Interoperability, Integration, and Distributed Data

Eric Kansa
UC Berkeley / OpenContext.org

Unless otherwise indicated, this work is licensed under a Creative Commons Attribution 3.0 License <http://creativecommons.org/licenses/by/3.0/>
Why this Topic Matters

- Lot’s of emphasis on semantics and standardization
- Need greater recognition for how **Web architecture** impacts (re)use potential of data
Key Concepts for the Web

Key issues for interoperability

• Resources: Anything that can be identified and retrieved with a URL/URI.
  - Static documents or files
  - Dynamically generated results of a query
  - Abstract Concepts

• Representations: Different views or expressions of a resource in different formats
  - Mobile phone view
  - Machine-readable view
Web Resources

Dynamic service / query results (Google)

Abstract concept ("place") in Pleiades
Web Resources

Dynamic service / query results (Google)

Abstract concept (“place” in general) in Pleiades
Web Resources

Key Common Requirement:
URL/URI
Representations
Representations
Representations of a Resource

- Browser (XHTML) and “Machine-readable” (ArchaeoML-XML) representations
Representations of a Resource

Nomisma.org (coin hordes)

Adapted from Inventory of Greek Coin Hoards

247 Diokles, on the coast 20 km. S of Aegina. Aegina. 500s

Burial: 2. 348 B.C. C1

Contents: 8000 drachms

Korone: 1 triobol

Athenian League: 3 triobols

Korone: 1 triobol

Phalere: 4 triobols

Athenian League: 25 triobols

Thasos: 7 triobols

Athens: 5 drachms

Athens: 6 tetradrachms

Phalere: 445 triobols (12 early, 320 late)

ATH Coin Shop 73545

ATH Coin Shop 83222

ATH Coin Shop 34143

Athens: 1 triobol

Athenian League: 150 triobols

ATH Coin Shop 77520

ATH Coin Shop 30011

ATH Coin Shop 47354

Corone: 2 tetradrachms

Athenian League: 351 triobols (122 early, 359 late)

Corone: 225 triobols

Athenian League: 1 quinarius (Syd. 698 - 89/2 B.C.; intrusion?)

Corone: 6 quinaries

Information from N. Price; details of above on
sale in BM
Representations of a Resource

- RDFa: Machine-readable data mixed in HTML
Why Care?

Key issues for interoperability

- URL/URI as a primary key:
  Let other people reference your data, anywhere on Web (!!)
  - Opens up silos
  - Let others add semantic value
  - Key for data citation

- Machine-Readable Representations: Share data that can be processed
  - Enable cross-collection analysis
  - New interfaces, visualizations
Data portability with Web standards (Atom, JSON, KML) and more specialized standards (ArchaeoML, RDF, OAI-PMH)
Web Representations: Visualizations
- **Incremental, Contestable Approach**
  - “Good” semantics depend on priorities / research questions.

- **Distributed**
  - Open, portable data lets others add semantic value
  - Open for data integration with other data sources on Web
• **ArchaeoML**
  - Abstracted standard, retains original typologies / terminologies
  - Baseline integration in Open Context

• **Granularity**
  - All data records have unique URL/URIs
  - Reference/index according to any ontology, anywhere on Web

• **Web services**
  - Precision queries to find resources meeting analytically meaningful criteria
Atom: Standard Container
Atom: Standard Container

URL/URI's of Resources
Feeds as Meaningful Lists

- Results of precision queries (with Boolean expression) expressed as a **Web resource**
- Can get sets of resources defined by precise analytic criteria
Feeds as Meaningful Lists

Open Context
Web-based research data publication

Open Context currently has 30888 items, filtered by the following criteria:
- Contained in: Jordan OR Turkey OR Israel
- Category: Animal Bone

Results: items 1 to 10 out of 30888 items

- Domuztepe Excavations
- Animal Bone: Domuztepe 19623
  - User Created Tags: Domuztepe dogs

- Petra Great Temple Excavations
- Animal Bone: Bone GT-0825
  - Context: Jordan / Petra Great Temple / Lower Transition / Special Project 65 / Louvre I / Seq. PF05129

- Petra Great Temple Excavations
- Animal Bone: Bone GT-3432
  - Context: Jordan / Petra Great Temple / Upper Transition / Special Project 64 / Louvre I / Seq. SP04037

- Domuztepe Excavations
- Animal Bone: DT04-4851
  - Context: Turkey / Domuztepe 1 / Louvre I / Seq. 30020
Feeds as Meaningful Lists
Feeds as Meaningful Lists
Feeds as Meaningful Lists

You can add your own layers of meaning (semantic models)
You can probably do better things with these data than we can!
Plain or Semantic Web

Linked Data Platform (LDP) Working Group Charter

The mission of the Linked Data Platform (LDP) Working Group is to produce a W3C Recommendation for HTTP-based (RESTful) application integration patterns using read/write Linked Data. This work will benefit both small-scale in-browser applications (WebApps) and large-scale Enterprise Application Integration (EAI) efforts. It will complement SPARQL and will be compatible with publishing Linked Data, bringing the data integration features of RDF into software development.


<table>
<thead>
<tr>
<th>End date</th>
<th>1 June, 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidentiality</td>
<td>Proceedings are public</td>
</tr>
<tr>
<td>Chairs</td>
<td>Arnaud Le Hors, IBM, Erik Wilde, EMC</td>
</tr>
<tr>
<td>Staff Contact</td>
<td>Eric Prud'hommeaux 0.15 FTE, Yves Lafon 0.15 FTE</td>
</tr>
<tr>
<td>Teleconference Schedule</td>
<td>One 60-90 minute call per week, plus tasks</td>
</tr>
<tr>
<td>Face-to-Face Meetings</td>
<td>Three expected during the course of the group, needed to help the group reach its goals. These meetings will use teleconferencing facilities, but effective participation generally requires attending in person, so participants should budget for travel.</td>
</tr>
</tbody>
</table>

Formats:
- **RDF**: Data model (graph), expressed as RDF/XML (pain), N3, Turtle. Preferred for Linked Data, but not easy to use by mortals
- **JSON**: Easy for mashups, but informal (no namespaces)
- **Mixed**: Pragmatic approach. Make data available in a variety of formats good for different consumers / applications, don't forget **CSV**!
Some Good Starting Points

- **HTTP URIs**: Good way to identify things (can also add persistent ids like DOIs)
- **REST(ish)**: REST can become a religious debate. But making resources (esp. structured data) available by following links is a good rule of thumb.
- **Link to other data**: Good practice, and it helps build collaborative ties, trust

---

**Noah Webster**

*Image Credit: “Nostri-Imago” via Flickr (CC-By)*

http://www.flickr.com/photos/nostri-imago/3429417174
Conclusions

“How I learned to stop worrying and love the Web”