

Linked Data for researchers, YCBA Pilot Project



Lec Maj

Head of Information Technology & Innovation

Yale Center for British Art (YCBA)

lec.maj@yale.edu

Who and what are we working on?

- Began with access to online collections via website [2010]
- Extended XML data provider, [2011] data used by:



- In 2012 began LOD work, connecting with other repository data sets in a pilot project.



What is going on with this data?

Google hadleigh castle john constable Lec Maj

Web Images Maps Shopping More Search tools

John Constable artwork

The Hay Wain 1821

The Cornfield 1826

Salisbury Cathedral... 1829

Salisbury Cathedral... 1823


Hadleigh Castle 1829

Hadleigh Castle (painting) - Wikipedia, the free encyclopedia
[en.wikipedia.org/wiki/Hadleigh_Castle_\(painting\)](http://en.wikipedia.org/wiki/Hadleigh_Castle_(painting))
John Constable visited **Hadleigh Castle** in 1814 and made a minute drawing of the castle as preparation for ten oil sketches and a single painting. The oil ...
You've visited this page 2 times. Last visit: 2 months ago.

'Sketch for 'Hadleigh Castle', John Constable | Tate
www.tate.org.uk/.../constable-sketch-for-hadleigh-castle-n048...
John Constable Sketch for 'Hadleigh Castle' c.1828-9. John Constable, 'Sketch for 'Hadleigh Castle' c.1828-9 · Full screen · License this image. John Constable ...
You've visited this page 2 times. Last visit: 3 months ago.

John Constable and Hadleigh Castle - Hadleigh & Thundersley ...
www.hadleighhistory.org.uk/page_id__159_path__0p13p2p2...
12 Oct 2011 – 'John Constable and Hadleigh Castle' by Chris Worpole from Hadleigh History.

Hadleigh Castle
Painting



Hadleigh Castle is an oil painting by the English painter John Constable. John Constable visited Hadleigh Castle in 1814 and made a minute drawing of the castle as preparation for ten oil sketches and a single painting. Wikipedia

Artist: John Constable
Location: Yale Center for British Art
Created: 1829

Answer: it requires domain expert

TATE

Tate Britain

Tate Modern

Tate Liverpool

Tate St Ives

Search



Home Art & artists Visit What's on Context & comment Join & support Learn Research About Shop

Find art and artists

Artist, title, subject, medium etc



John Constable

Sketch for 'Hadleigh Castle' c.1828-9



Share this artwork

YALE CENTER FOR BRITISH ART

Search this site



About Us | News | Get Involved

Collections Exhibitions Events Research Education Architecture Visiting

Home -> Collections -> Search -> Results -> Hadleigh Castle

Search Details

Creator John Constable, 1776-1837, British

Title Hadleigh Castle

Alternate Titles Sketch of Hadleigh Castle
Verso: Study of Five Horned Cattle

Date 1828 to 1829

Medium Oil on millboard

Dimensions 7 7/8 x 9 7/16 inches (20 x 24 cm)

Credit Line Yale Center for British Art, Paul Mellon Collection

Accession Number B2001.2.141

Collection Paintings and Sculpture

Link to This Record <http://collections.britishart.yale.edu/vufind/Record/1668069>

« Prev #2 of 8 Next »



cropped to image, recto, unframed

Public Domain | Download

- Image from Tate
- Data from YCBA
- Different objects
- BM, V&A, etc. have related objects as well

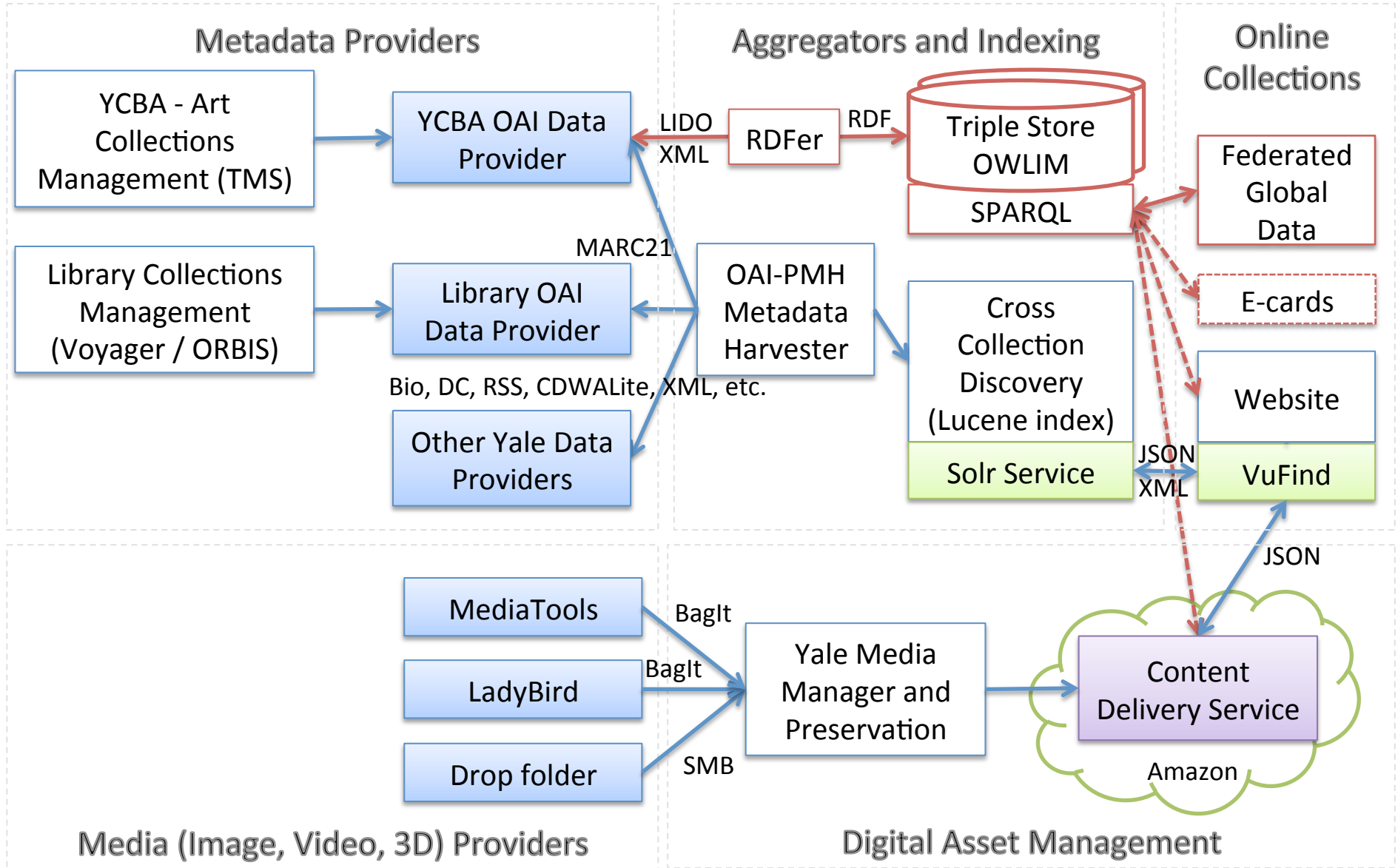
Data for Research

- Answer in-depth questions
- Refine and collect better data
- Current research
 - Spreadsheets
 - Paper print outs
 - Google search
 - Image search
- Research Questions
 - What related types of objects are in existence
 - What drawings were created and owned by Paul Sandby (analysis of image markings to identify marks on them)
 - History map of ownership migration for Sandby's drawings
 - Compare compositions of water colors to other objects to help speculate artistic models (how was the work influenced, how it is treated by various artists)
- Tools
 - Need visual tools for domain experts to analyze data
 - Combination of maps, timelines, etc. without technical expertise and complicated tools



<http://collection.britishart.vale.edu/id/object/178>

Architecture Integration



Data



<http://collection.britishart.vale.edu/id/object/57163>

- Schema and Data Complexity:
 - XML all kinds of schemas / vocabularies (12 mapped over 150 fields)
 - In pilot project CIDOC-CRM Ontology (Conceptual Reference, no limit to describing data)
- Scale (at the moment metadata only):
 - Pilot Project: YCBA (100K XML objects \approx 20M RDF triples)
 - Assuming current mapping of 1 XML object \approx 200 RDF triples
 - 3 Museums (2-3M XML objects \approx 400-600M RDF triples)
 - Painting, Prints, Drawings, Books, Rare Books, Sculptures, Bones, Cake...
 - Professional Schools (1M XML objects \approx 200M RDF triples)
 - Law, SOM, Drama, Music...
 - 21 Libraries (16M XML objects \approx 3.2Billion RDF triples)
 - Book, Rare Books, Maps, Sketchbooks, Journals...
- Type of queries we expect:
 - NLP, multi-join, federated (in/out of cloud)
 - in-depth research oriented, visualization of data for domains (maps, timelines,...)

Key Challenges

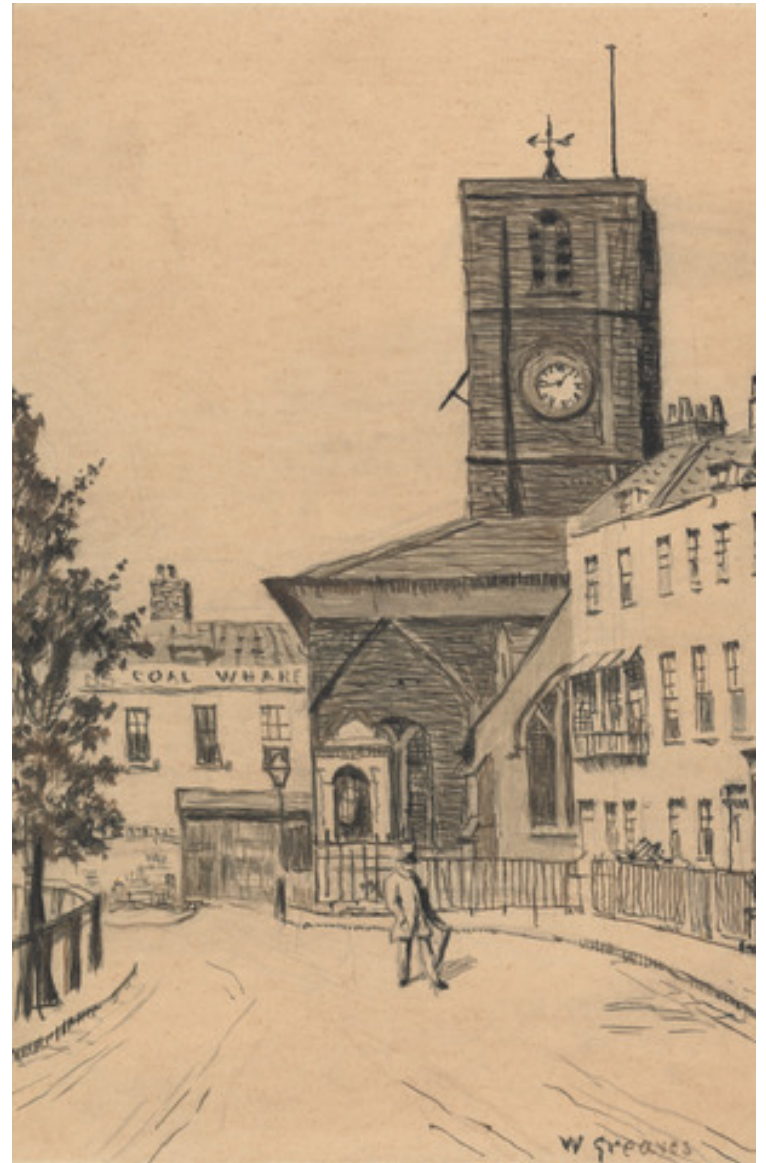
- Loading data without running out of time or computing resources (memory)
 - RDF triples updated (serialized) slow
 - Parallel asynchronous updates (multi-threaded)
- Need to know what is happening under the hood
 - Indexing / distributed processing
 - Commits
- Queries
 - Federation across global repositories
 - Dynamic Ontology Matching
- Analyzing natural language questions for translation to SPARQL, voice Q/A
- 3rd Party / Partner real-time data integration



<http://collection.britishart.yale.edu/id/object/236>

Limitations (computing vs. time)

- Performance Expectation
 - Query: Google like performance (3 seconds at most to process queries)
 - Loading data, cannot take longer than next load cycle
 - Offload whatever takes too long (multi-thread, cluster, cloud)
- Functionality
 - Multi-threaded and parallel data loading from cloud (ie: 100 servers)
 - Local indexing vs. in-cloud distributed indexing (ran into problems with CCD)
- Interoperability
 - Federation support across product and other products
 - URI mapping / lookups
 - Various ontology mapping (CRM, EDM,...)



<http://collection.britishart.yale.edu/id/object/10386>

What to expect from benchmarks



<http://collection.britishart.yale.edu/id/object/37735>

- Configuration recommendations
- Processing benchmarks
 - Load: triples per sec
 - Inference
 - Indexing
 - Poor query examples and timings (# of joins)
- Scale and hardware
 - Physical vs. VM vs. Cloud
- At what point will system start breaking down
- Performance measurement tool to diagnose problems in real time