Achieving Meaningful Interoperability for Web-Based Scholarship

Herbert Van de Sompel
Los Alamos National Laboratory
@hvdsomp

Michael L. Nelson
Old Dominion University
@phonedude_mln

Cartoon by:
Patrick Hochstenbach
Evolution of Our Thinking about Interoperability

- Interoperability
  - Repository centric
  - Web centric
    - Descriptive - RDF
    - Navigational - HTTP Links
Evolution of Our Thinking about Interoperability

interoperability

- repository centric
  - OAI-PMH
  - OAI-ORE
- web centric
  - descriptive - RDF
  - navigational - HTTP Links

Memento
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
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

• To a large extent, interoperability across 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

- Interoperability
  - Repository centric
  - Web centric
  - Descriptive - RDF
  - Navigational - HTTP Links
• 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
Those Were the Days

Herbert Van de Sompel & Michael L. Nelson
CNI Membership Meeting, Washington, DC, 12/14/2015
Those Were the Days
The OAMH protocol is a low-barrier interoperability specification for the recurrent exchange of metadata between systems
the Metadata Harvesting protocol

service provider

harvester

6 Requests

replies

data provider

repository

herbert van de sompel
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 GET with GetRecord verb

http://an.oa.org/OAI-script?
A repository replies to a request with an *incomplete list* and a `resumptionToken`;
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
Evolution of Our Thinking about Interoperability

- Interoperability
  - Repository centric
  - Web centric
  - Descriptive - RDF
  - Navigational - HTTP Links
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
Evolution of Our Thinking about Interoperability

interoperability

- repository centric
- web centric

- descriptive - RDF
- navigational - HTTP Links
• 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?

http://www.openarchives.org/ore/1.0/toc
Aiming for New Levels of Cross-Repository Functionality
TICER, Digital Libraries a la Carte, Tilburg, The Netherlands, August 22 2006
Herbert Van de Sompel
ORE Will Allow Web Crawlers to Unambiguously Recover CDO Structure from the Web Graph
Express the \texttt{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
- Typed Link
- Controlled Vocabularies for Typed Links

W3C Architecture of the World Wide Web

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

- interoperability
  - repository centric
  - web centric
    - descriptive - RDF
    - navigational - HTTP Links
• 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

RFC7089 - http://mementoweb.org/guide/rfc/
Memento: Access Versions via the Original URI and a Datetime

Select Date
Dec 14 2005

Dec 12 2005

Internet Archive
Original Resource and Mementos
Bridge from Present to Past

Original Resource

URI-R

HTTP Link
timegate

TimeGate

URI-G

URI-M_1

URI-M_2

Mementos

Apr 10 2001
representation

Aug 15 2007
representation

current
representation

Herbert Van de Sompel & Michael L. Nelson
CNI Membership Meeting, Washington, DC, 12/14/2015
Bridge from Present to Past
Bridge from Past to Present
Tools of the Web-Centric Interoperability Trade – HTTP Stack

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

- Typed Link
- Controlled Vocabularies for Typed Links

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)
  • Finder 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
Signposting the Scholarly Web
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
HTTP HEAD on URI of Paper, As Is

curl -I http://www.dlib.org/dlib/november15/vandesompel/11vandesompel.html

HTTP/1.1 200 OK
Date: Tue, 17 Nov 2015 10:32:51 GMT
Server: Apache/2.2.15 (CentOS)
Last-Modified: Mon, 16 Nov 2015 17:16:55 GMT
ETag: "1c037b-bdd2-51e9e99a5cfc0"
Accept-Ranges: bytes
Content-Length: 48594
Content-Type: text/html; charset=UTF-8
HTTP HEAD on URI of Paper, Signposted with “author” Link

curl -I http://www.dlib.org/dlib/november15/vandesompel/11vandesompel.html

HTTP/1.1 200 OK
Date: Tue, 17 Nov 2015 10:32:51 GMT
Server: Apache/2.2.15 (CentOS)
Last-Modified: Mon, 16 Nov 2015 17:16:55 GMT
ETag: "1c037b-bdd2-51e9e99a5cfc0"
Accept-Ranges: bytes
Content-Length: 48594
Link: <http://orcid.org/0000-0002-0715-6126> ; rel="author"
; type="application/rdf+xml"
Content-Type: text/html; charset=UTF-8
Follow Typed Link in Search of Author Profile

curl -i -H "Accept: application/rdf+xml" http://orcid.org/0000-0002-0715-6126

HTTP/1.1 307 Temporary Redirect
Content-Type: application/rdf+xml; qs=2; charset=UTF-8
Location: https://pub.orcid.org/experimental_rdf_v1/0000-0002-0715-6126
Connection: keep-alive
Obtain Author Profile

curl https://pub.orcid.org/experimental_rdf_v1/0000-0002-0715-6126

<rdf:RDF>
  <rdf:Description rdf:about="http://orcid.org/0000-0002-0715-6126">
    <foaf:page rdf:resource="https://twitter.com/hvdsomp"/>
    <rdfs:label rdf:type="http://www.w3.org/ns/prov#Person">
      Van de Sompel</rdfs:label>
    <foaf:givenName>Herbert</foaf:givenName>
    <foaf:page rdf:resource="https://www.youtube.com/user/hvdsomp"/>
    <foaf:page rdf:resource="http://www.slideshare.net/hvdsomp"/>
  </rdf:Description>
</rdf:RDF>
Scenario 1: The PID, the Splash Page, the Stuff

https://www.youtube.com/watch?v=deejMy4-zTU
Herbert Van de Sompel & Michael L. Nelson
CNI Membership Meeting, Washington, DC, 12/14/2015
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
Herbert Van de Sompel & Michael L. Nelson
CNI Membership Meeting, Washington, DC, 12/14/2015
Herbert Van de Sompel & Michael L. Nelson
CNI Membership Meeting, Washington, DC, 12/14/2015
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
Scenario 2: Resource Versions
Herbert Van de Sompel & Michael L. Nelson

CNI Membership Meeting, Washington, DC, 12/14/2015

http://mementoweb.org/guide/howto/
Herbert Van de Sompel & Michael L. Nelson

CNI Membership Meeting, Washington, DC, 12/14/2015

http://mementoweb.org/guide/howto/
This Allows a Machine Agent …

- To access the version that was operational at a specified date/time
- To navigate between versions
- To understand what the generic URI is from which, at any moment in time, the then-current resource version is available
- To obtain a standardized list of versions and their creation date/times
Scenario 3: The (Code) Snapshot in Zenodo
GitHub

S

Snapshot

DOI$_x$

Human Landing Page

S1

S2

S3

Z

S1 is S at t1

S2 is S at t2

S3 is S at t3

Z is a snapshot of S3, which is S at time t3
GitHub

S

Snapshot

DOI

describedby canonical

Human Landing Page

item collection

S1 is S at t1
S2 is S at t2
S3 is S at t3

Z is a snapshot of S3, which is S at time t3
GitHub

S

Snapshot

DOI

described by canonical

Human Landing Page

item collection

S1 is S at t1

S2 is S at t2

S3 is S at t3

Z is a snapshot of S3, which is S at time t3
GitHub

S

duplicate

duplicate

S1 is S at t1
S2 is S at t2
S3 is S at t3

GitHub

Snapshot

DOI

describedby canonical

Human Landing Page

collection

item

original

memento-datetime: t3

Z is a snapshot of S3, which is S at time t3

S

Herbert Van de Sompel & Michael L. Nelson
CNI Membership Meeting, Washington, DC, 12/14/2015
Herbert Van de Sompel & Michael L. Nelson  
CNI Membership Meeting, Washington, DC, 12/14/2015
This Allows a Machine Agent …

- To understand that the snapshot is a duplicate of a specific version of the code in GitHub

- To navigate from the snapshot to the current state of the GitHub repository

- To navigate from the snapshot to the version of the code in GitHub that was operational at a specified date/time (if, eventually, GitHub supports the Memento protocol)

- To navigate from the specific version of the code in GitHub to its snapshot

- To understand that a DOI is associated with a specific version of GitHub code
See Also

Enhancing integrated environmental modelling by designing resource-oriented interfaces

Carlos Granell \( ^{a,*} \), Laura Díaz \( ^{b} \), Sven Schade \( ^{a} \), Nicole Ostländer \( ^{a} \), Joaquín Huerta \( ^{b} \)

\( ^{a} \) Institute for Environment and Sustainability, European Commission – Joint Research Centre, Ispra, Italy
\( ^{b} \) Institute of New Imaging Technologies, Universitat Jaume I, Castellón, Spain

Abstract

Integrated environmental modelling is gaining momentum for addressing grand scientific challenges such as monitoring the environment for change detection and forecasting environmental conditions along with the consequences for society. Such challenges can only be addressed by a multi-disciplinary approach, in which socio-economic, geospatial, and environmental information becomes inter-connected. However, existing solutions cannot be seamlessly integrated and current interaction paradigms prevent mainstream usage of the existing technology. In particular, it is still difficult to access...
Achieving Meaningful Interoperability for Web-Based Scholarship

Herbert Van de Sompel
Los Alamos National Laboratory
@hvdsomp

Michael L. Nelson
Old Dominion University
@phonedude_mln

Cartoon by:
Patrick Hochstenbach