Designing a Prototype Digital Repository for Archaeological Information at the ASCSA

CNI Fall Meeting,
Washington D.C., December 8, 2008
About the ASCSA

• Founded in 1881
• A private foundation, relying on endowment
• Ca. 100 employees
• Split between Princeton, NJ, and Athens
• Four missions: teaching, research, archaeological exploration, and publication
• The largest of 17 foreign archaeological institutes in Athens (including DAI, EfA, BSA)
Managing Committee of ca. 300 scholars from 180 North American colleges and universities

| American Numismatic Society | Amherst College | Arcadia University | Arizona State University | Austin Peay State University | Bard College | Barnard College | Boston College | Boston University | Brevard College | Brigham Young University | Brock University | Brown University | Bryn Mawr College | Bucknell University | Buffalo State College | California State University, Fresno | California State University, Long Beach | Carleton College | Case Western Reserve University | City University of New York | Clark University | Colgate University | College of Charleston | College of New Jersey | College of the Holy Cross | College of William and Mary | Columbia University | Connecticut College | Cornell University | Creighton University | Dartmouth College | Davidson College | DePauw University | Dickinson College | Duke University | Dumbarton Oaks Research Library | Emory University | Fairfield University | Florida State University | Fordham University | Franklin and Marshall College | Gettysburg College | George Mason University | George Washington University | Georgetown University | Grand Valley State University | Grinnell College | Gustavus Adolphus College | Hamilton College | Hampden-Sydney College | Harvard University | Hollins University | Hunter College | Illinois State University | Illinois Wesleyan University | Indiana University | Indiana University (IUPUI) | Institute for Advanced Study | Institute of Fine Arts | Iowa State University | Johns Hopkins University | Kennesaw State University | Lawrence University | Louisiana State University | Loyola College in Maryland | Loyola University of Chicago | Macalester College | McGill University | McMaster University | Michigan State University | Middlebury College | Mount Holyoke College | Nebraska Wesleyan University | New York University | Northwestern University | Oberlin College | Ohio State University | Ohio University | Ohio Wesleyan University | Pemboke College | Pennsylvania State University | Pitzer College | Pomona College | Princeton University | Providence College | Purdue University | Radcliffe College | Randolph College | Randolph-Macon College | Rhodes College | Rutgers | San Francisco State University | Savannah College of Art and Design | Scripps College | Smith College | Smithsonian Institution | Southern Methodist University | Southwestern University | Stanford University | State University of New York-Buffalo | Swarthmore College | Sweet Briar College | Temple University | Texas A & M University | Texas Christian University | Texas Tech University | Trinity College | Trinity University | Tufts University | Tulane University | Union College | University of Arizona | University of Arkansas | University of British Columbia | University of California, Berkeley | University of California, Davis | University of California, Irvine | University of California, Los Angeles | University of California, Santa Barbara | University of Chicago | University of Cincinnati | University of Colorado | University of Florida | University of Georgia | University of Illinois at Chicago | University of Illinois, Urbana-Champaign | University of Iowa | University of Kansas | University of Manitoba | University of Mary Washington | University of Maryland | UMBC | University of Massachusetts | University of Michigan | University of Minnesota | University of Mississippi | University of Missouri, Columbia | University of Missouri, St. Louis | University of Montreal | University of Nebraska | UNC, Chapel Hill | UNC, Greensboro | University of North Florida | University of Notre Dame | University of Oklahoma | University of Pennsylvania | University of Pittsburgh | University of Rhode Island | University of Richmond | University of South Dakota | University of South Florida | University of Southern California | University of Southern Indiana | University of Tennessee | University of Texas at Austin | University of Toronto | University of Vermont | University of Virginia | University of Washington | University of Waterloo | University of Wisconsin | Valparaiso University | Vanderbilt University | Vassar College | Virginia Polytechnic Institute | Wabash College | Washington University | Wayne State University | Wellesley College | Wesleyan University | Westminster College | Wichita State University | Wilfrid Laurier University | Willamette University | William Marsh Rice University | Williams College | Wright State University | Yale University |
EU-funded digitization project 2006-2008

Digital assets now include:

• 7,000 archaeological photos (glass plates and negatives)
• 11,000 images from the Dragoumis papers (letters and photos)
• 33,000 images from the Gennadius scrapbooks (newspaper clippings, photos, ephemera of various sorts)
Excavations at Ancient Corinth (since 1896)
Digital Assets

Digital surrogates:
(EU funded digitization project; to preserve on-site archive)
• 135,000 digitized photographs
• 200,000 pages of excavation notebook pages
• 7,000 maps and plans

Born digital:
• Databases, digital photos, CAD models, GIS datasets
**Excavations at the Athenian Agora (since 1931)**

**Digital Assets**

**Born digital (PHI funding):**

- **Databases:**
  - Objects (164,000 records)
  - Features (7,530 records)
  - Conservation Reports (93,730 records)
  - Photography Metadata (46,740 records)
  - Plans Metadata (3,510 records)
  - Samples and Analyses (1,500 records)
  - Bibliography (815 records)

- **Image Files** (tif, jpeg, Photoshop, camera raw)
  - Final Files (46,740)
  - Work Files (37,730)
  - Museum QTVR (1,261 objects, at least 108 photographs per object = 136,188 images)
  - Museum digital photography (1,261 objects, 2-3 images per object)

- **Plan Files** (tif, jpeg, autocad, illustrator)
  - Final Files (811)
  - Work Files (7,050)

- **Raw Excavation Files** created during the excavation season
  - (field databases, images, drawings, notes, reports, etc.)
  - 54 GB of data with an average of 5GB/year added

**Digital surrogates (EU 2009-):**

- 180,000 pages of Excavation Notebooks (25,000 records for features database)
- 5,000 pages of end of season reports
- 10,000 sheets of drawings (digitization, vectorization)
- 60,000 photos (negatives, prints, glass plates)
- 100,000 coin envelopes
- 120,000 catalog cards
... and don’t forget the published materials, the administrative data, the scientific datasets, etc.

- 40,000 digitized pages of *Hesperia* and *Hesperia* Supplements
- 20,000 digitized pages of *Agora* and *Corinth* monographs
- Now born-digital publications and online-only supplementary materials for *Hesperia*
- Digital datasets and images from Wiener Laboratory for archaeological science.
- Administrative databases, financial records (all with access issues)
- Library databases, electronic journals and books.
Affiliated archaeological projects
3 permits a year. Each season generates ca. 40 GB, and often doesn’t have a secure place to store that data.

e.g., Mount Lykaion (Sanctuary of Zeus)
US sponsors: University of Pennsylvania Museum and University of Arizona

e.g., Mitrou (prehistoric settlement)
US sponsor: University of Tennessee
Active ASCSA Archaeological Projects
(particular concentration in Corinthia and Crete)
2005 realization: An increasing amount of digital stuff, including more and more born digital. What to do?

Application to the Mellon Foundation for help in “the organization and deployment of information resources.” $290,000 in funding received June 2006. Invaluable recommendations of Beth Forrest Warner and Susan Perry.

- **Website redesign.** A way of thinking through our structure.
- **Coordinated service model.** Efficiencies / redeployed staff.
- **Staff training and development.** New skills for a digital world.
- **Information architecture development.**
“The effort to develop a digital library architecture that can serve the ongoing needs of the School presents all of the problems which confront academic research library collections, while adding the new complications of handling large, heterogeneous collections of excavation data”

Thornton Staples, November 30, 2006
What are the particular challenges of managing archaeological data in digital form?

• Many different types of file format.
• Large datasets (at least for the humanities).
• Durability a major issue since born digital data is truly irreplaceable.
• Lack of common metadata standards and a tradition of silos.
• Intellectual property considerations similar to art history plus a need to restrict unpublished data.
• Data is socially constructed. History of interpretation crucial.
• Information doesn’t make sense in isolation, but only in context.
These issues informed the design of the prototype

• Focus on archaeological data, but good for all materials.
• Primary goal: data from different excavations in single system. The metadata systems for different excavations have strong overlaps but specific semantics. Preserving formats of source databases rather than imposing a single system.
• Presents data in a form that makes sense to archaeologists. Some content is surrogate information for entities which no longer exist (units of excavation, referred to as baskets). Some content represents speculations about what originally existed (lots and features). Various relationships among entities are possible, and can change.
• Durable and flexible XML-based architecture. Each archeological entity is an XML file, using a non-standard schema. Image descriptions and relationships given in the entity files. Relationships among the entities are asserted from the “child” entity file explicitly. Entity relationships are indexed as RDF triples in a “standard” Mulgara index. Full-text metadata indexed as RDF triples in a “Lucene” Mulgara index.
Existing excavation databases are mapped to a data dictionary (in future, CIDOC or VRA Core 4 schema)

<table>
<thead>
<tr>
<th>Data Dictionary</th>
<th>Agora</th>
<th>Corinth Objects</th>
<th>Corinth Unregistered Finds</th>
<th>Corinth Unregistered Flots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Title</td>
<td>objhead</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Authority</td>
<td>Authority</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grids</td>
<td>Grids</td>
<td>localbspot</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mints</td>
<td>Mints</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regions</td>
<td>Regions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Storage</td>
<td>Storage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XyzPolygons</td>
<td>XyzPolygons</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronology</td>
<td>Chronology</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DateOfDiscovery</td>
<td>DateOfDiscovery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NotebookDates</td>
<td>NotebookDates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Periods</td>
<td>Periods</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Category</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classifications</td>
<td>Classifications</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Keywords</td>
<td>Keywords</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ArchiveNotes</td>
<td>ArchiveNotes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ConservationStatus</td>
<td>ConservationStatus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Context</td>
<td>Context</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DileOlvers</td>
<td>DileOlvers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DileRevers</td>
<td>DileRevers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GreekDescription</td>
<td>GreekDescription</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Handling</td>
<td>Handling</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Context</td>
<td>Context</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing</td>
<td>Writing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>who</td>
<td>who</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>writing</td>
<td>writing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>decor</td>
<td>decor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>technique</td>
<td>technique</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>consor</td>
<td>consor</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>other</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>held</td>
<td>held</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pot</td>
<td>pot</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>held</td>
<td>held</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A network of independent XML information objects (some shown below). Indexed on the fly.
Next step: turning digital projects into digital program
(currently we have only two IT staff, and no project managers)

• How many staff do we need to add? In which departments? At what level? With what skills?

• What are our hardware and software needs? How can we keep these costs under control?

• Should we continue to develop a Fedora repository, or should we go for something simpler? At least in the short term.

• What kind of permanent governance structure should we put in place to coordinate and prevent departmental drift?
Three possible scenarios for implementing digital repository (whether prototype or something simpler)

**Scenario 1:** Go it alone, perhaps outsourcing some functions to a commercial vendor?

**Scenario 2:** Partner with one / some of our Managing Committee institutions?

**Scenario 3:** Partner with other foreign archaeological schools and the Greek Ministry of Culture?
Scenario 1: Go it alone
(with own staff or vendor, e.g., BePress, VTLS)

Pro:
• Control of our data. This is a particular concern of excavators.
• Flexibility to build a bespoke system that really works for classical archaeology.
• Makes ASCSA look attractive to granting agencies.

Con:
• Potentially huge costs. Can we afford it?
• Lack of skills within the organization. Can we hire them?
• Danger of perpetuating silos, and building an idiosyncratic system that turns out not to be interoperable.
Scenario 2: Partner with North American institution(s)
(e.g., California Digital Library, University of Kansas, New York University, Archaeoinformatics.org)

**Pro:**
- Access to best practices and a well-maintained infrastructure.
- North American funders value collaborations.
- Potentially most cost-effective.

**Con:**
- Potentially boxed into a system unsuitable for our kind of data.
- How do we prevent being subsumed, and dropping to the bottom of the pile?
- Association with one institution may alienate others.
Scenario 3: Collaboration with partners in Greece (e.g., DAI, Polytechneion, Ministry of Culture)

Pro:
• Closest to the data. Best opportunity to reveal new links / create regional perspectives.
• Access to substantial European funding.
• Politically sensible.

Con:
• Different national and organizational cultures of potential partners may lead to tension.
• Danger of being sucked into a “pan-European gargleblaster”
• Such projects have a poor track record of sustainability.
Collaboration between ASCSA and DAI a logical one (Application for NEH-DFG grant to explore this – Fall 09)
Thank you to the Andrew W. Mellon Foundation and the Greek Ministry of Culture

Charles Watkinson
cwatkinson@ascsa.org
Tel: 609 683 0800 x 21
http://www.ascsa.edu.gr

(presented on behalf of my colleagues in the information Architecture Team who have done all the work: Tarek Elemam, Bruce Hartzler, James Herbst, Carol Stein, Thornton Staples)