found
  410 Gone

• The new PURL server allows the creation of PURLs with these HTTP responses:

  **303 See Other**
  Used to identify **physical** and **conceptual** resources
PURLs for Real-World Resources

This allows clients to automatically resolve Semantic Web identifiers without confusing them with information resources.

Example:
Resolve a conceptual resource

I want to find out about Moby Dick
Example: Resolve a conceptual resource

1) Moby Dick is a concept, not a particular information resource.
2) You may find out more about it at http://en.wikipedia.org/wiki/Moby_Dick
Best Practices: Baby Steps

- Identify resources by URIs.
- Use “See Also” PURLs to ensure cross-boundary data integrity.
- Don’t reinvent the wheel: Reuse common public and partner URIs before minting your own.

Best Practices: Next Steps

- Share a PURL service or use more than one.
- Coordinate at the information space level, not the code or even API levels.
- Local control, global access.
Running and Leaping

- Experiment with Active PURLs.
- An Active PURL participates in the provision of metadata about resources it represents.
- Come talk to us later.

PURL Community

- http://purlz.org/
- http://purl.org/
  - Documentation, source code, mailing lists
- See also: