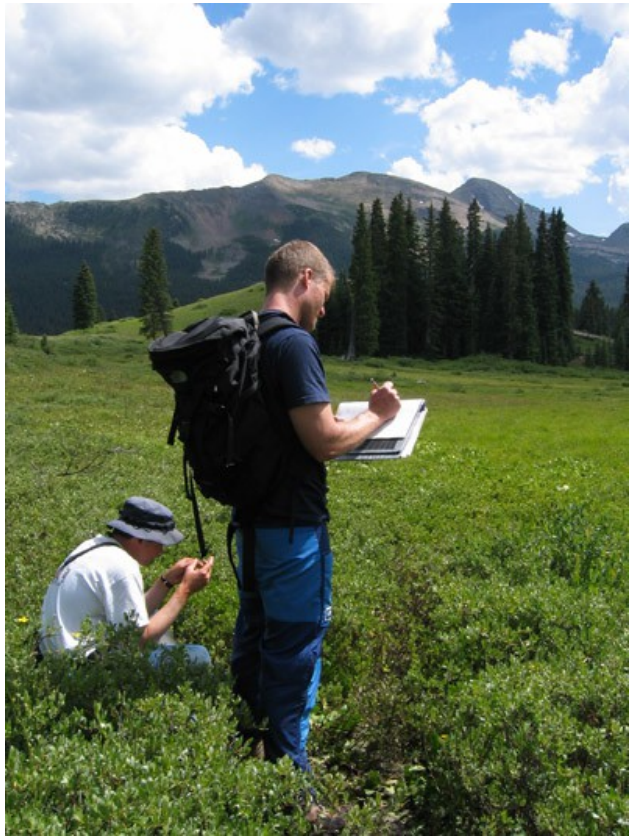
Interoperability, Integration, and Distributed Data

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



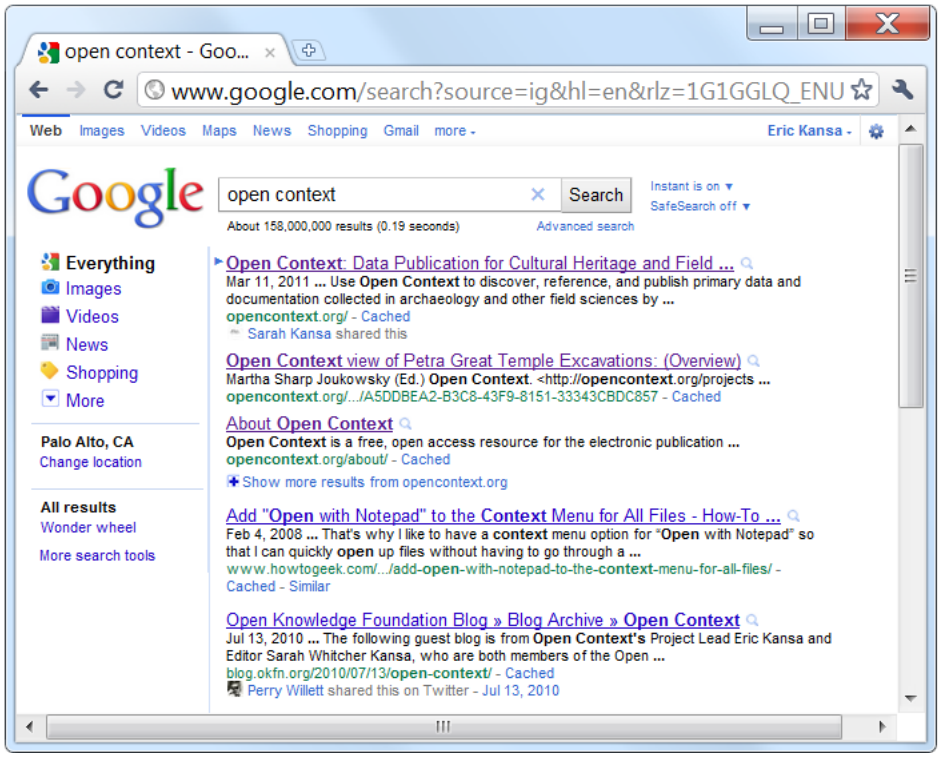
Key Concepts for the Web

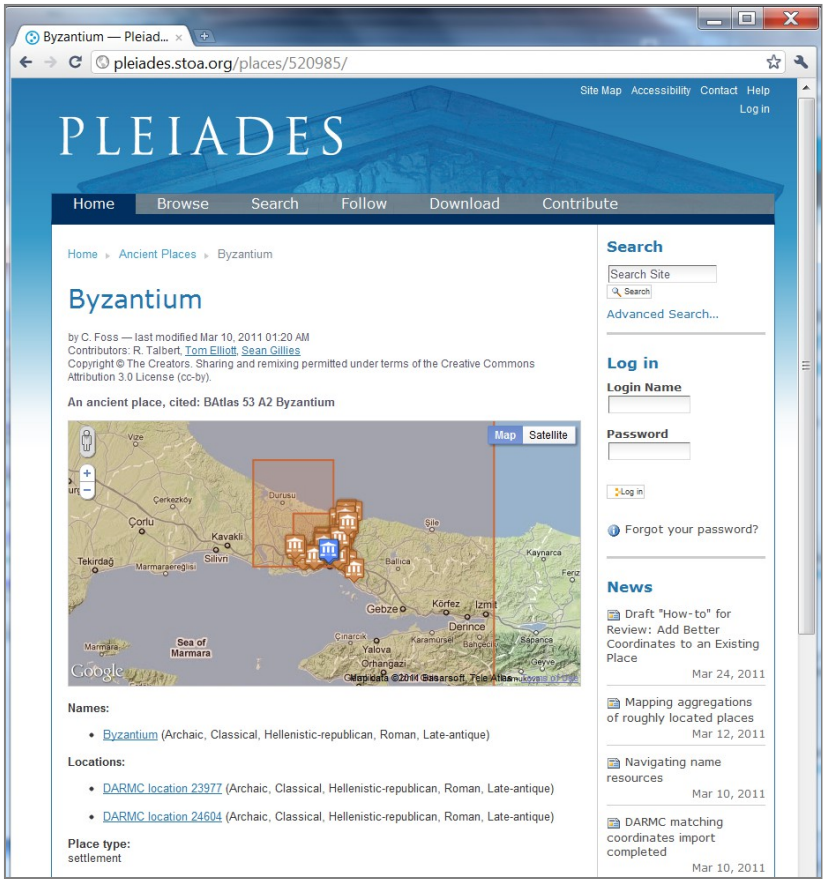
Key issues for interoperability

- Resources: Anything that can be indentified 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

Example in discussion

Web Resources

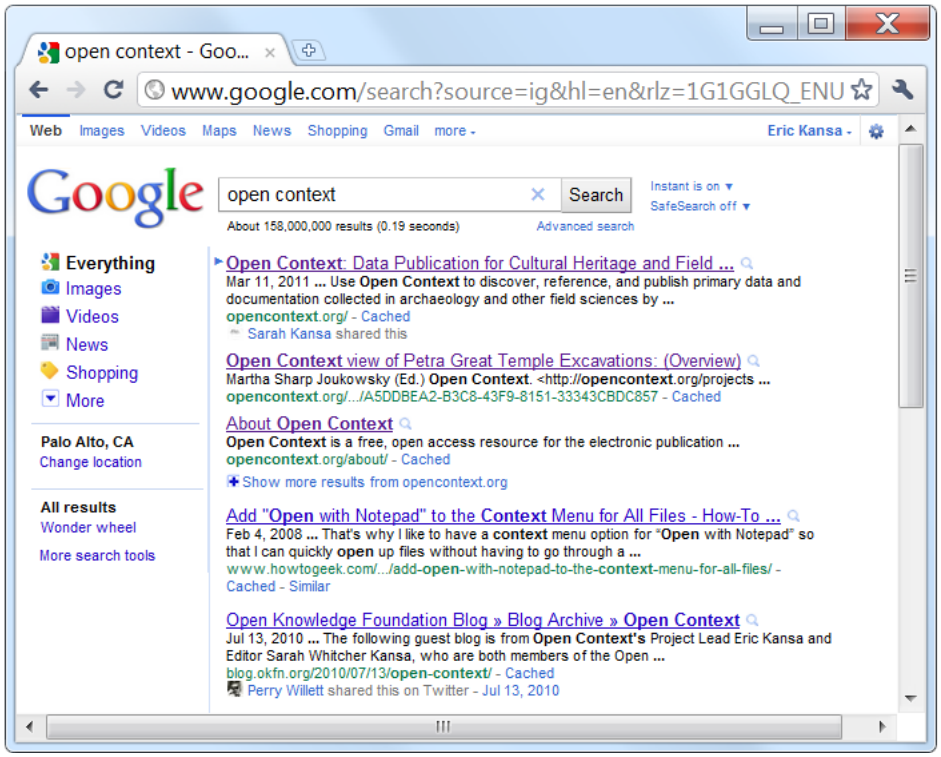


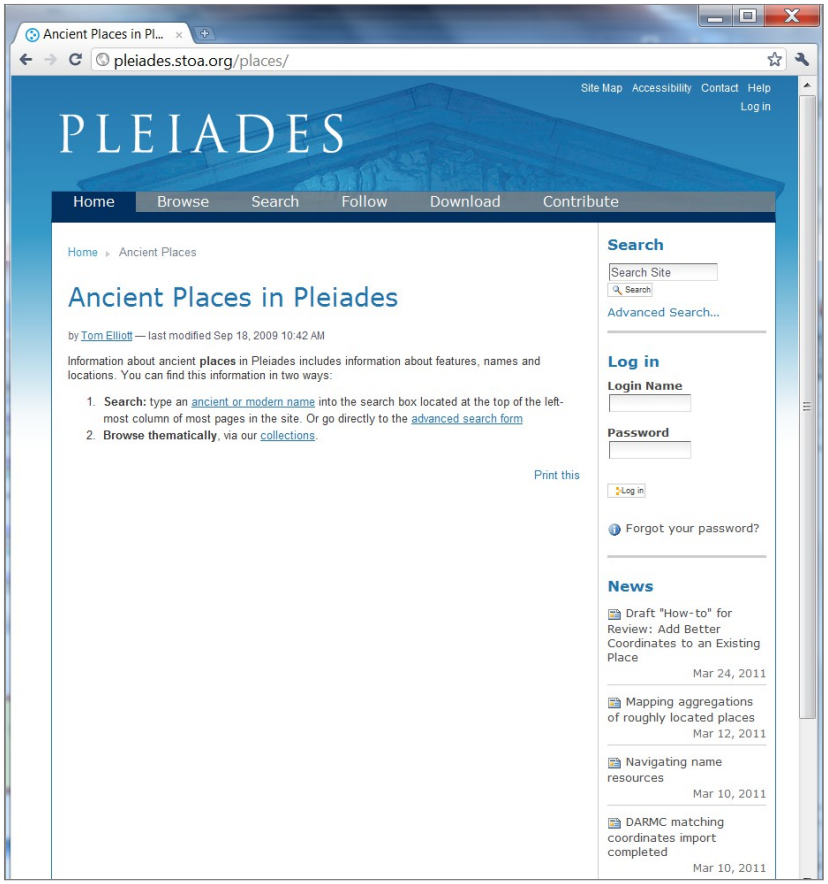


Dynamic service / query results (Google)

Abstract concept (“place”) in Pleiades

Web Resources

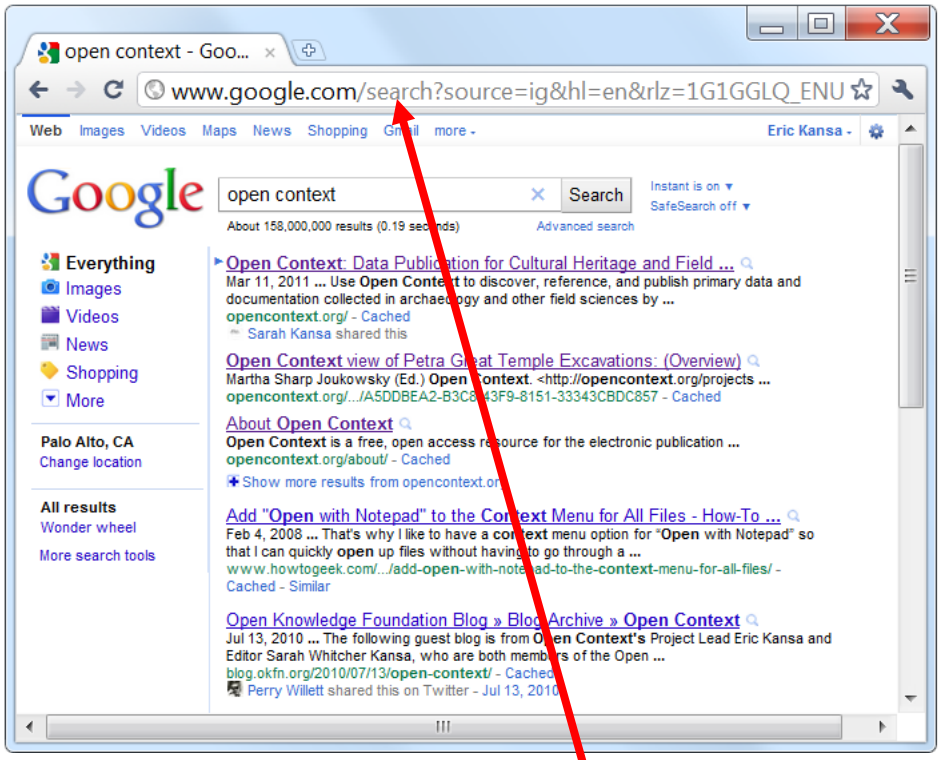


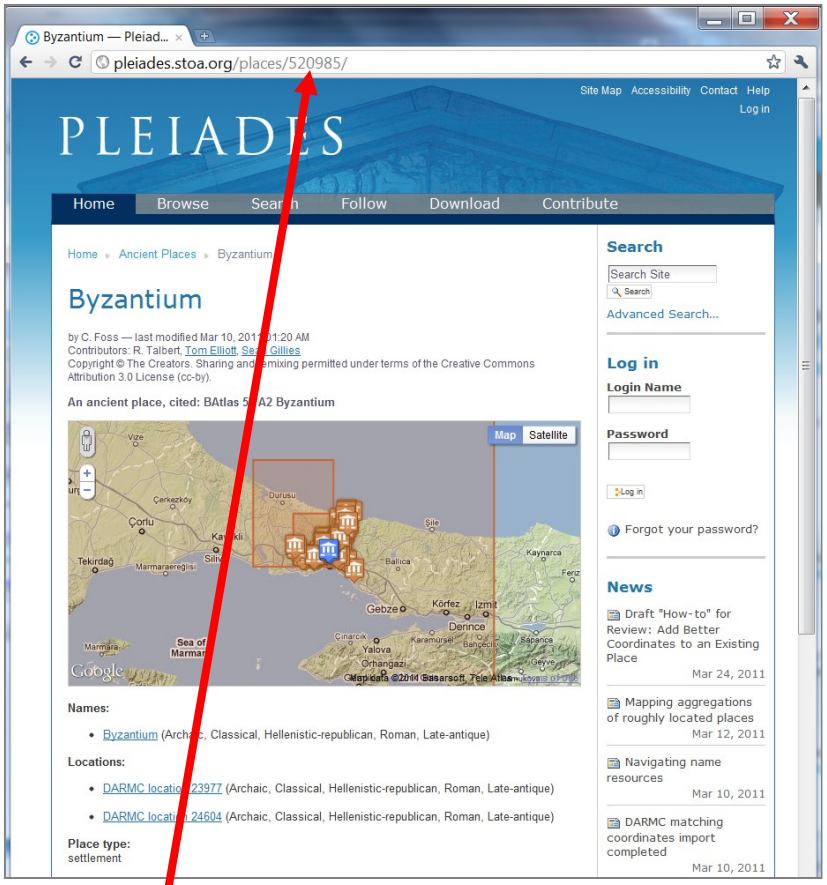


Dynamic service / query results (Google)

Abstract concept ("place" in general) in Pleiades

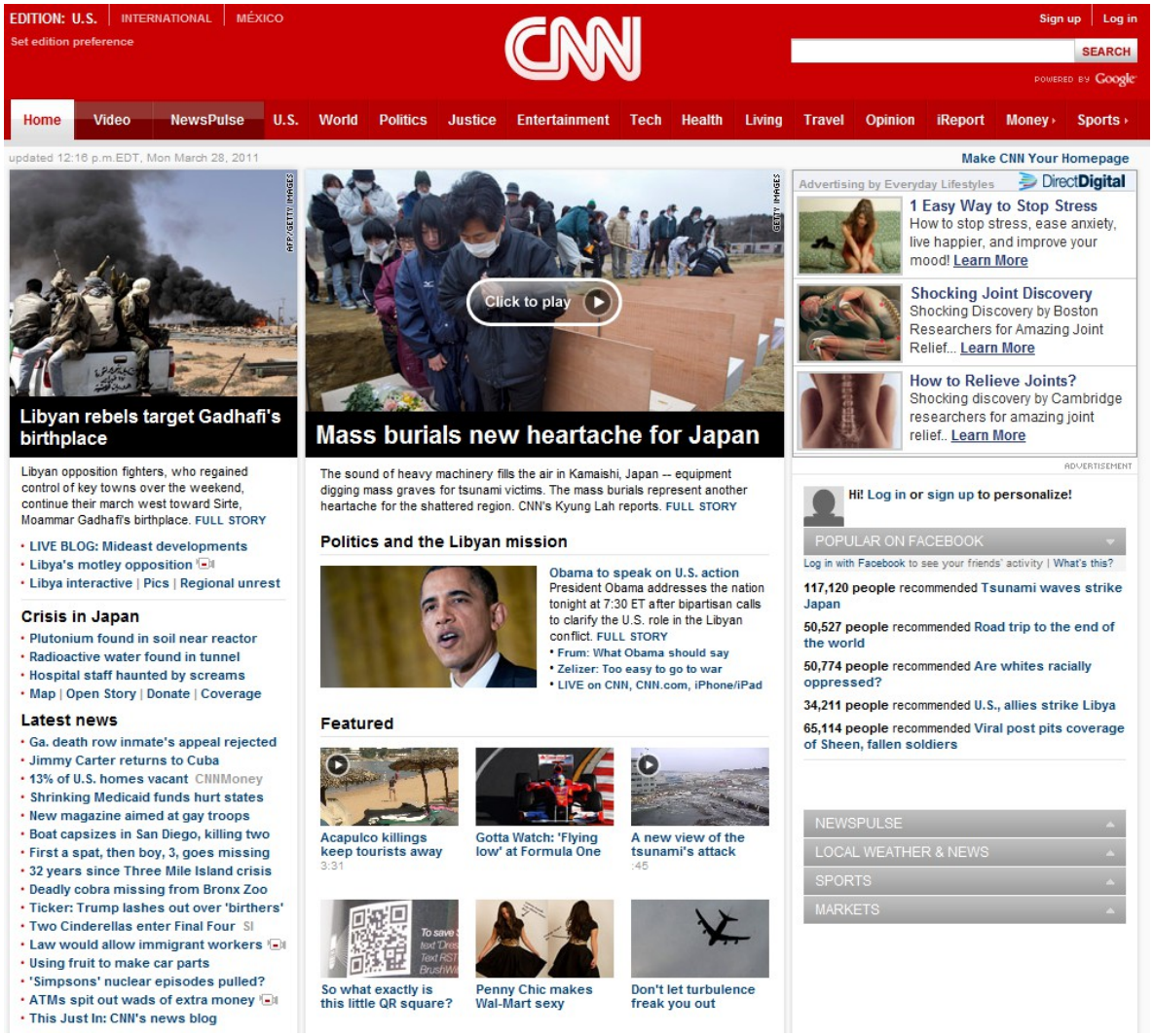
Web Resources



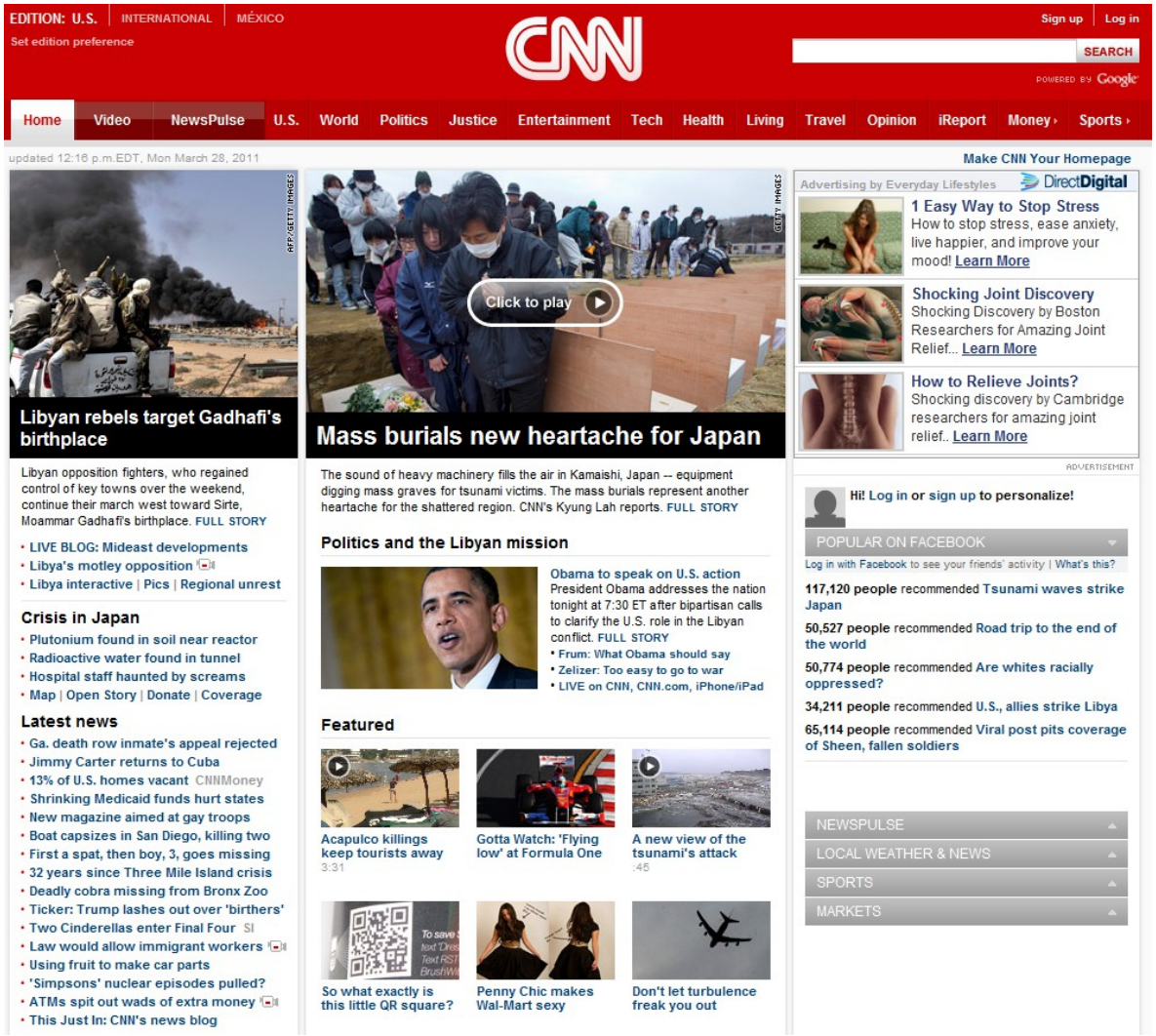


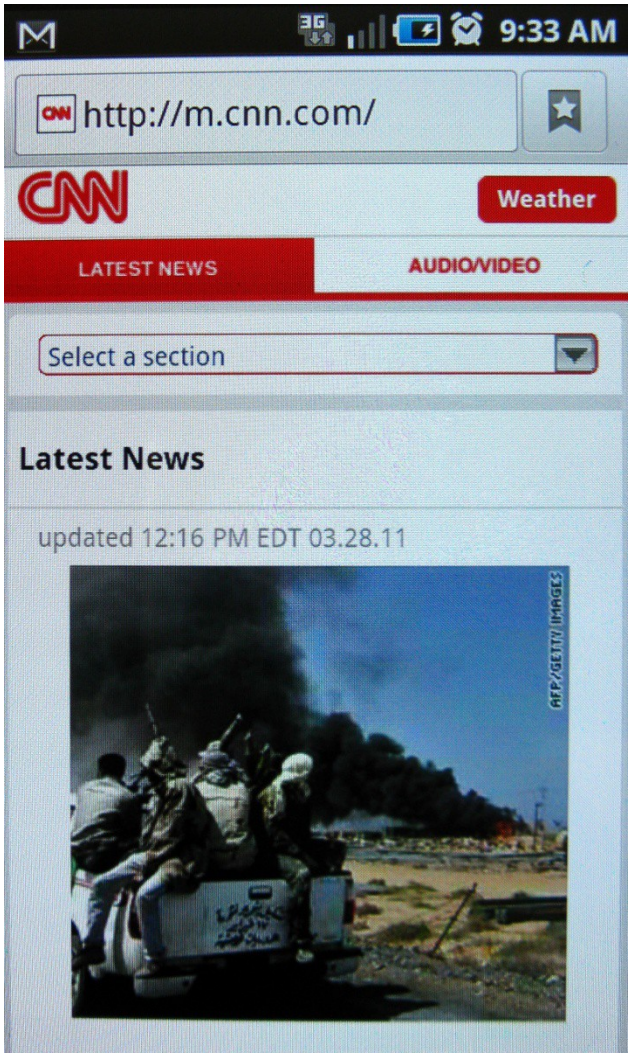
Key Common Requirement:
URL/URI

Representations

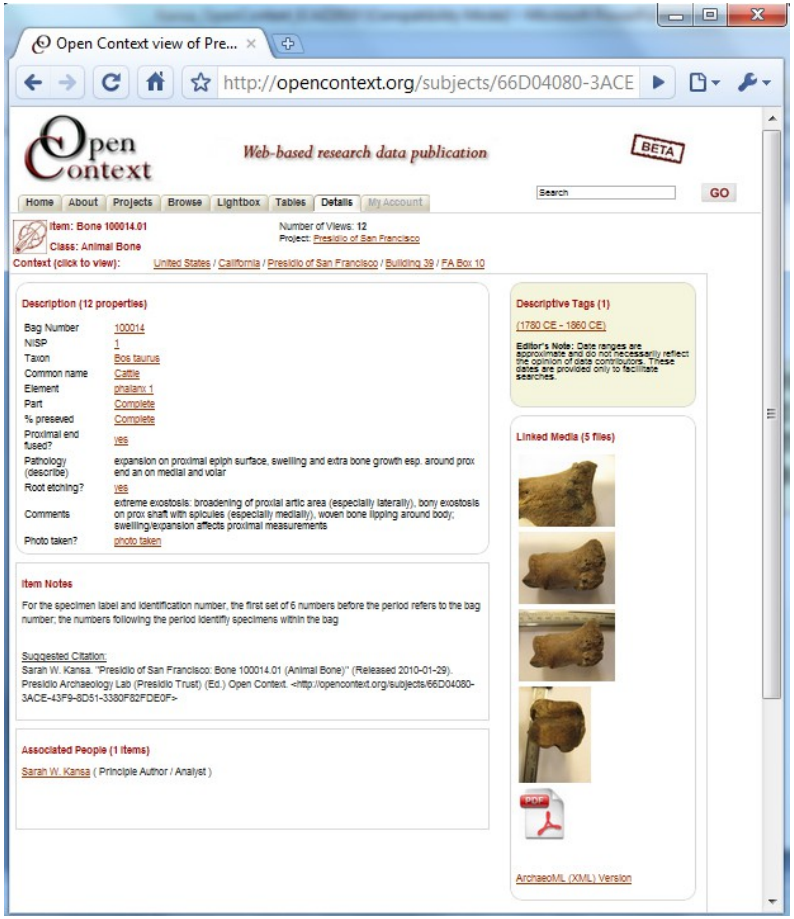


Representations





Representations of a Resource



This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
− <arch:spatialUnit UUID="D5D2BD55-55A2-445F-9258-37A36AAC343E" ownedBy="C5B4F73B-5EF8-4099-590E-B0275EDBA2A7">
  − <arch:name>
      <arch:string>Bone 100013.04</arch:string>
    </arch:name>
  − <oc:item_class>
      <oc:name>Animal Bone</oc:name>
    − <oc:iconURI>
        http://www.opencontext.org/database/ui_images/med_oc_icons/animal_bone_50x50.jpg
      </oc:iconURI>
    </oc:item_class>
  − <arch:observations>
    − <arch:observation obsNumber="1">
      − <arch:properties>
        − <arch:property>
            <arch:variableID>7764853A-0A18-41DB-B663-8E8685CB2F1C</arch:variableID>
            <arch:integer>100013</arch:integer>
            <oc:propid href="http://opencontext.org/properties/79283804-5075-411B-757A-1CE33FE0560A">79283804-5075-411B-757A-1CE33FE0560A</oc:propid>
            <oc:var_label type="integer">Bag Number</oc:var_label>
            <oc:show_val>100013</oc:show_val>
          </arch:property>
        − <arch:property>
            <arch:variableID>37D59894-C1D4-41B4-C88B-DFC5A0C3FB6E</arch:variableID>
            <arch:integer>1</arch:integer>
            <oc:propid href="http://opencontext.org/properties/B41B61DA-0CD6-4D87-EE41-E821C4BDC753">B41B61DA-0CD6-4D87-EE41-E821C4BDC753</oc:propid>
            <oc:var_label type="integer">NISP</oc:var_label>
            <oc:show_val>1</oc:show_val>
          </arch:property>
        − <arch:property>
            <arch:variableID>A9172FBF-9AEF-4F7E-0BD5-DC23EC72F541</arch:variableID>
            <arch:valueID>9CE2D1A6-AD67-46CF-61A3-341CDA5ACED1</arch:valueID>
            <oc:propid href="http://opencontext.org/properties/BA9534BD-FC23-4037-954D-F415F732B618">BA9534BD-FC23-4037-954D-F415F732B618</oc:propid>
            <oc:var_label type="nominal">Taxon</oc:var_label>
            <oc:show_val>Bos taurus</oc:show_val>
          </arch:property>
        − <arch:property>
            <arch:variableID>262078F0-8B76-4B5B-9BEC-F74D191E4EF7</arch:variableID>
            <arch:valueID>D73A48C8-C0F6-427C-1615-159FDD200407</arch:valueID>
            <oc:propid href="http://opencontext.org/properties/5E764397-3125-4552-046F-A084E0219A7A">5E764397-3125-4552-046F-A084E0219A7A</oc:propid>
            <oc:var_label type="nominal">Common name</oc:var_label>
```


- Browser (XHTML) and “Machine-readable” (ArchaeoML-XML) representations

Representations of a Resource

Nomisma.org: [home] [all mints] [all hoards] [all authorities] [all ids] "Common currency for digital numismatics."

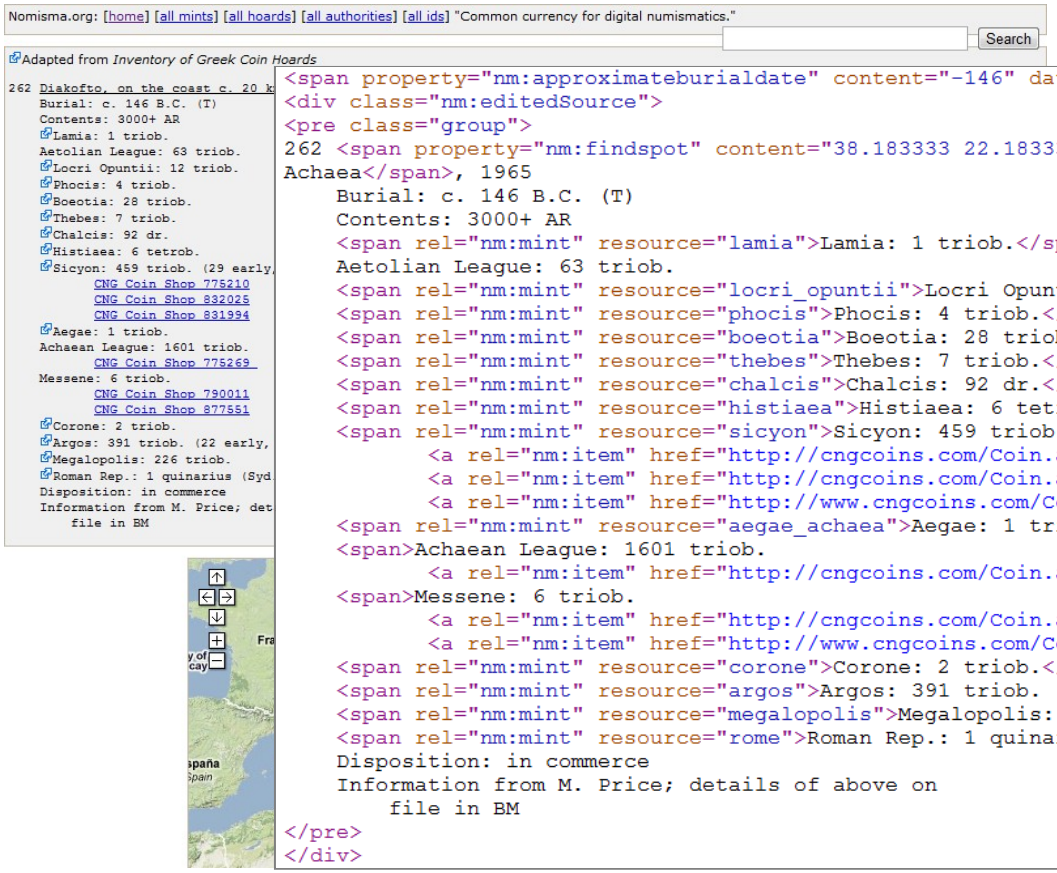
Search

Adapted from *Inventory of Greek Coin Hoards*

262 Diakofto, on the coast c. 20 km. E of Aegium, Achaea, 1965
Burial: c. 146 B.C. (T)
Contents: 3000+ AR
Lamia: 1 triob.
Aetolian League: 63 triob.
Locri Opuntii: 12 triob.
Phocis: 4 triob.
Boeotia: 28 triob.
Thebes: 7 triob.
Chalcis: 92 dr.
Histiaea: 6 tetrob.
Sicyon: 459 triob. (29 early, 430 late)
CNG Coin Shop 775210
CNG Coin Shop 832025
CNG Coin Shop 831994
Aegae: 1 triob.
Achaean League: 1601 triob.
CNG Coin Shop 775269
Messene: 6 triob.
CNG Coin Shop 790011
CNG Coin Shop 877551
Corone: 2 triob.
Argos: 391 triob. (22 early, 369 late)
Megalopolis: 226 triob.
Roman Rep.: 1 quinarius (Syd. 609a = 83/2 B.C.; intrusion?)
Disposition: in commerce
Information from M. Price; details of above on file in BM

- Nomisma.org (coin hordes)

Representations of a Resource




```
<span property="nm:approximateburialdate" content="-146" datatype="xsd:date"/>
<div class="nm:editedSource">
<pre class="group">
262 <span property="nm:findspot" content="38.183333 22.183333">Diakofto, on the coast c. 20 km. E of Aegium,
Achaea</span>, 1965
    Burial: c. 146 B.C. (T)
    Contents: 3000+ AR
    <span rel="nm:mint" resource="lamia">Lamia: 1 triob.</span>
    Aetolian League: 63 triob.
    <span rel="nm:mint" resource="locri_opuntii">Locri Opuntii: 12 triob.</span>
    <span rel="nm:mint" resource="phocis">Phocis: 4 triob.</span>
    <span rel="nm:mint" resource="boeotia">Boeotia: 28 triob.</span>
    <span rel="nm:mint" resource="thebes">Thebes: 7 triob.</span>
    <span rel="nm:mint" resource="chalcis">Chalcis: 92 dr.</span>
    <span rel="nm:mint" resource="histiaea">Histiaea: 6 tetrob.</span>
    <span rel="nm:mint" resource="sicyon">Sicyon: 459 triob. (29 early, 430 late)
           <a rel="nm:item" href="http://cngcoins.com/Coin.aspx?CoinID=89134">CNG Coin Shop 775210</a>
           <a rel="nm:item" href="http://cngcoins.com/Coin.aspx?CoinID=148616">CNG Coin Shop 832025</a>
           <a rel="nm:item" href="http://www.cngcoins.com/Coin.aspx?CoinID=148615">CNG Coin Shop 831994</a></span>
    <span rel="nm:mint" resource="aegae_achaea">Aegae: 1 triob.</span>
    <span>Achaean League: 1601 triob.
           <a rel="nm:item" href="http://cngcoins.com/Coin.aspx?CoinID=89154">CNG Coin Shop 775269 </a></span>
    <span>Messene: 6 triob.
           <a rel="nm:item" href="http://cngcoins.com/Coin.aspx?CoinID=105434">CNG Coin Shop 790011</a>
           <a rel="nm:item" href="http://www.cngcoins.com/Coin.aspx?CoinID=168066">CNG Coin Shop 877551</a></span>
    <span rel="nm:mint" resource="corone">Corone: 2 triob.</span>
    <span rel="nm:mint" resource="argos">Argos: 391 triob. (22 early, 369 late)</span>
    <span rel="nm:mint" resource="megalopolis">Megalopolis: 226 triob.</span>
    <span rel="nm:mint" resource="rome">Roman Rep.: 1 quinarius (Syd. 609a = 83/2 B.C.; intrusion?)</span>
    Disposition: in commerce
    Information from M. Price; details of above on
        file in BM
</pre>
</div>
```

- RDFa: Machine-readable data mixed in HTML

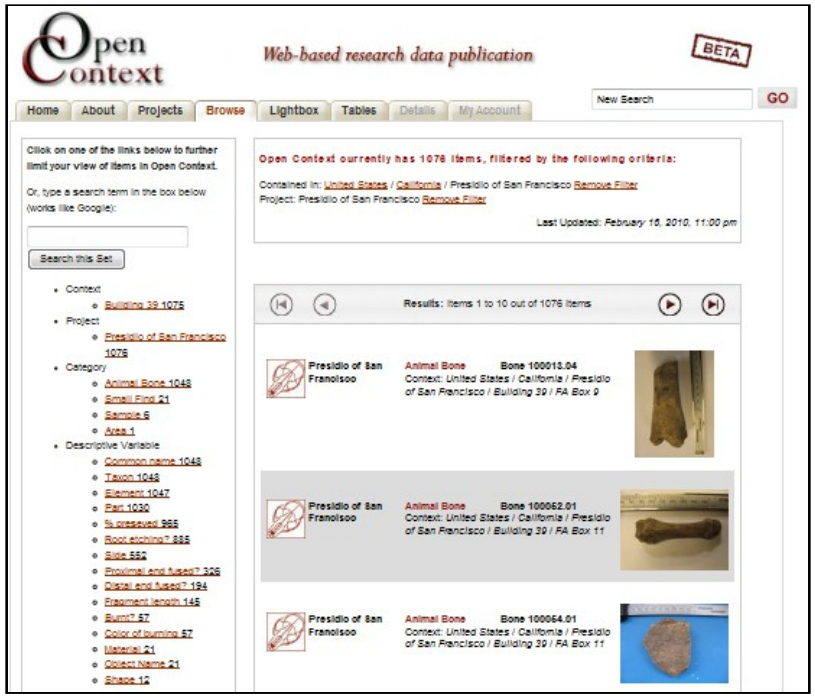
Why Care?

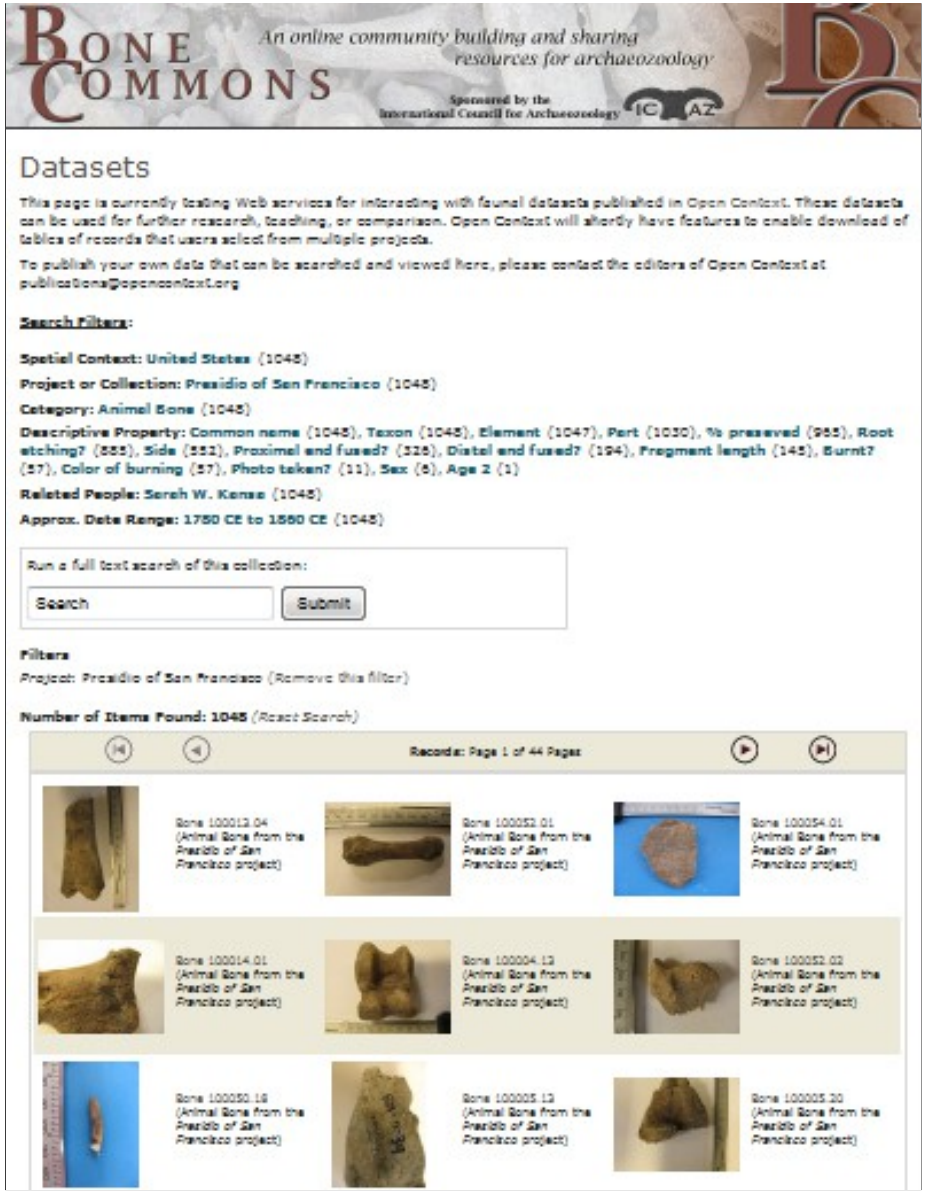
Key issues for interoperability

- 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

Open Context

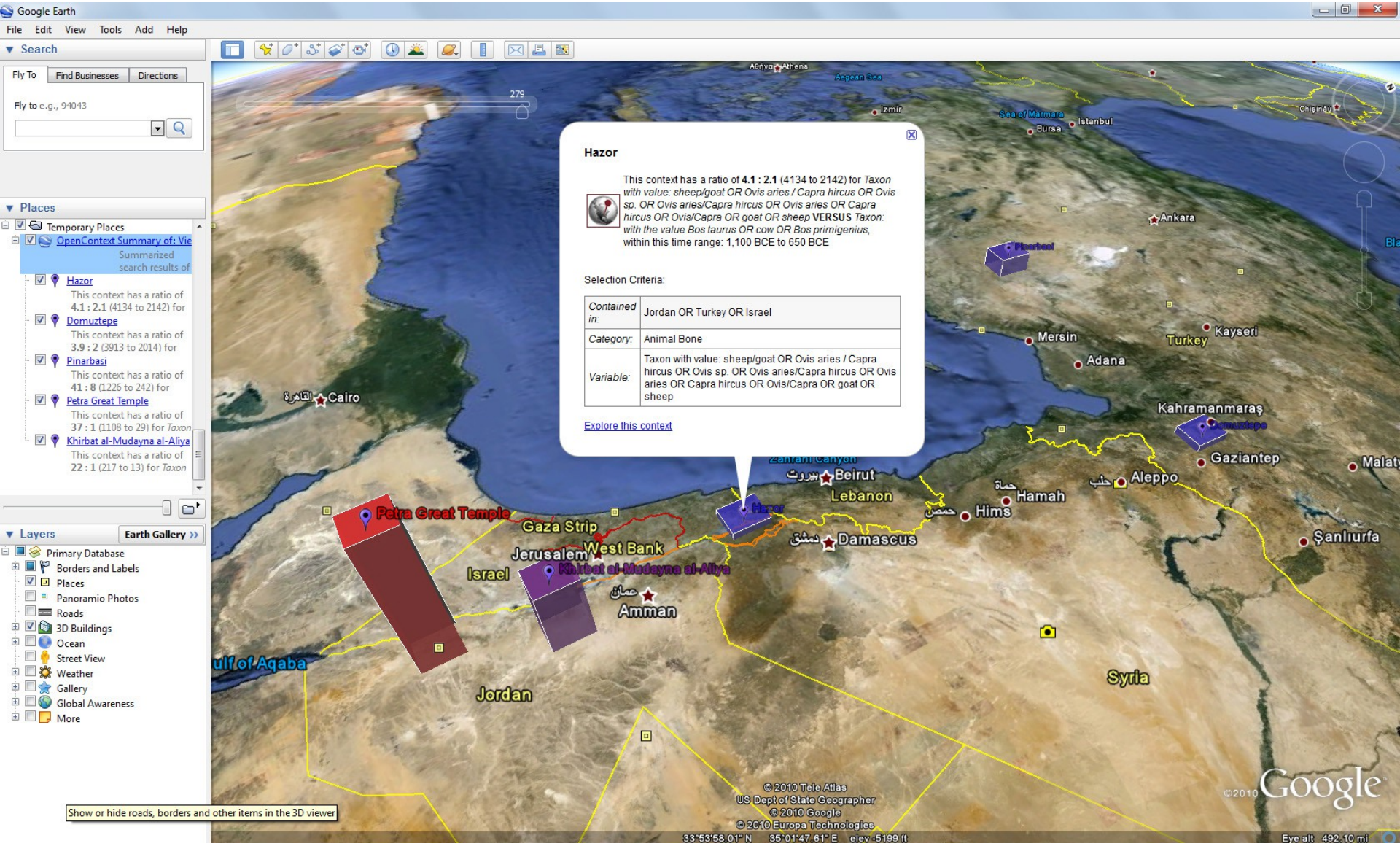
Working toward these goals

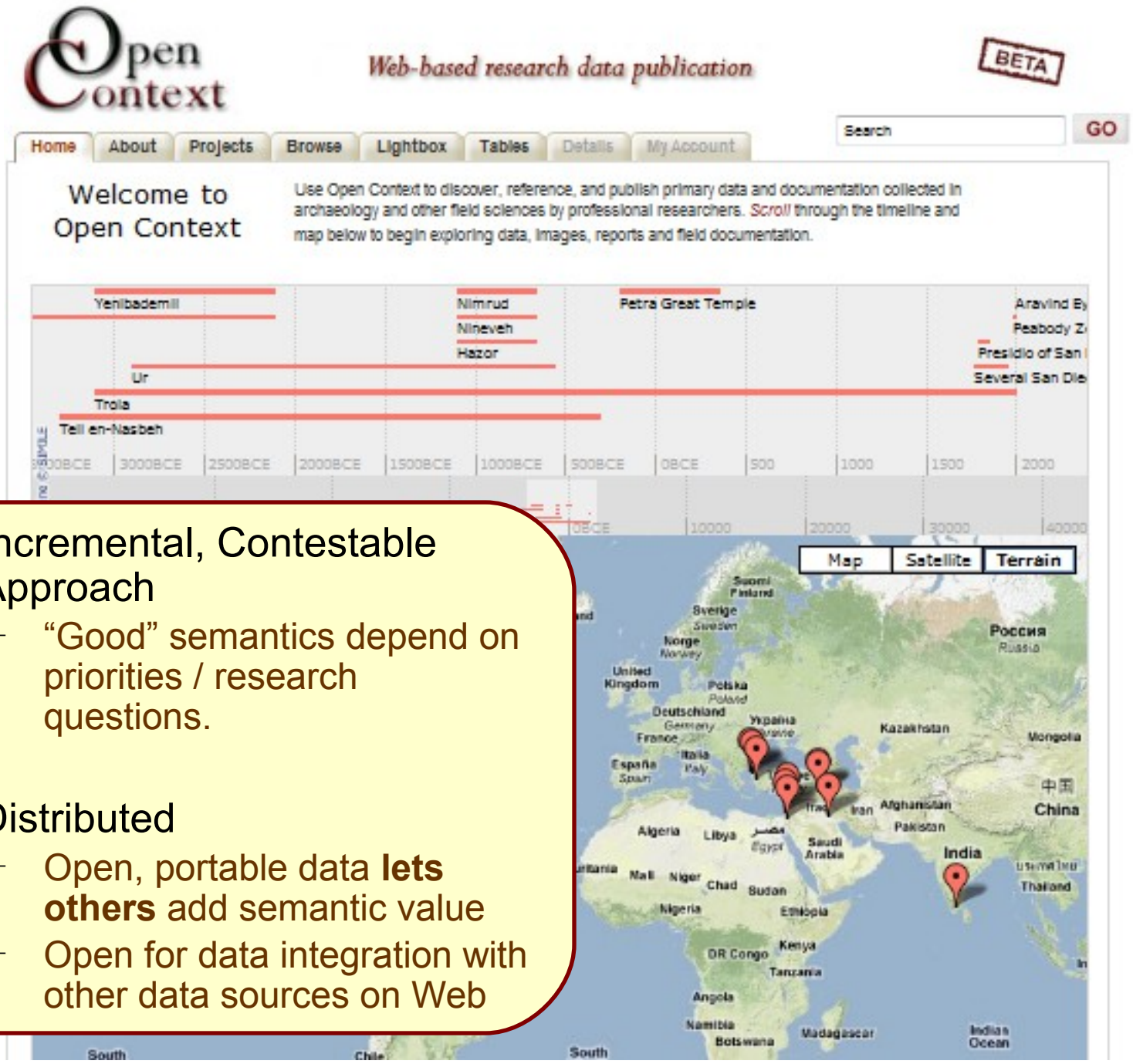




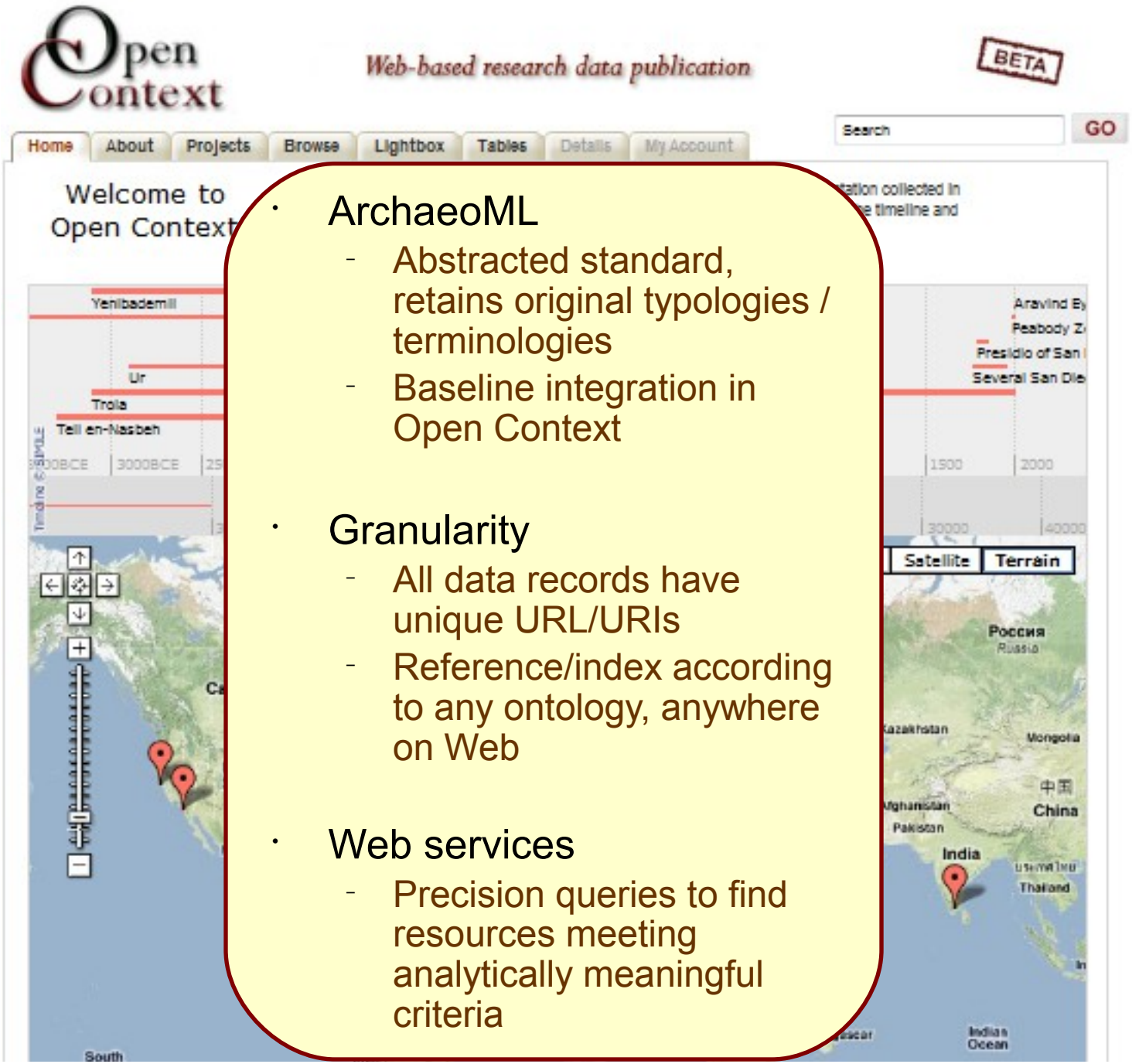
Data portability with Web standards (Atom, JSON, KML) and more specialized standards (ArchaeoML, RDF, OAI-PMH)

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

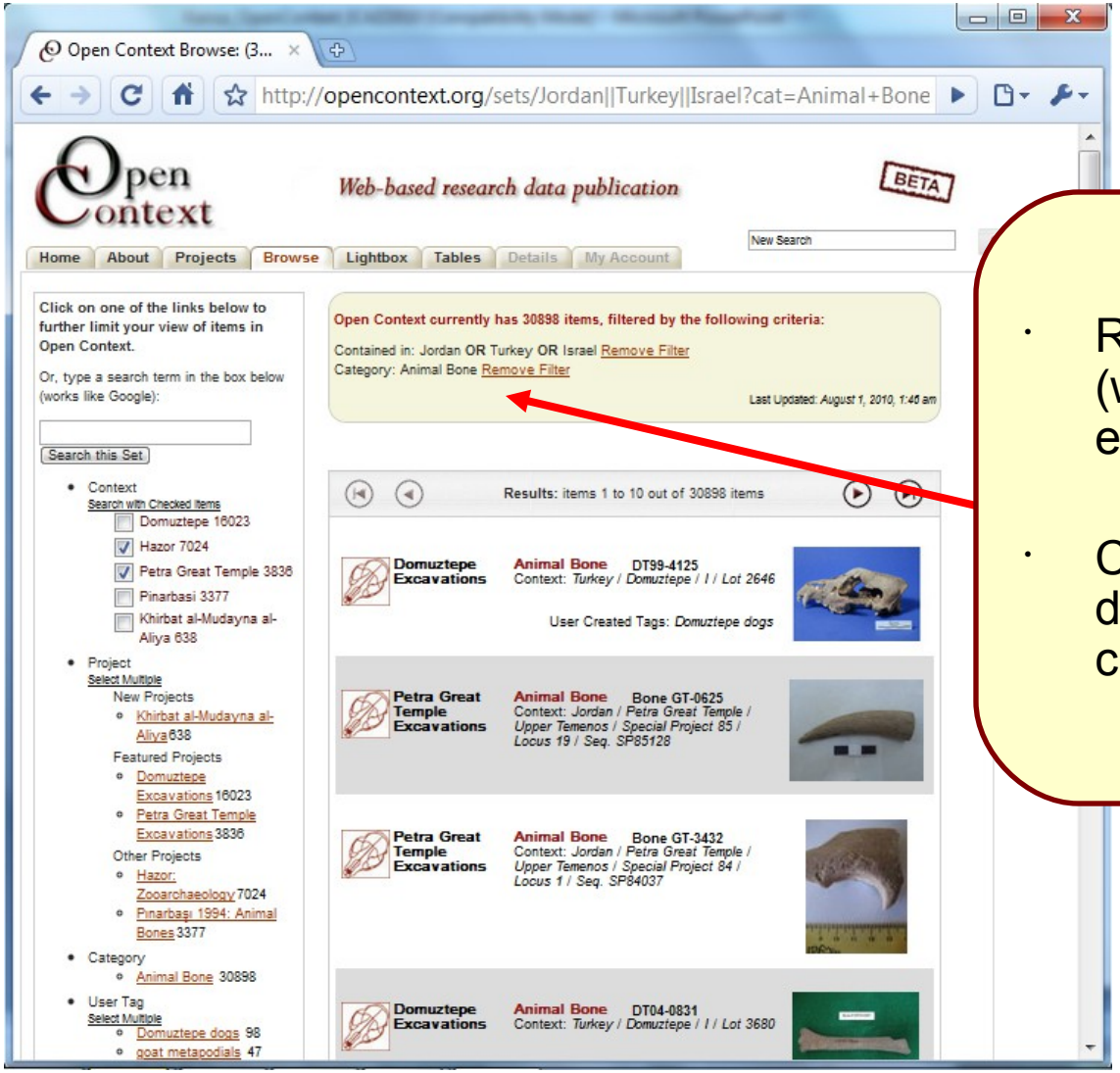


- ArchaeoML
  - Abstracted standard, retains original typologies / terminologies
  - Baseline integration in Open Context
- Granularity
  - All data records have unique URL/URIs
  - Reference/index according to any ontology, anywhere on Web
- Web services
  - Precision queries to find resources meeting analytically meaningful criteria

Atom: Standard Container

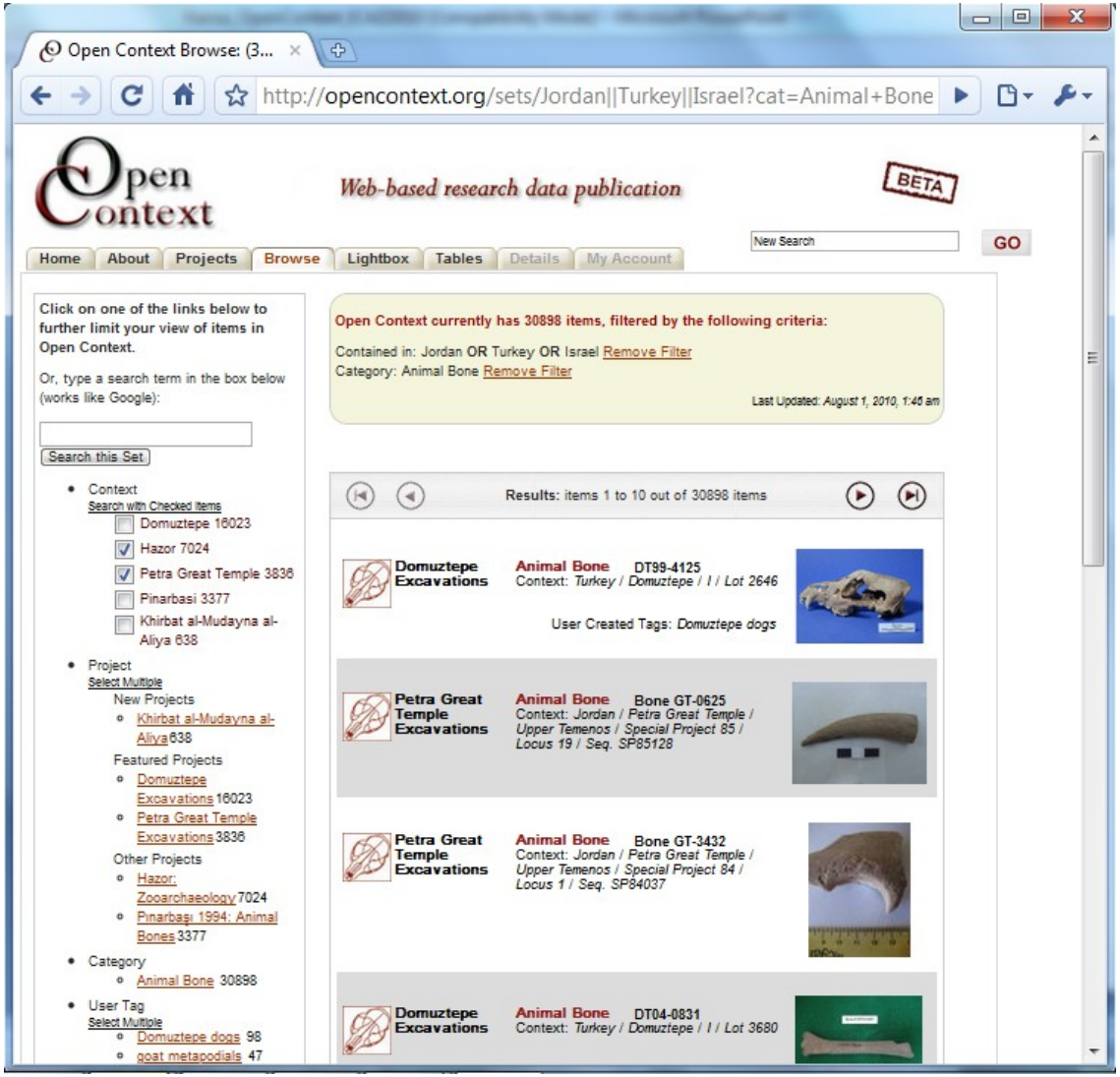
Atom: Standard Container

Feeds as Meaningful Lists

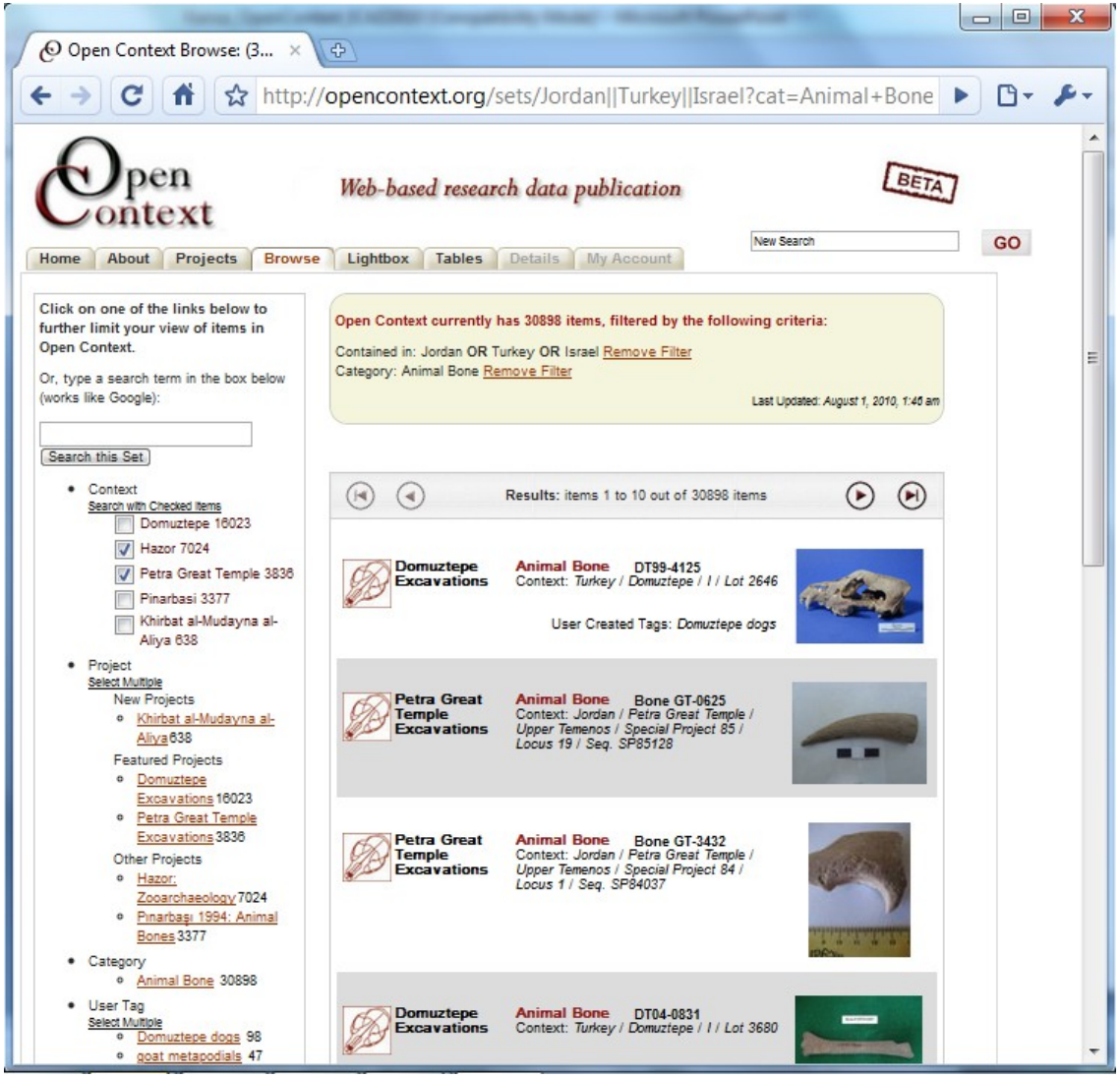


- Results of precision queries (with Boolean expression) expressed as a **Web resource**
- Can get sets of resources defined by precise analytic criteria

Feeds as Meaningful Lists



Feeds as Meaningful Lists

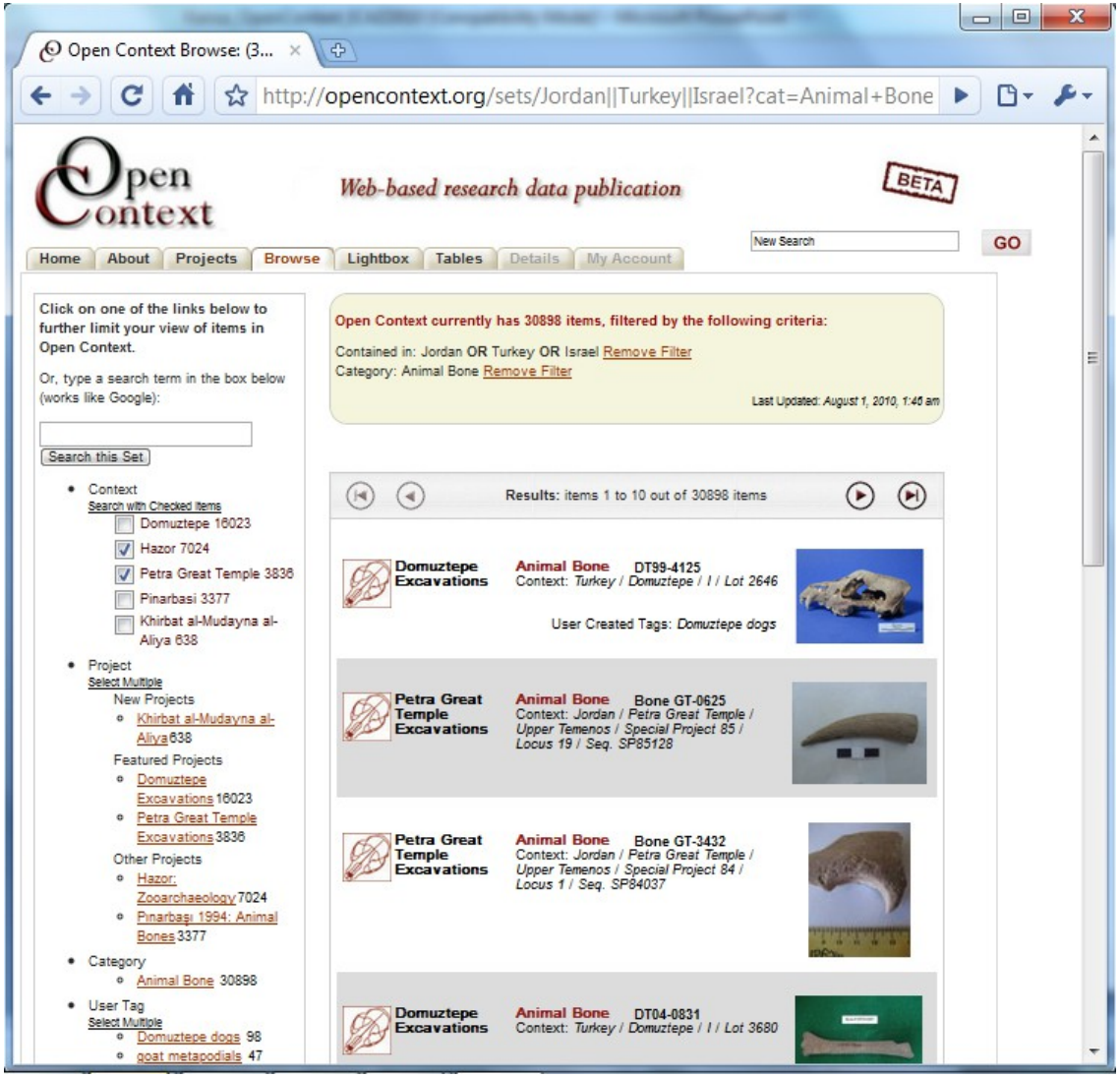


University of California

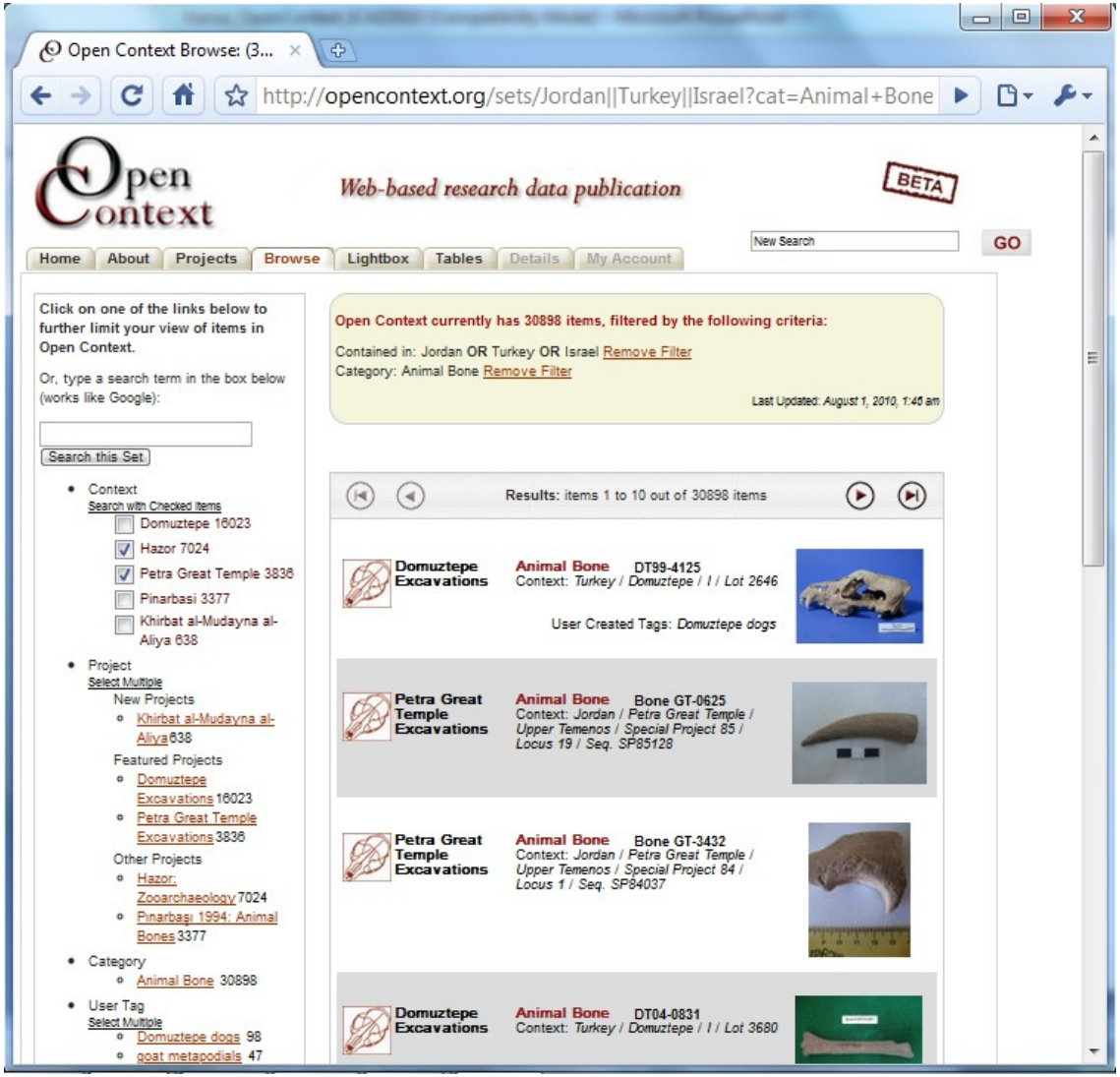
CDL

California Digital Library

Feeds as Meaningful Lists

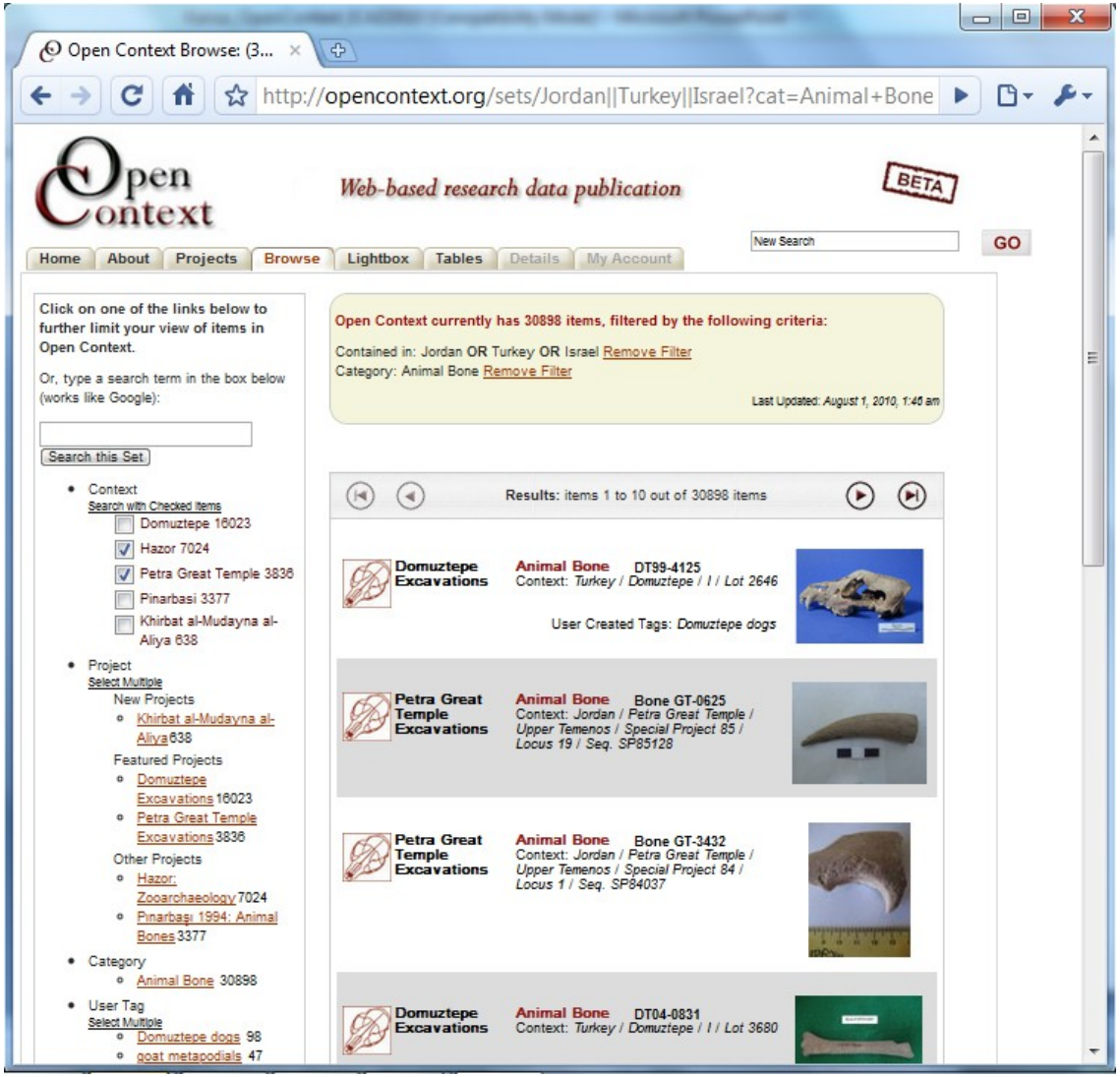


Feeds as Meaningful Lists

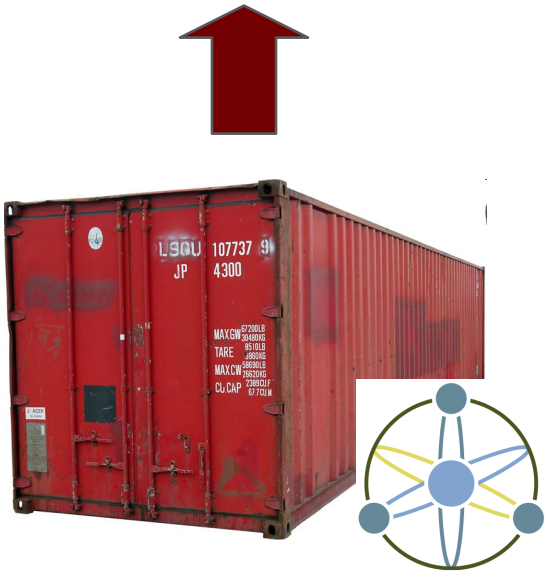


You can add your own layers of meaning (semantic models)

Feeds as Meaningful Lists



You can probably do better things with these data than we can!



Plain or Semantic Web

W3C Technology and Society domain

- Background
- Scope
- Deliverables
- Dependencies and Liaisons
- Participation
- Communication
- Decision Policy
- Patent Policy
- About this Charter

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 compat
publishing Linked Data, bringing the data integration features of RD
software development.

Join the Linked Data Platform (LDP) Working Group.

End date	1 June, 2014
Confidentiality	Proceedings are public
Chairs	Arnaud Le Hors, IBM Erik Wilde, EMC
Staff Contact	Eric Prud'hommeaux 0.15 FTE Yves Lafon 0.15 FTE
Teleconference Schedule	One 60-90 minute call per week, plus task
Face-to-Face Meetings	Three expected during the course of the group, as 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.

Formats:

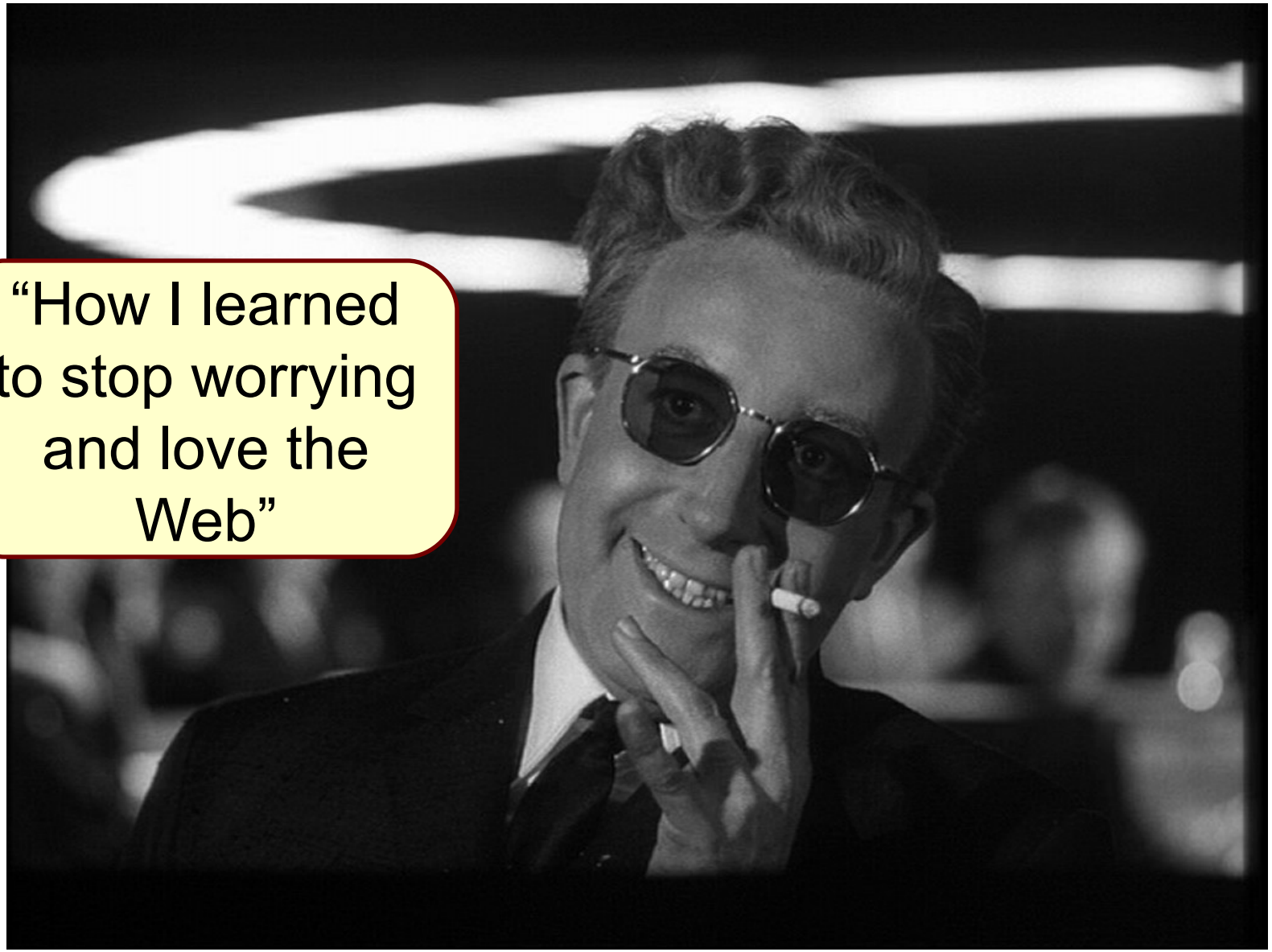
- **RDF:** Data model (graph), expressed as RDF/XML (pain), N3, Turtle. Preffered for Linked Data, but not easy to use by mortals
- **JSON**: Easy for mashups, but informal (no namespaces)
- **Mixed**: Pragmatic approach. Make data available in a variety of formats good for different consumers / applicatins,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

Pragmatism

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) availble 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