Making Square Pegs Fit
The Power of Working Together

Ching-Hsien Wang
wangch@si.edu

http://Collections.si.edu

Smithsonian Institution
Making Square Pegs Fit

The Power of Working Together

Primary Team Members:

Andrew Gunther,
Jim Felley,
George Bowman,
Randy Arnold,
Mike Trigonoplos.
The Public perception of the Smithsonian: One Institution
Background

Smithsonian consists of

- 19 museums,
- 20 libraries,
- 14 archives,
- 1 National Zoo,
- 1 Astrophysical Observatory
- Research centers in Panama, Boston, New York, Maryland, and Virginia.
Diverse Material Types
No Unified Access To Smithsonian’s Collections
Seeking for a Perfect Solution
Scope of the Collections Search Center

- 2.7 million records,
- 276,000 images,
- 40 data sources from Libraries, Archives and Museum collection databases.

Books, serials, trade catalogs, photographs, paintings, sculptures, manuscripts, letters, postage stamps, postcards, sound recordings, posters, decorative arts, ceramics, maps, portraits, scientific specimens, rockets, airplanes, etc.
Features of the Collections Search Center

- **Keyword searching**, 
- **View options**, 
- **Slideshow**, 
- **Expand/collapse view**, 
- **Share this**, 
- **Facet categories**, 
- **Time-slider**, 
- **Blogs**
Objects and Materials Working Together

Warren MacKenzi, potter

- 6 ceramic objects from American Art Museum
- 10 books about Warren Mackenzi and American potters from Library
- 3 interview transcripts from Archives,
- 1 sound recording of Oral History from Archives,
- 3 letters written by Warren Mackenzi from Archives,
- Two more related collections from Archives,
Objects and Materials Working Together

Birds in Panama

- 24,800 specimen from National Museum of Natural History,
- 13,100 collected or donated by Alexander Webmore
- 8 institutional records with online finding aids from archives,
- 37 photographs of him or taken by him while working in the field,
- 4 collections includes diaries, manuscripts of Mr. Wetmore,
- 2 Oral History Interviews,
- 86 books he owned,
- 3 portrait paintings of Mr. Wetmore.
Objects and Materials Working Together

**Alexander Calder, Artist**

- National Postal Museum (5)
- Hirshhorn Museum and Sculpture Garden (48)
- Smithsonian American Art Museum (32)
- National Portrait Gallery (30)
- Smithsonian Institution Libraries (192)
- Photograph Archives, Smithsonian American Art Museum (301)
- Archives of American Art (160)
- Archives of American Gardens (15)
- Smithsonian Institution Archives (9)
Road Map and Approach

- **Artesia Digital Asset Management System**
- **NMAfA**
- **FSGA**
- **NMAAHC**
- **NPG**
- **SAAM**
- **CHNDM**
- **NASM**
- **NPM**
- **ACM**

**Horizon 8**
- **SIL**
- **Archives**
- **SAAM - ARI**
- **SAAM - Juley**
- **SAAM - AECI**
- **Bibliographies**
- **History of St. Airplanes**

**EMu 11**
- **Anthropology**
- **Botany**
- **Entomology**
- **Mineral**
- **Paleobiology**
- **Birds**
- **Fishes**
- **Herpetology**
- **Mammals**

**TMS 10**
- **NMAfA**
- **NPM**
- **ACM**
- **NMAH**
- **NASM**
- **FSGA**
- **NMAAHC**
- **NPG**
- **SAAM**
- **CHNDM**

**Mimsy 1**
- **AAA**
- **Other**

**Artesia**
Digital Asset Management System
Road Map and Approach
Process Flow Diagram

Horizon

Data Extract and Transformation

XML documents

Data standardization Processing

Solr

Lucene Index

Output data In XML

Output data In JSON

Output data In Python

Smithsonian Photographic Initiative

Online Exhibition

Collections Search Center

Education Interface

Open Access Applications

Digital Library

Data Extract and Transformation

XML documents

Museum TMS

Data Extract and Transformation

XML documents

Museum EMU

Data Extract and Transformation

XML documents

Digital Archives

Data Extract and Transformation

XML documents
Making it work

- Prototype a smaller system including library, archives and Art inventory project in 2007 with 8 Horizon DBs,

- Metadata Index Model creates the framework and data structure for bibliographic, archival, three dimensional objects and scientific specimens

<table>
<thead>
<tr>
<th>Title/Object name</th>
<th>Object Type</th>
<th>Culture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identifier</td>
<td>Publisher</td>
<td>Set Name</td>
</tr>
<tr>
<td>Physical Description</td>
<td>Name</td>
<td>Date Source</td>
</tr>
<tr>
<td>Notes</td>
<td>Language</td>
<td>Credit Line</td>
</tr>
<tr>
<td>Taxonomic Name</td>
<td>Topic</td>
<td>Object Rights</td>
</tr>
<tr>
<td>Place</td>
<td>Record Link</td>
<td>Online Media Group</td>
</tr>
</tbody>
</table>
Making it Work

• **Free-text** elements and structured elements creates the balance of keyword search and browse control,

• Attribute elements that allow flexibility and **control of display**, 

  **Name data element**
  Any people, groups (except cultures), titled presentations (exhibitions, expeditions) associated with the object or resource.

  `<freetext category="name" label="Author">`
  `<freetext category="name" label="Creator">`
  `<freetext category="name" label="Artist">`
  `<freetext category="name" label="Maker">`
  `<freetext category="name" label="Sitter" role="sitter">`
Devoting Efforts into Working Together Across Smithsonian

• Look at data beyond the surface and dig deep for commonality.
  – Library, Archives and Museums are three separate professions with different cataloging standards,

• Encourage participation by addressing unit specific concerns and seek solutions,
  – Listen very carefully, and address every concern!
Devoting Efforts into Working Together Across Smithsonian

• Start with willing partners and show case rewarding results to influence others
  – Focus on positive element instead of fighting with the negatives.
  – Started with SIRIS users (Horizon databases)
  – Postal Museum, Portrait Gallery and American Art Museum are our early museum implementers.
Devoting Efforts into Working Together Across Smithsonian

• Focus on collaboration and avoid competition.
  – Generate links back to home site and increase web traffic to museum web sites.

• Use standards whenever possible to move forward
  – MARC, CDWLITE, MODS, Dublin Core,
  – AAT, LCSH, ICZN, ICBN
Devoting Efforts into Working Together Across Smithsonian

• Use technology to accommodate differences
  – Create flexible data structure to accommodate special cases. Free-text vs. structure data elements.
  – Custom programming to standardize data:
    • Scrub data at data extraction time (database specific rules),
    • Supply data elements to cover assumed data elements,
    • Create and apply data filters at data ingest time for mass standardization across the institution
Into the Nitty-Gritty

• Facet terms transformation from MARC headings

Name facet from Main Entry tag 100

100 1 $aCaldenby, Claes,$d1946-$tAsplund.$lEnglish
Name=Caldenby, Claes,

100 1 $aEllington, Duke,$d1899-1974
Name=Ellington, Duke
Into the Nitty-Gritty

- Facet terms transformation from hierarchical terms

Topic facet from Subject tag 650

650 00 $aArt $y20th century $x Criticism and interpretation.

- Topic=Art,
- Date= 20th century,
- Top= Criticism and interpretation

650 00 $aArt$zAlabama$zBirmingham.

- Topic=Art,
- Place=Alabama,
- Place=Birmingham,
- Object type= Periodicals
Into the Nitty-Gritty

• Facet terms transformation from hierarchical terms

Object Type facet from form & genre tag

655 $aPhotographs $y1850-1900 $vBlack-and-white photoprints

Object type=Photographs,
Date=1850-1900,
Object type= Black-and-white photoprints

655 $aPostcards $y20th century $zUnited States

Object type=Postcards,
Date=20th century
Place=United States
Into the Nitty-Gritty

- Data massage when we extract data from original databases
  - Transform data based on local database specifics,
    - Example: First/Last name re-order,
    - Expand abbreviations,
    - Separate or concatenate data values
  - Supply assumed data content,
    - Example: stamps, works of art, American Indians, American art, type specimens,
Into the Nitty-Gritty

- Data massage when we ingest data into the mass index

*Object type terms (~3500)* standardized using AAT as a guide

There are 166 terms mapped into “Photographs”

Examples:

- acetate negative
- Acetate film
- Negatives (photographic)
- Aerial views
- albumen print
- Aerial shots of countryside
- Ambrotypes
- Autochrome process
- b&w negatives
- Carbro-color prints
- b&w negatives
- Banquet camera photographs
- Cellulose nitrate
- Chromogenic color prints
- Black-and-white transparencies
- Chromogenic processes
- Blueprint process
- negatives Acetate film
- Cyanotypes
- Daguerreotypes
- Dye destruction process
- Glass plates
- Interpositives
- Kodachrome
• Data massage when we ingest data into the mass index

*Date* are transformed into standard date ranges: Decades: 1500-present, Centuries: 0 – 1500, and Millennia: BEC.

Examples:

1945 >> 1940s, 1865-1890 >> 1860s, 1870s and 1890
ca. 1756 >> 1750s, 20th Century >> 1900s, 1910s, 1920s, ... 1990s
January 25th, 1877 >> 1870s, Yuan dynasty (1279 - 1368) >> 1200s, 1300s,
195x >> 1950s, 1934-55 >> 1930s, 1940s, 1950s
Remaining Challenges

• More data to load (3 million more records) for the initial phase:
  – American History Museum, 860,000 records
  – Department of Botany: 784,720 records
  – Department of Invertebrate Zoology: 918,568 records
  – Department of Mineral Sciences: 383,812 records
  – Department of Paleobiology: 589,696 records
  – Division of Fishes: 326,767 records
  – Division of Amphibians and Reptiles: 557,435 records
  – Division of Mammals: 579,232 records
  – Smithsonian Tropical Research Institute 150,000 records
Remaining Challenges

- Implement Geo-Location code and map filter
- Build more web applications using data in EDAN,
- Standardized Topic, Culture terms
- Explore hierarchical facets for object type, data source, and Taxonomical terms
- And more...
Question?

Ching-Hsien Wang
wangch@si.edu

http://Collections.si.edu
Smithsonian Institution