Linked Data, Publication, and the Life Cycle of Archaeological Information

Eric C. Kansa (@ekansa)
UC Berkeley D-Lab
& Open Context

2014-2015 Harvard Center for Hellenic Studies & German Archaeological Institute Research Fellow

Unless otherwise indicated, this work is licensed under a Creative Commons Attribution 3.0 License <http://creativecommons.org/licenses/by/3.0/>
Welcome to Open Context
Open Context reviews, edits, and publishes archaeological research data and archives data with university-backed repositories, including the California Digital Library.

Data Sharing as Publication
• Started in 2007
• Open data (mainly CC-By)
• Archiving by California Digital Library
• Part of a broader reform movement in scholarly communications
Reforming scholarly communications

1. Why Linked Data?
2. Challenges of (Linked) Data in academic settings
3. Case studies in integrating Linked Data in research communication
Reforming scholarly communications

1. Why Linked Data?
2. Challenges of (Linked) Data in academic settings
3. Case studies in integrating Linked Data in research communication
Web of Data (2011)

Need Archaeology on the Map
Contributions should not be isolated from other communities
Bootstrapping Problem

- Benefits mainly theoretical until you get lots of Linked Data
- Need research community with skill sets to use Linked Data
Metcalfe's Law

\[ Y = X^*(X-1) \]

**Critical Mass**

Value

Number of Nodes
Stable Web URI:
Reference this to disambiguate between “Alexandria” (Egypt) and other places called “Alexandria” (many of which are also ancient)
Pelagios:
Heat map of museum collections, archives, databases referencing places in Pleiades
(PIs Leif Isaksen, Elton Barker)
Perio.do

- Gazetteer of assertions about “periods” (place-time entities)
- Not a controlled vocabulary, but Linked Data friendly way to author and reference scholarly assertions about periods
- NEH funded (PIs Adam Rabinowitz, Ryan Shaw, Eric Kansa)

Periodization

Source

<table>
<thead>
<tr>
<th>Title</th>
<th>The Uruk countryside: the natural setting of urban societies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year published</td>
<td>1972</td>
</tr>
<tr>
<td>Creators</td>
<td>Adams, Robert McC. (Robert McCormick), 1926-</td>
</tr>
<tr>
<td>Contributors</td>
<td>Nissen, Hans Jörg.</td>
</tr>
</tbody>
</table>

Periods

<table>
<thead>
<tr>
<th>Name</th>
<th>Start</th>
<th>Stop</th>
<th>Coverage</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jemdet Nasr</td>
<td>3100 B.C.</td>
<td>2900 B.C.</td>
<td>Iraq</td>
<td>See also Charvat 2002</td>
</tr>
<tr>
<td>Early Uruk</td>
<td>3500 B.C.</td>
<td>3200 B.C.</td>
<td>Iraq</td>
<td>See also Liverani 2006</td>
</tr>
<tr>
<td>Late Uruk</td>
<td>3200 B.C.</td>
<td>3000 B.C.</td>
<td>Iraq</td>
<td>See also Liverani 2006</td>
</tr>
</tbody>
</table>

Add period
Reforming scholarly communications

1. Why Linked Data?
2. Challenges of (Linked) Data in academic settings
3. Case studies in integrating Linked Data in research communication
Commercial interests and public policy

Neoliberalism:
(Loosely associated ideologies / assumptions / interests)

“Surely you were aware when you accepted the position, Professor, that it was publish or perish.”

Conditions of academic labor
Serials expenditures percentage increase over 1986

- serial expenditures %+
- consumer price index %+

Source: The Occasional Pamphlet - Harvard University
(http://blogs.law.harvard.edu/pamphlet/2013/01/29/why-open-access-is-better-for-scholarly-societies/)
OPEN ACCESS
Neoliberalism:  
(Loosely associated ideologies / assumptions / interests)
Neoliberalism: Taylorism, “Audit Culture” and fierce job/grant competition

↓

Data contributions don’t count!

Image Credit: Wikimedia Commons (Public Domain)
Ironies of data:
Publications counted as data, but data don’t count!
Contingent Employment

Tenure-Track Faculty Members Yield to Part-Timers
Shown are national totals for all degree-granting institutions for selected years.

<table>
<thead>
<tr>
<th>Year</th>
<th>Full-time tenured and tenure track</th>
<th>Part time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1975</td>
<td>56.8%</td>
<td>43.2%</td>
</tr>
<tr>
<td>1989</td>
<td>30.2%</td>
<td>69.8%</td>
</tr>
<tr>
<td>1995</td>
<td>46.8%</td>
<td>53.2%</td>
</tr>
<tr>
<td>2007</td>
<td>56.8%</td>
<td>43.2%</td>
</tr>
</tbody>
</table>

Source: U.S. Department of Education, IPEDS Fall Staff Survey. Compiled by the American Association of University Professors. By Ron Coddington

Expanding Public Access to the Results of Federally Funded Research

Posted by Michael Stebbins on February 22, 2013 at 12:04 PM EST

The Obama Administration is committed to the proposition that citizens deserve easy access to the results of scientific research their tax dollars have paid for. That's why, in a policy memorandum released today, OSTP Director John Holdren has directed Federal agencies with more than $100M in R&D expenditures to develop plans to make the published results of federally funded research freely available to the public within one year of publication and requiring researchers to better account for and manage the digital data resulting from federally funded scientific research. OSTP has been looking into this issue for some time, soliciting broad public input on multiple occasions and convening an interagency working group to develop a policy. The final policy reflects substantial inputs from scientists and scientific organizations, publishers, members of Congress, and other members of the public—over 65 thousand and of whom recently signed a We the People petition asking for expanded public access to the results of taxpayer-funded research.
Need more carrots!
1. Citation, credit, intellectually valued
2. Research outcomes (new insights from data reuse!)
Reforming scholarly communications

1. Why Linked Data?
2. Challenges of (Linked) Data in academic settings
3. Case studies in integrating Linked Data in research communication
1. Referenced by US National Science Foundation and National Endowment for the Humanities for Data Management
2. “Data sharing as publishing” metaphor
Need to consider the wider research community (inside & outside universities)
Digital Index of North American Archaeology (DINAA)

1. ~500,000 site records curated by state officials
2. Key (Linked Data!) reference for N. American archaeology
3. PIs/Co-Pis: David G. Anderson, Joshua Wells, Eric Kansa, Sarah Kansa, Stephen Yerka
# Descriptive Variables and Value(s)

<table>
<thead>
<tr>
<th>Descriptive Variable</th>
<th>Value(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Culture</td>
<td>Paleolithic, 10000 B.C.-8500 B.C.</td>
</tr>
<tr>
<td></td>
<td>Early Archaic</td>
</tr>
<tr>
<td></td>
<td>African-American</td>
</tr>
<tr>
<td></td>
<td>Historic-Unspecified</td>
</tr>
<tr>
<td></td>
<td>American 1821-present</td>
</tr>
<tr>
<td></td>
<td>Nineteenth C. American 1821-1899</td>
</tr>
<tr>
<td></td>
<td>Twentieth C American</td>
</tr>
<tr>
<td></td>
<td>Prehistoric-Unspecified</td>
</tr>
<tr>
<td>Site Type</td>
<td>Refuse (historic)</td>
</tr>
<tr>
<td></td>
<td>Other</td>
</tr>
<tr>
<td></td>
<td>Artifact scatter-low density</td>
</tr>
<tr>
<td></td>
<td>Historic town</td>
</tr>
<tr>
<td></td>
<td>Single Artifact</td>
</tr>
<tr>
<td></td>
<td>Road segment (historic)</td>
</tr>
<tr>
<td></td>
<td>Chimney ruin</td>
</tr>
<tr>
<td></td>
<td>Features-subsurface</td>
</tr>
<tr>
<td></td>
<td>Village (prehistoric)</td>
</tr>
<tr>
<td></td>
<td>Agriculture/farm</td>
</tr>
<tr>
<td></td>
<td>Land-terrestrial</td>
</tr>
</tbody>
</table>

## Site Name
ETNA TURPENTINE CAMP

<table>
<thead>
<tr>
<th>SHPO evaluation of NR eligibility</th>
<th>eligible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of SHPO evaluation</td>
<td>2009-03-27</td>
</tr>
<tr>
<td>National Register Reference Number</td>
<td>09001055</td>
</tr>
</tbody>
</table>
Digital Index of North American Archaeology (DINAA)

1. Rich metadata (cultures, chronology, site-types)
2. Reduced precision location data (site security, legal)
3. Data modeling challenges (using GeoJSON-LD, CIDOC-CRM, event models)
Using site file data to examine the impacts of sea level rise

In 100 years, 19,676 sites will be covered!
DINAA to help "bootstrap" a linked data ecosystem for North American archaeology
EOL Computable Data Challenge
(Ben Arbuckle, Sarah W. Kansa, Eric Kansa)
Large scale data sharing & integration for exploring the origins of farming.
Funded by EOL / NEH
1. 300,000 bone specimens
2. Complex: dozens, up to 110 descriptive fields
3. 34 contributors from 15 archaeological sites
4. More than 4 person years of effort to create the data!
Relatively collaborative bunch, Ben Arbuckle cultivated relationships & built trust over years prior to EOL funding.
Raw Data: Idiosyncratic, sometimes highly coded, often inconsistent.
Raw Data Can Be Unappetizing

See DIPIR study: http://dipir.org
Sometimes data is better served cooked.
Publishing Workflow

Improve / Enhance
1. Consistency
2. Context (intelligibility, interoperability)
Introduction

Understanding Epipaleolithic hunter-gatherer lifestyles and changes in their subsistence patterns in relation to environmental fluctuations is pivotal in understanding one of the milestones in the human evolution, namely, the transformation from exploitation of wild plant and animal resources to the production of domestic variants of these resources.

Öküzini Cave was initially subject to limited and occasional investigation by Ismail Kılıç Kökten from the mid-1950s to 1973. Excavations were resumed by a short-lived Turkish-German collaboration in 1985. Large-scale and systematic excavations in Öküzini were conducted by a large international team under the direction of the Museum of Antalya and supervised by İlhan Yalçınkaya between 1989 and 1999. The last project resulted in exhaustive studies including geology, lithic techno-typology, archaeobotany, archaeometry, and paleoethnology of the site, and was published as a monograph. The new excavations revealed 13 discrete geological horizons (GH 0 through XI) within a 3.5-meter Epipaleolithic sequence included a mixed protohistoric or Neolithic-Chalcolithic level disturbed by human burials preceded by Epipaleolithic layers that have been subdivided into four cultural phases or archaeological units (AUS 1, 2, 3, 4) based on characteristics of the lithic assemblages. The designated AUSs at Öküzini cover a temporal range from 16,460 to 12,000 uncalibrated years BP or from 19,790 to 12,900 calibrated years BP.

The primary focus of the zooarchaeological research at Öküzini was to examine a series of related topics such as: (1) Assemblage composition and characterization, (2) Changes in animal exploitation patterns and hunting strategies through time, (3) Mobility patterns, site function and inter-site variation, and (4) Periodicity in animal exploitation.

Methodology

The recording process involved two stages:

1. General documentation of the entire assemblage for the purpose of assemblage characterization (e.g., degree of fragmentation, skeletal part representation, etc.). This level included every element, element portion, and non-identified fragments and splinters recovered.
2. Particular documentation of pre-determined attributes in relation to the particular questions that are being asked (e.g., kill-off patterns, seasonality, etc.). This level targeted selected elements and portions such as mandibles with teeth, loose mandibular teeth, pelvic acetabula, and all limb epiphyses.

Potential Applications of Data

These data add to the increasing body of evidence for epipaleolithic subsistence in western Anatolia.
<table>
<thead>
<tr>
<th>Element</th>
<th>Taxon</th>
<th>Certainty</th>
<th>Symmetry</th>
<th>Place</th>
<th>Cond of Frag</th>
<th>Number of</th>
<th>Z</th>
<th>AA</th>
<th>AB</th>
<th>AE</th>
</tr>
</thead>
<tbody>
<tr>
<td>116</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>25</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>25</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>116</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>116</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>116</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>116</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>33</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>116</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>116</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>116</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>116</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>116</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>95</td>
<td>25</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>31</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>25</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0.5</td>
<td>15</td>
<td>6</td>
</tr>
<tr>
<td>21</td>
<td>15</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>9</td>
</tr>
</tbody>
</table>
"Ovis orientalis"
http://eol.org/pages/311906/

- Documentation
- Review, editing
- Annotation

Ovis orientalis

Wild sheep

Code: 14

Sheep, wild

Code: 70

O. orientalis

Code: 15

Sheep (wild)

Code: 16
Ovis orientalis

Brief Summary

Learn more about this article

Biology

Mouflon tend to feed early in the morning and in the evening, resting during the day under an overhanging bush or rock, where they are well hidden. Mouflon are gregarious and form non-territorial herds grazing on grasses, plants, shrubs, and flowers. They are highly vocal animals with a variety of calls and barks.

IUCN threat status: Vulnerable (VU)
1. Needed to mint new concepts like “sheep/goat”
2. Vocabularies need to be responsive + dynamic, esp. for multidisciplinary uses
**Bone J8c22.31**

**Class:** Animal Bone

**Archaeological Context:**
- **Tuczeń / Öküzini Cave / VII / Square J8c**

---

### Descriptive Properties (23)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Archaeological Horizon</td>
<td>22</td>
</tr>
<tr>
<td>Osteo Id</td>
<td>Metatarsus III + IV</td>
</tr>
<tr>
<td>Taxonomic Id</td>
<td>Ovis orientalis</td>
</tr>
<tr>
<td>Articulation</td>
<td>Separate</td>
</tr>
<tr>
<td>Symmetry</td>
<td>Right</td>
</tr>
<tr>
<td>Sex</td>
<td>Unidentified</td>
</tr>
<tr>
<td>Pathology</td>
<td>No pathology</td>
</tr>
</tbody>
</table>

---

**Linked Data:**

- Open Context Zooarchaeology Annotations-Has anatomical identification :: UBERON (Uber Anatomy Ontology)-fused metatarsal bones 3 and 4
- Biological Taxonomy Vocabulary-Has Biological Taxonomy :: Encyclopedia of Life-Ovis orientalis
- Open Context Zooarchaeology Annotations-Has fusion character :: Open Context Zooarchaeology Annotations-Distal epiphysis unfused

---

**Suggested Citation:**

---

**Linked Media (0)**
ArchaeoML (XML Version)
Version-control (Github, XML Data)

---

**Copyright Licensing:**

Content is licensed with Creative Commons Attribution License. Attribution Required. Citation and hyperlinks for online uses.
Uber anatomy ontology

Class: fused metacarpal bones 3 and 4

- Term IRI: http://purl.obolibrary.org/obo/UBERON_0013587
- Definition: An element formed from the fusion of metacarpal 3 and metacarpal 4.

Annotations

- has_obo_namespace: uberon
- has_exact_synonym: fused metacarpal 3/4
- has_related_synonym: metacarpal 3+4
- id: UBERON:0013587
- comment: In many arthrodactyls, the 3rd & 4th metatarsals are fused, creating a main metatarsal.

Equivalents

- metacarpal bone and (has_fused_element some metacarpal bone of digit 3) and (has_fused_element some metacarpal bone of digit 4)

Class Hierarchy

- Thing
  - anatomical entity
    - material_anatomical_entity
    - anatomical_structure
      - organ
        - skeletal_element
          - bone_element
            - bone of appendage girdle complex
              - bone of free limb or fin
                - limb bone
                  - autopod bone
                    - digitopodium bone
                      - metapodium bone
                        - fused metapodial bones 3 and 4
                          - fused metatarsal bones 3 and 4
                            - fused metacarpal bones 3 and 4

Superclasses & Asserted Axioms

- fused metapodial bones 3 and 4
1. Needed a controlled vocabulary for skeletal anatomy

2. Better data modeling than common in zooarchaeology, adds quality
Linking to UBERON

1. Models links between anatomy, developmental biology, and genetics
2. Unexpected links between the Humanities and Bioinformatics!
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>TP6704.F1</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Thalassia</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>TP6704.F2</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Thalassia</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>TP6704.F3</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Thalassia</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Metcalfe's Law

\[ Y = X^*(X-1) \]

Critical Mass
- *Not* a neat model of progress to adopt a more productive economy. Very different, sometimes piecemeal adoption in different regions.

### Easy to Align

1. Animal taxonomy
2. Skeletal elements
3. Sex determinations
4. Side of the animal
5. Fusion (bone growth, up to a point)
Hard to Align (poor modeling, recording)
1. Tooth wear (age)
2. Fusion data
3. Measurements

Despite common research methods!!
Professional expectations for data reuse

1. Need better data modeling (than feasible with, cough, Excel)
2. Data validation, normalization
3. Requires training & incentives for researchers to care more about the quality of their data!
Data are challenging!

1. Decoding takes 10x longer
2. Data management plans should also cover data modeling, quality control (esp. validation)
3. More work needed modeling research methods (esp. sampling)
4. Editing, annotation requires lots of back-and-forth with data authors
5. Data needs investment to be useful!
One does not simply walk into Mordor Academia and share usable data…
Final Thoughts

Data require intellectual investment, methodological and theoretical innovation.

Institutional structures are poorly configured to support data-powered research.

New professional roles needed, but who will pay for it?
Thank you!

Special Thanks!

Harvard Center for Hellenic Studies & the German Archaeological Institute (DAI)