Reminiscing about Fifteen Years of Interoperability Efforts

Herbert Van de Sompel
@hvdsomp

Cartoon by:
Patrick Hochstenbach

Herbert Van de Sompel
Data and Computing Infrastructures for Open Scholarship, Paris RDA, 22/09/2015
Research Communication & Research Process on the Web

• A highly distributed activity
  • No “winner takes all”
  • No single player that can dictate the rules for the entire environment

• 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

• 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
Repository-Centric Interoperability Paradigm

Address interoperability challenges from the perspective of a node

- The node at the center of the universe
The OAMH protocol is a low-barrier interoperability specification for the recurrent exchange of metadata between systems.
the Metadata Harvesting protocol

6 Requests

service provider

Replies

data provider

herbert van de sompel
Repository-Centric Interoperability Paradigm

Address interoperability challenges from the perspective of a node

- The node at the center of the universe
- Piggybacking on the web without truly embracing its core technologies
3.1.1.1 Encoding an OAI-PMH request in a URL for an HTTP GET

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://an.oa.org/OAI-script?

HTTP GET with GetRecord verb
A repository replies to a request with an *incomplete list* and a `resumptionToken`.
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
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?
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 – OAI-ORE

- 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, PROV, Research Objects, …

- Address a specific problem
- Provide extensive descriptive expressiveness
- Typically based on publishing additional 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
- Appropriate for in-depth interoperability among select nodes?
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 to Temporal Resource Versions

Original Resource

TimeGate

Mementos

HTTP Link original
HTTP Link timegate
HTTP Link original

HTTP GET Accept-Datetime: t_i
HTTP GET Accept-Datetime: t_j

URI-R
URI-G
URI-M_1
URI-M_2

HTTP Link memento

Herbert Van de Sompel
Data and Computing Infrastructures for Open Scholarship, Paris RDA, 22/09/2015
Tools of the Web-Centric Interoperability Trade - Memento

- Resource
- URI
- HTTP as the API
- Representation
- Media Types
- Link
- Content Negotiation, i.e. for preferred Datetime
- 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:

- Address a broad problem
- Provide coarse expressiveness
- Typically based on publishing links that support discovery of resources of certain types
- Low implementation barrier because of familiar technology stack
- Significant Return On Investment
- Appropriate for coarse interoperability among nodes?
Signposting the Scholarly Web

https://www.youtube.com/watch?v=deejMy4-zTU
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

https://dx.doi.org/10.1016/j.envsoft.2012.04.013
Scenario 1: The PID, the Splash Page, the Stuff
Response to HTTP GET 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
Human Landing Page

DOI

canonical

ORCID_a

ORCID_b

A

JPEG

HTML

PDF

B

Herbert Van de Sompel
Data and Computing Infrastructures for Open Scholarship, Paris RDA, 22/09/2015
Herbert Van de Sompel
Data and Computing Infrastructures for Open Scholarship, Paris RDA, 22/09/2015
Scenario 2: The (Code) Snapshot in Zenodo
S

C

a snapshot of S at t3

S1 @t1  S2 @t2  S3 @t3  S4 @t4
S

original

C

Memento-Datetime: t3

S1
@t1

S2
@t2

S3
@t3

S4
@t4
Scenario 3: Establishing Connection Between Paper and Data
mentions D

P

HTTP HEAD

D

does not mention P
HTTP POST

source=P
target=D
Scenario 4: Citation Notification
Herbert Van de Sompel
Data and Computing Infrastructures for Open Scholarship, Paris RDA, 22/09/2015
HTTP GET
accept: application/rdf+xml
Reminiscing about Fifteen Years of Interoperability Efforts

Cartoon by: Patrick Hochstenbach

Herbert Van de Sompel @hvdsomp