



Discovery Center for Libraries, Archives and Museums

Ching-hsien Wang

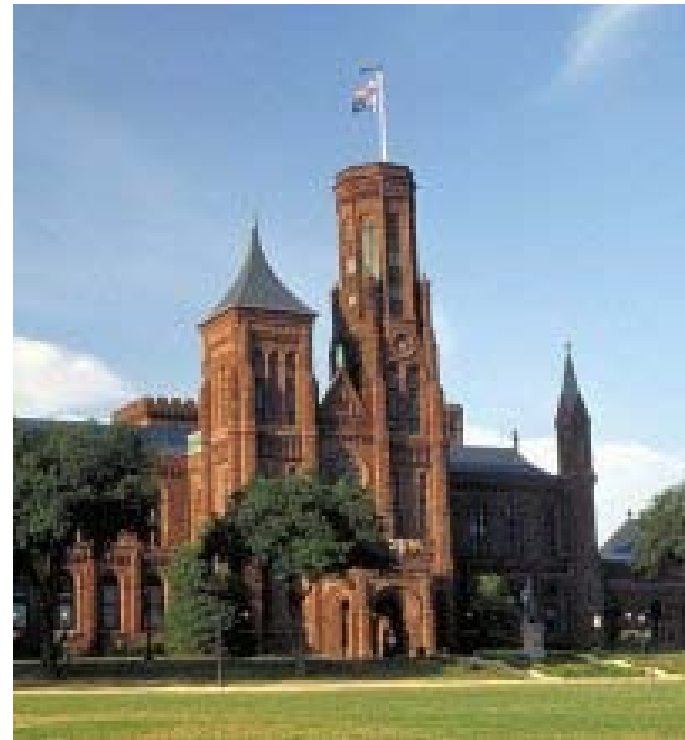
Jim Felley

Smithsonian Institution

Office of

Chief Information Officer

<http://collections.si.edu>

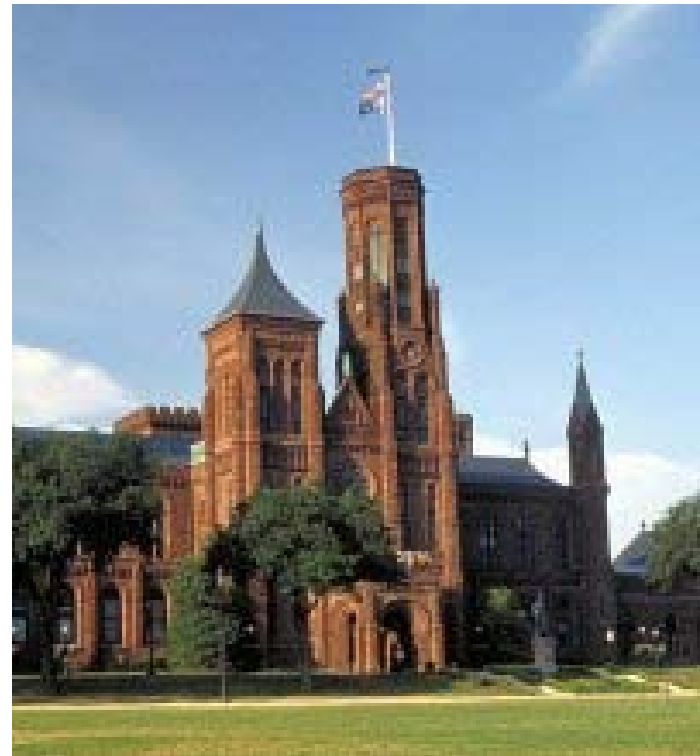




Discovery Center for Libraries, Archives and Museums

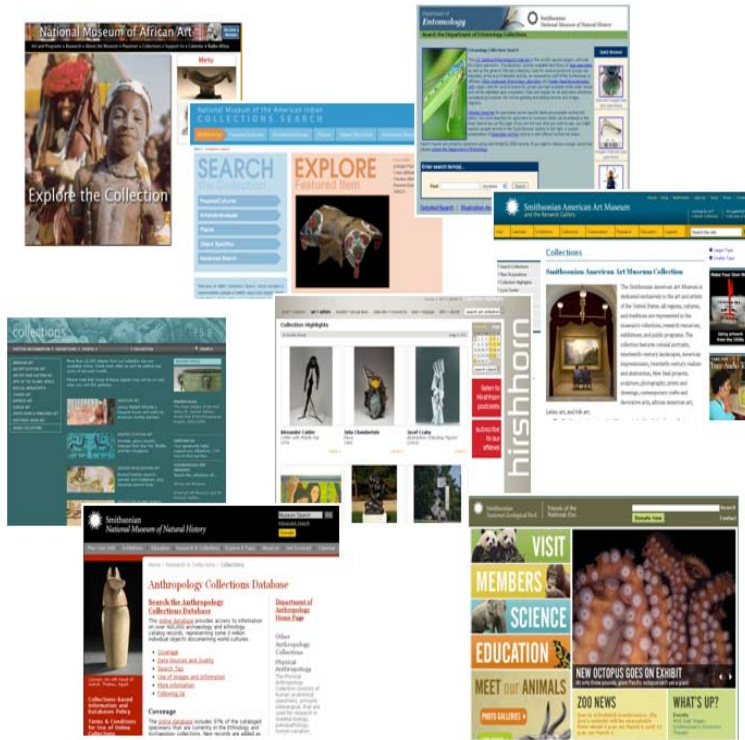
Rich and Complex Smithsonian

- 19 museums,
- 20 libraries,
- 14 archives,
- 1 National Zoo,
- 1 Astrophysical Observatory
- Research centers in Panama, Boston, New York, Maryland, and Virginia.





No Unified Access To Smithsonian's Collections



From Astronomy to Arts

- 137 million objects,
- 650 Silo databases
- 150 public web sites



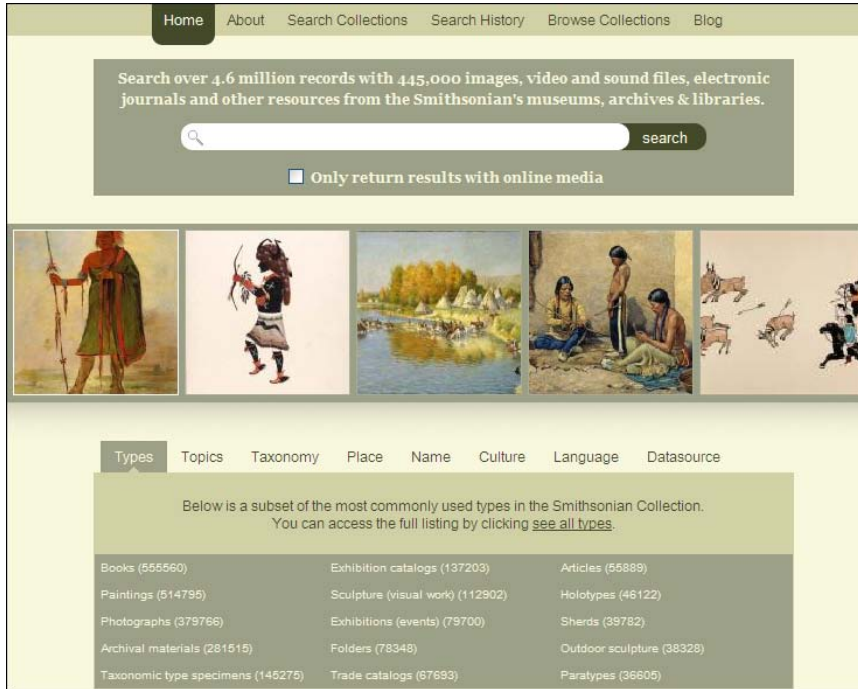
Core Team to a Solution



- Andrew Gunther,
- Jim Felley,
- Mike Trigonoplos
- George Bowman
- Randy Arnold
- Ching-hsien Wang



Scope of the Collections Search Center



The screenshot shows the Smithsonian Collections Search Center website. At the top, there is a navigation menu with links for Home, About, Search Collections, Search History, Browse Collections, and Blog. Below the menu is a search bar with a magnifying glass icon and a 'search' button. A checkbox labeled 'Only return results with online media' is visible below the search bar. The main content area features a row of five small images: a Native American figure with a bow, a Native American figure with a bow, a landscape painting of a river, a Native American scene with people, and a collection of small images including a dog and a person. Below the images is a section titled 'Types' with a list of categories and their counts:

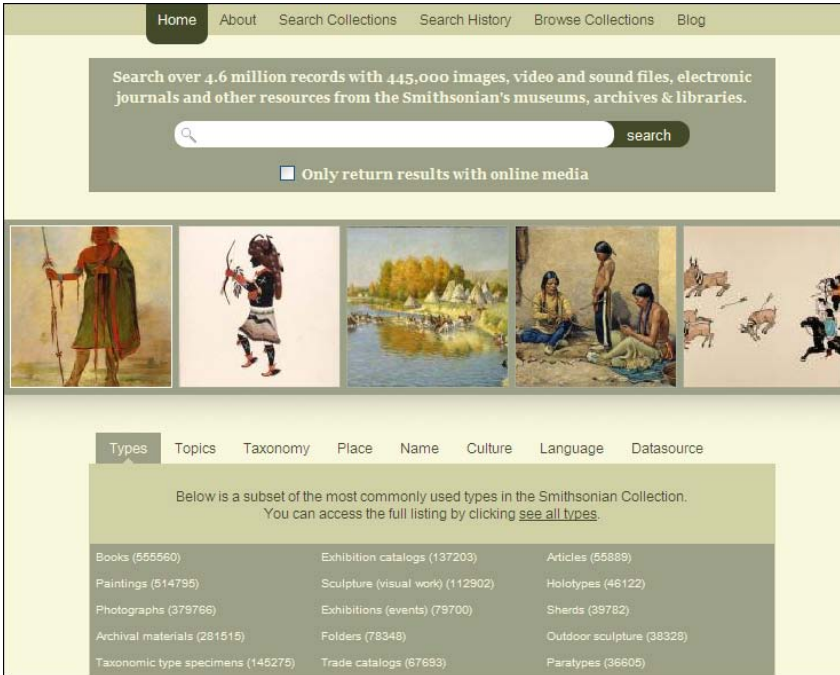
Types	Topics	Taxonomy	Place	Name	Culture	Language	Datasource
Books (555560)	Exhibition catalogs (137203)	Articles (55889)					
Paintings (514795)	Sculpture (visual work) (112902)	Holotypes (46122)					
Photographs (379766)	Exhibitions (events) (79700)	Sherds (39782)					
Archival materials (281515)	Folders (78348)	Outdoor sculpture (38328)					
Taxonomic type specimens (145275)	Trade catalogs (67693)	Paratypes (36605)					

- 5.4 million records,
- 460,000 images,
- 30 databases from Libraries, Archives Museum and Research Centers.

Photographs, paintings, manuscripts, letters, postage stamps, scientific specimens, rockets, airplanes, postcards, sound recordings, posters, decorative arts, ceramics, maps, sculptures, books, trade catalogs, etc.



Features of the Collections Search Center



The screenshot shows the Smithsonian Collections Search Center interface. At the top, there is a navigation bar with links: Home, About, Search Collections, Search History, Browse Collections, and Blog. Below the navigation bar is a search area with a text input field containing a magnifying glass icon and a 'search' button. A checkbox labeled 'Only return results with online media' is located below the search field. Below the search area is a row of five small images: a person in a green cloak, a person with a bow, a landscape with a river, a person sitting on the ground, and a group of people with animals. Below the images is a 'Types' section with a navigation bar containing links: Types, Topics, Taxonomy, Place, Name, Culture, Language, and Datasource. Below the navigation bar is a text box stating: 'Below is a subset of the most commonly used types in the Smithsonian Collection. You can access the full listing by clicking [see all types](#).' Below the text box is a table of types and their counts.

Books (555560)	Exhibition catalogs (137203)	Articles (55889)
Paintings (514795)	Sculpture (visual work) (112902)	Holotypes (46122)
Photographs (379766)	Exhibitions (events) (79700)	Sherds (39782)
Archival materials (281515)	Folders (78348)	Outdoor sculpture (38328)
Taxonomic type specimens (145275)	Trade catalogs (67693)	Paratypes (36605)

- [Keyword searching,](#)
- [Facet browse & filtering,](#)
- [Display View options,](#)
- [Slideshow,](#)
- [Time-slider,](#)
- [Share this,](#)
- [Blogs](#)



The Result Example

Alexander Calder, Artist



Archival Collections

Books

Photographs

Drawings

Paintings

Sculptures

Stamps

Interviews

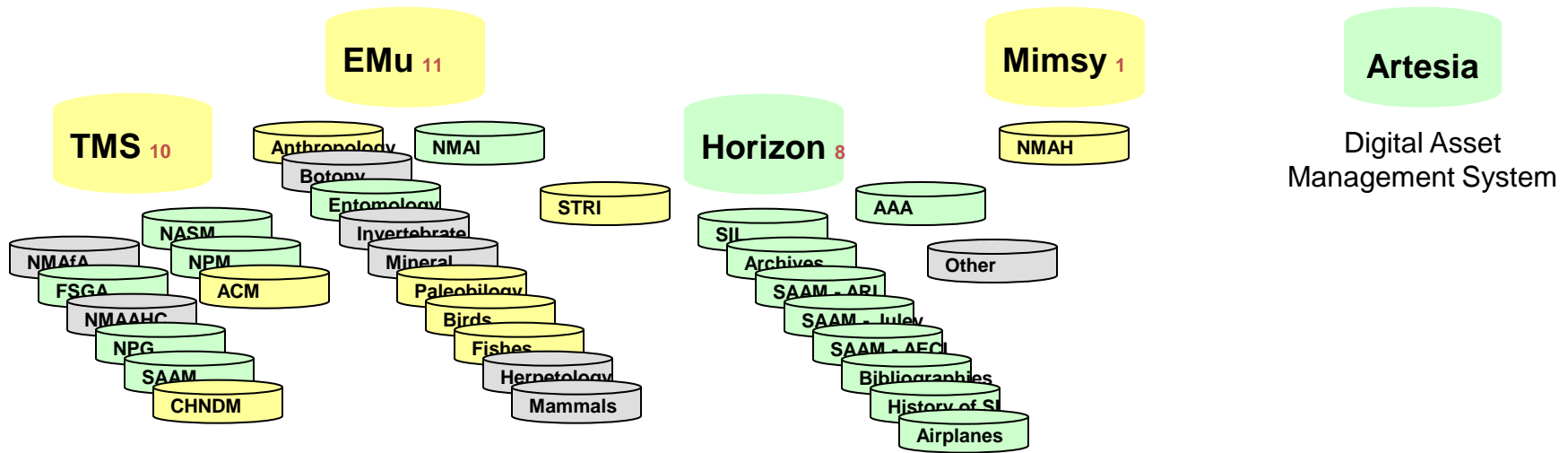
Sound Recordings

- 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)



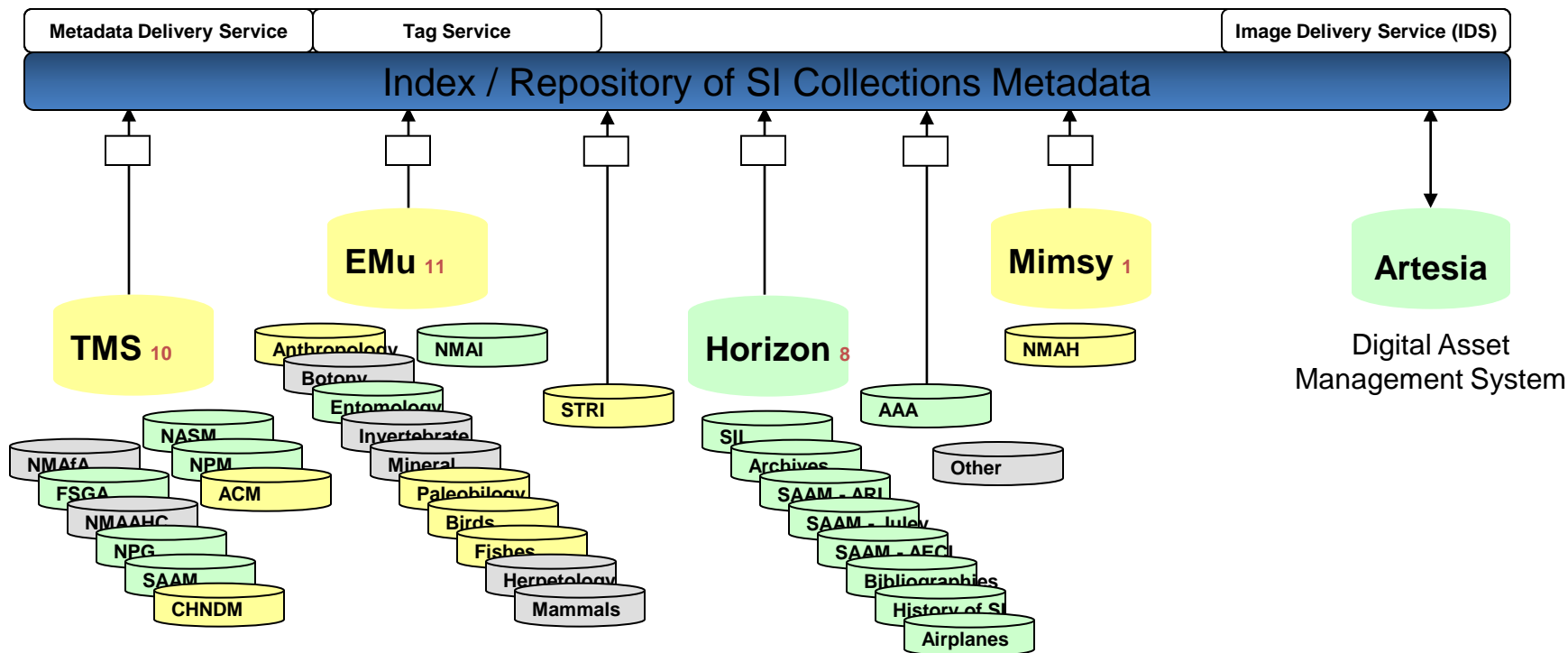
Smithsonian
Institution

Road Map to a Solution





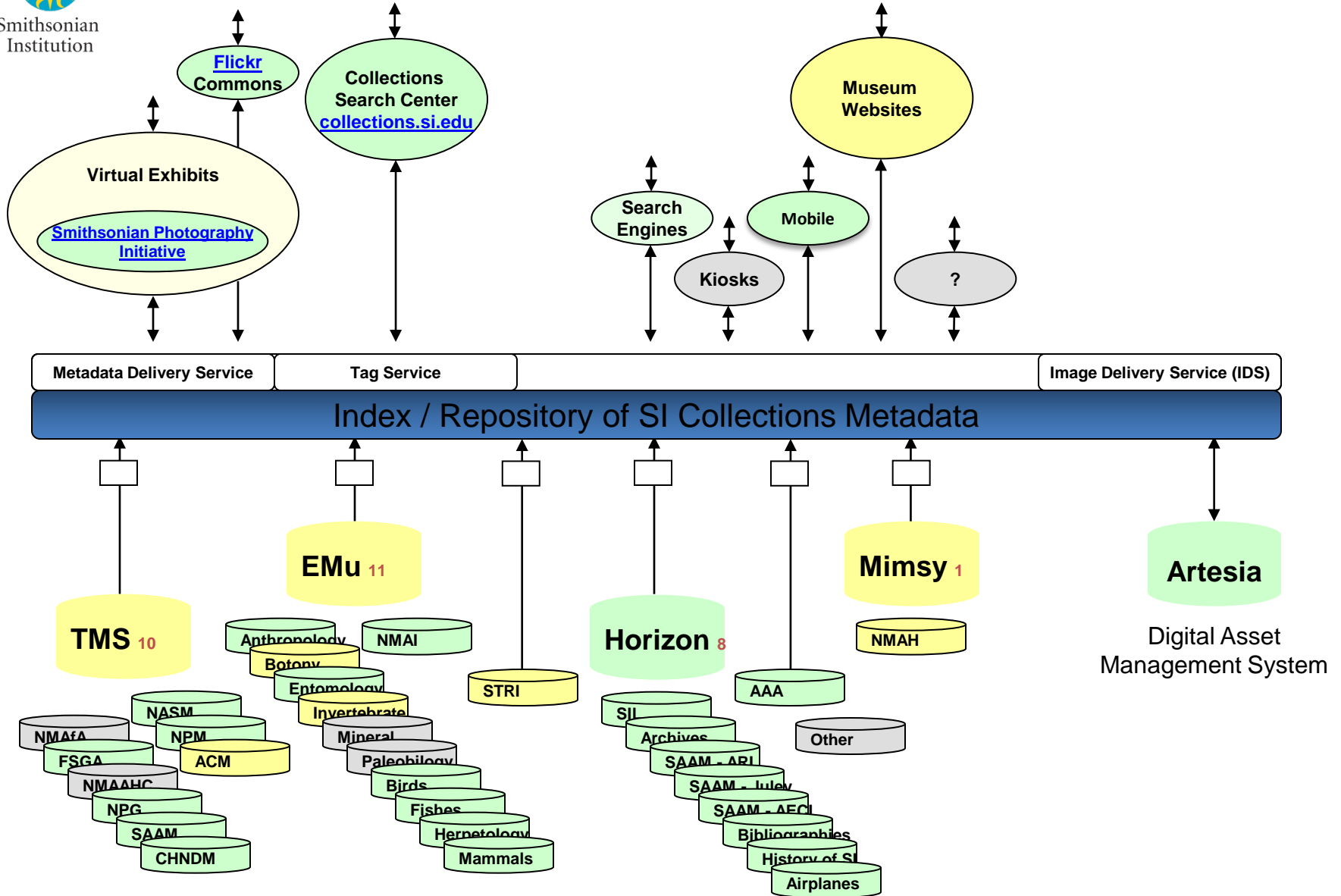
Enterprise Digital Access Network (EDAN)





Smithsonian Institution

Possibilities of using EDAN





Smithsonian
Institution

Set Data Standards

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

Title/Object name	Object Type	Culture
Identifier	Publisher	Set Name
Physical Description	Name	Date Source
Notes	Language	Credit Line
Taxonomic Name	Topic	Object Rights
Place	Record Link	Online Media Group



Configure Solr Schema

Three types of fields :

- Free text fields:

```
<field name="freetext" type="jsonField" indexed="false" stored="true" multiValued="false"
compressed="true" />
```

- Facet fields:

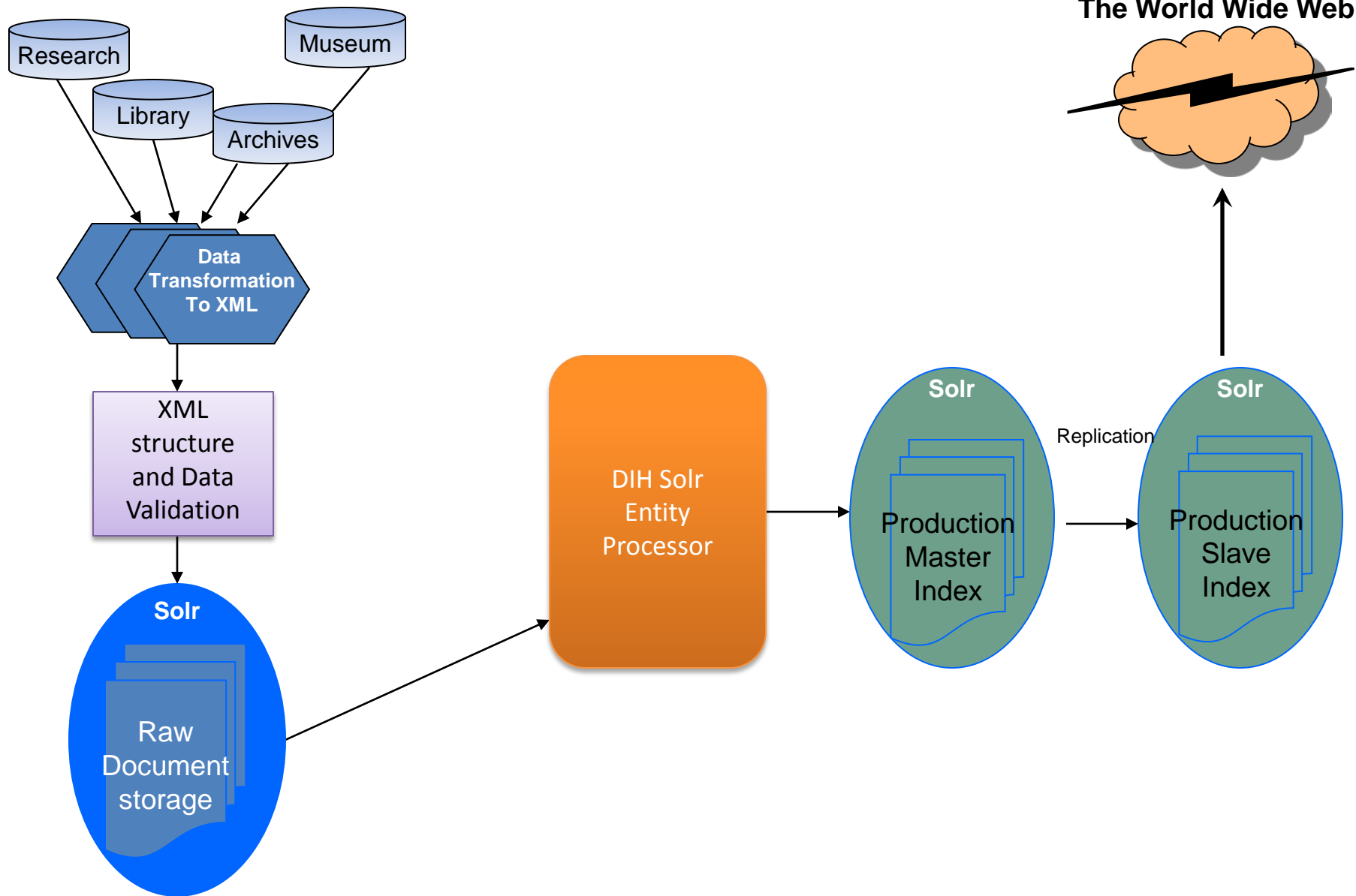
```
<field name="object_type" type="string" indexed="true" stored="false" multiValued="true" />
<field name="online_media_type" type="string" indexed="true" stored="false" multiValued="true" />
<field name="topic" type="string" indexed="true" stored="false" multiValued="true" />
```

- Record identification fields

```
<field name="record_ID" type="string" indexed="true" stored="true" multiValued="false" />
<field name="data_source" type="string" indexed="true" stored="true" multiValued="true" />
<field name="unit_code" type="string" indexed="true" stored="true" multiValued="true" />
```

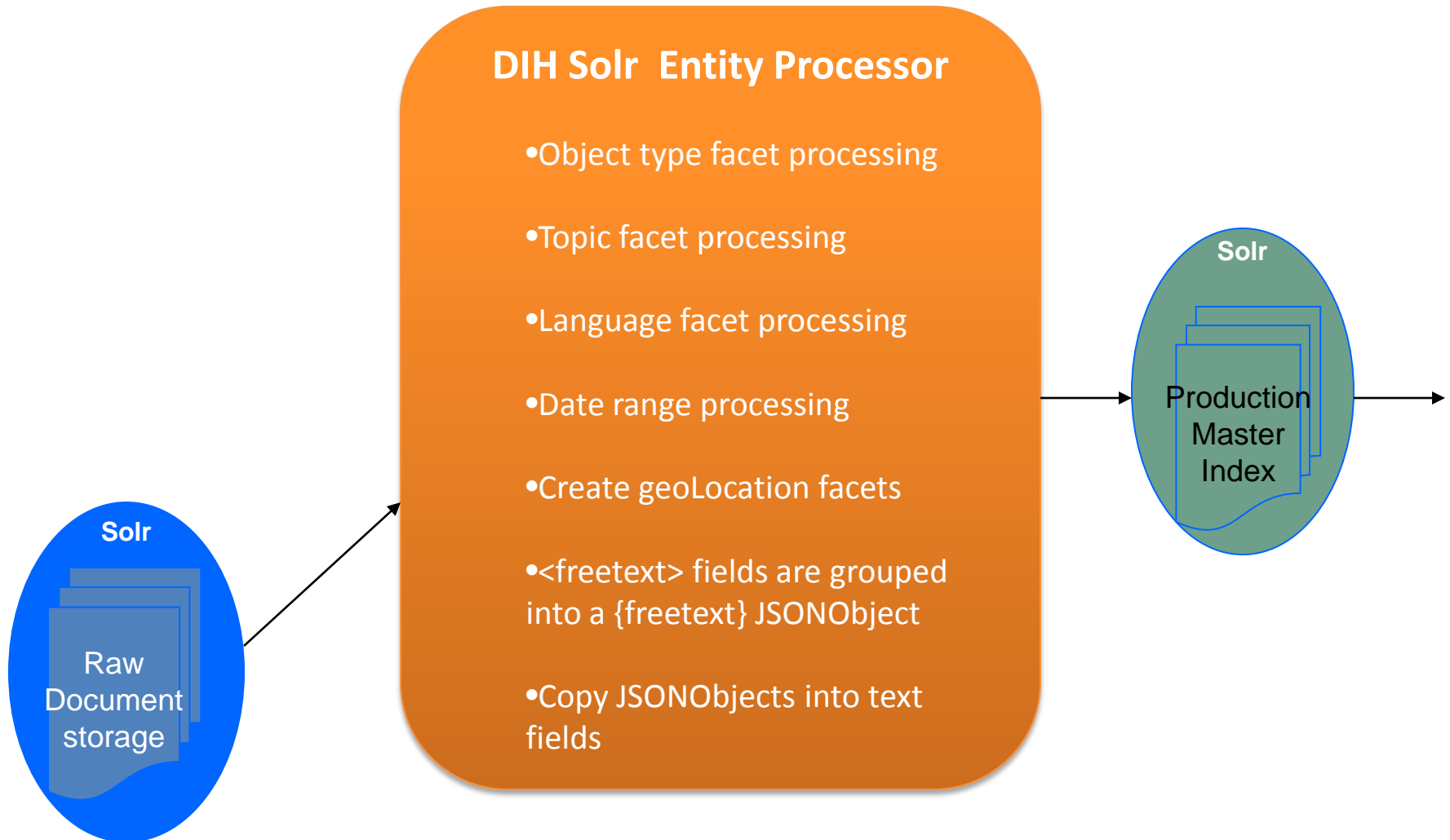


Activity Workflow





Using Data Input Handler





Raw Document Storage Index Schema Configuration

Store large block of data with simple configuration:

```
<field name="recordID" type="string" indexed="true" stored="true" />
```

```
<field name="loaderID" type="string" indexed="true" stored="true" />
```

```
<field name="deleted" type="boolean" indexed="true" stored="true" default="false" />
```

```
<field name="rawrec" type="XMLField" indexed="false" stored="true"  
compressed="true" />
```

```
<field name="timestamp" type="date" indexed="true" stored="true" default="NOW" multiValued="false" />
```



Where 'rawrec' is a CDATA block with the entire raw record.



Smithsonian
Institution

Data Stored in CDATA block

```

<freetext category="dataSource" label="Data Source">National Anthropological Archives</freetext>
<freetext category="identifier" label="Local number">NAA MS 7144</freetext>
<freetext category="identifier" label="Local number">NAA INV 10000000</freetext>
<freetext label="Creator" category="name">ANONYMOUS</freetext>
<freetext category="physicalDescription" label="Physical description">1 page 43 cm x 151.1 cm</freetext>
<freetext category="notes" label="Notes">The Korea and Manchuria Tour Guide Office (Sem-Man Annaijo) was a Japanese
    offices in Tokyo, Osaka, and Shimonoseki</freetext>
<freetext category="notes" label="Summary">The item is a piece of artwork that presents a bird's-eye view of Korea,
    regions and includes railroad routes and waterways. It is illustrated to show people, buildings and habitations, domestic
    fauna, and modes of transportation. Also included is a likeness of the Manchu emperor Pu-yi and a man hailing the
    Annotations, perhaps added by the collector, provide romanized versions of some place names and the title "Manchoukuo
    the mount</freetext>
<freetext category="notes" label="Cite as">Manuscript 7144, National Anthropological Archives, Smithsonian Institution</freetext>
<freetext category="objectType" label="Type">Maps</freetext>
<freetext category="objectType" label="Type">Collection descriptions</freetext>
<freetext category="objectType" label="Type">Maps, Pictorial</freetext>
<freetext category="objectType" label="Type">Posters</freetext>
<freetext label="Subject" category="name">Pu Yi, Henry President (1932-1934) and emperor (1934-1945) of Manchoukuo Born 1906 depicted
    Manchu</freetext>
<freetext category="culture" label="Culture">Japanese</freetext>
<freetext category="culture" label="Culture">Manchu</freetext>
<freetext category="place" label="Place">Korea</freetext>
<freetext category="place" label="Place">Manchuria</freetext>
<freetext category="date" label="Date">1933?</freetext>
<freetext category="setName" label="See more items in">Numbered manuscripts 1850s-1980s (some earlier)</freetext>

```





Object Type Facet Processing

Data pattern matching against an SQL database for controlled vocabulary

Example:

166 terms mapped into “*Photographs*”

acetate negative

Acetate film

Aerial views

albumen print

Ambrotypes

Autochrome process

Cellulose nitrate

Chromogenic color prints

Chromogenic processes

Blueprint process

Cyanotypes

Daguerreotypes

Glass plates

Interpositives

3500+ *Object type terms* are standardized



▼ Type		
Frequency	Alphabetical	
Archival materials	+ -	(284609)
Articles	+ -	(57127)
Books	+ -	(567275)
Catalogs	+ -	(34125)
Exhibition catalogs	+ -	(132422)
Exhibitions (events)	+ -	(81197)
Folders	+ -	(78348)
Holotypes	+ -	(64708)
Isotypes	+ -	(50860)
Landscapes	+ -	(180118)

[see all](#)



Smithsonian
Institution

Date Range Processing

Data pattern matching and date calculation

Dates are transformed into standard date ranges:

Decades: 1500-present,

Centuries: 0 – 1500,

Millennia: BCE.

Examples:

1945 >> 1940s,

195x >> 1950s

1865-1890 >> 1860s, 1870s and 1890

20th Century >> 1900s, 1910s, 1920s, ... 1990s

January 25th, 1877 >> 1870s

Yuan dynasty (1279 - 1368) >> 1200s, 1300s,

ca. 1756 >> 1750s,





Freetext Processing grouped into a JSONObject

<freetext>

```

<dataSource label="Data Source">National Anthropological Archives</dataSource>
<name label="Creator">ANONYMOUS</name>
<name label="Subject">Pu Yi, Henry President (1932-1934) and emperor (1934-1945) of Manchoukuo</name>
<physicalDescription label="Physical description">1 page 43 cm x 151.1 cm</physicalDescription>
<setName label="See more items in">Numbered manuscripts 1850s-1980s (some earlier)</setName>
<date label="Date">1933?</date>
<place label="Place">Korea</place>
<place label="Place">Manchuria</place>
<notes label="Notes">The Korea and Manchuria Tour Guide Office (Sem-Man Annaijo) was a Japanese travel agency with offices in Tokyo, Osaka, and Shimonoseki</notes>
<notes label="Summary">The item is a piece of artwork that presents a bird's-eye view of Korea, Manchuria, and nearby regions and includes railroad routes and waterways. It is illustrated to show people, buildings and habitations, domesticated flora and fauna, and modes of transportation. Also included is a likeness of the Manchu emperor Pu-yi and a man hailing the flag of Manchoukuo. Annotations, perhaps added by the collector, provide romanized versions of some place names and the title "Manchoukuo & Korea" on the mount</notes>
<notes label="Cite as">Manuscript 7144, National Anthropological Archives, Smithsonian Institution</notes>
<objectType label="Type">Maps</objectType>
<objectType label="Type">Collection descriptions</objectType>
<objectType label="Type">Maps, Pictorial</objectType>
<objectType label="Type">Posters</objectType>
<identifier label="Local number">NAA MS 7144</identifier>
<identifier label="Local number">NAA INV 10000000</identifier>
<culture label="Culture">Japanese</culture>
<culture label="Culture">Manchu</culture>

```

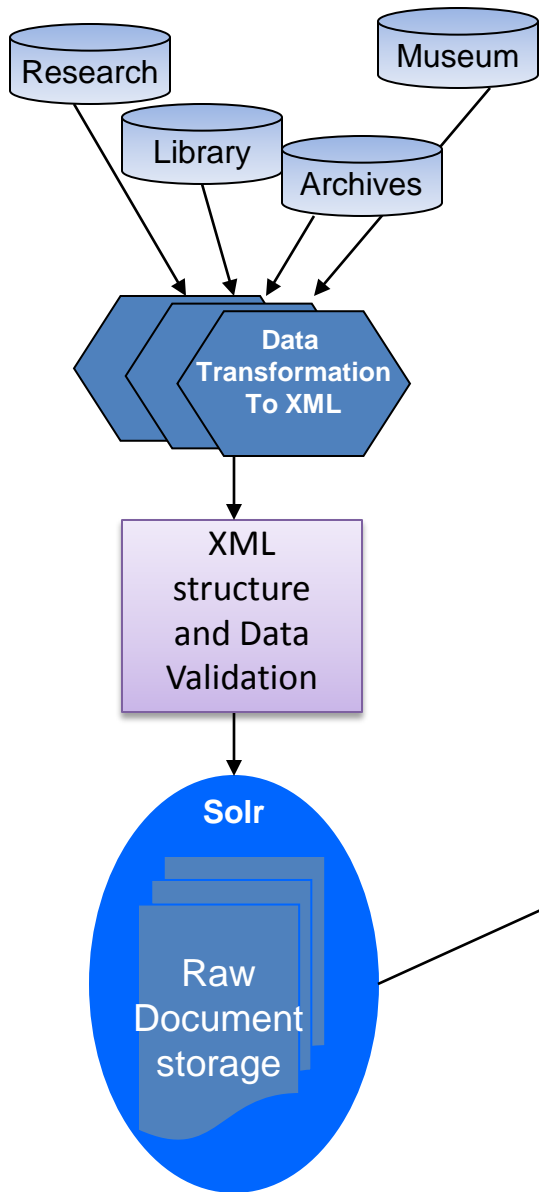


</freetext>

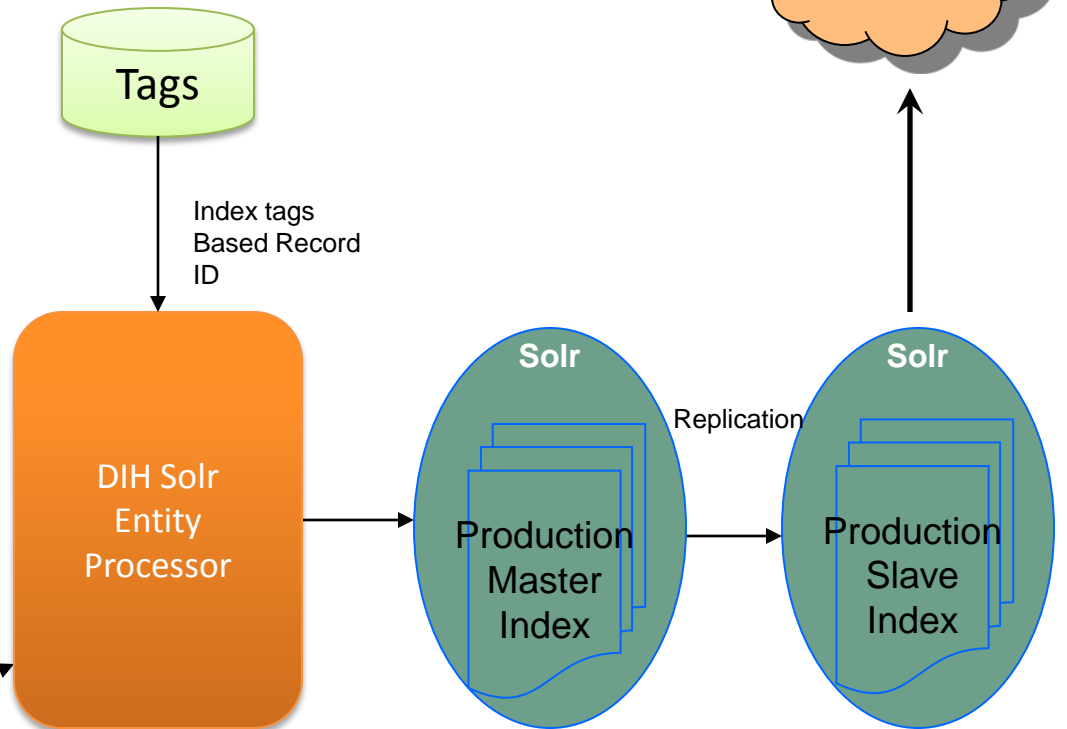


Smithsonian
Institution

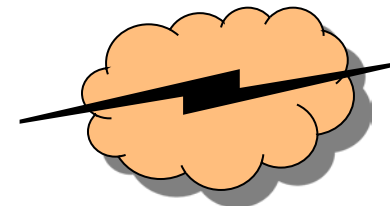
30
databases



Activity Workflow



The World Wide Web



What we are working on now

- Include more data from more databases
 - Department of Invertebrate Zoology: 918,568 records
 - Department of Mineral Sciences: 383,812 records
 - Department of Paleobiology: 589,696 records
 - Smithsonian Tropical Research Institute 150,000 records
 - American History Museum, 860,000 records
- Improve data quality
 - Clean up data in Name field
 - Enhance geographic data
- Implement map displays
 - Too much data is not a good thing
- Experiment with payloads
 - Make better distinction of our data for better searching



Question?

<http://Collections.si.edu>

Smithsonian Institution