

Digital scholarship and the cloud

Dr Adam Farquhar
Head of Digital Scholarship
The British Library

britishlibrary.typepad.co.uk/digital-scholarship
IDCC 2014-02-27, San Francisco



150 Million Items



Diamond Sutra

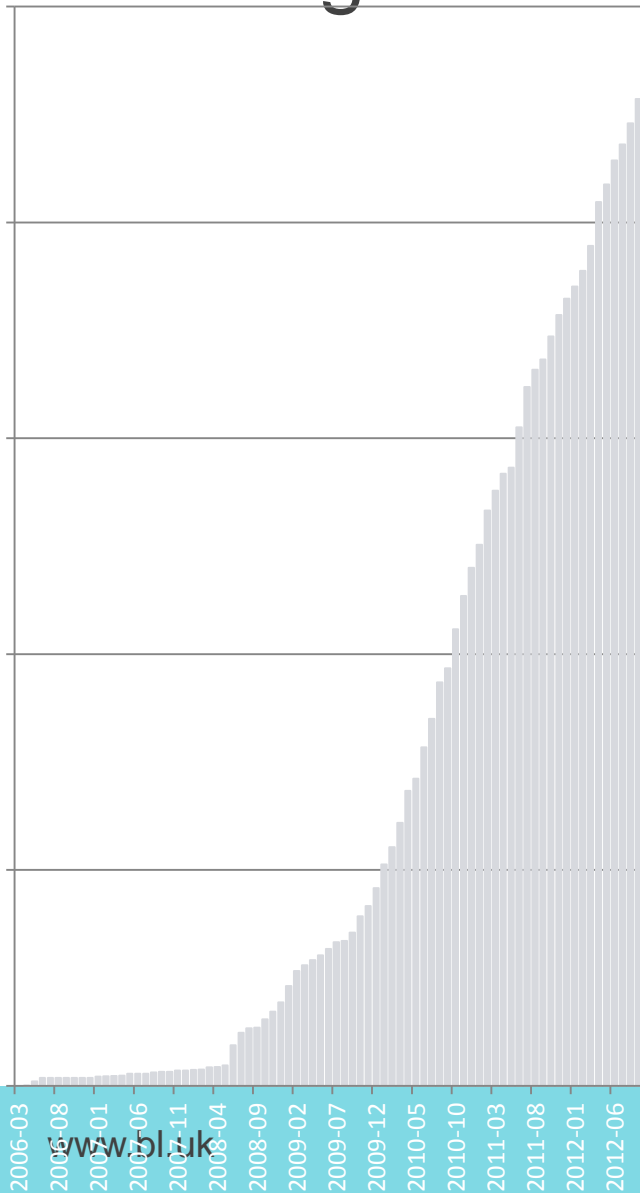
868AD

Blog 2011

Clearly this was the time to coax Marty to recall his Beat days in North Beach, and that's just what I did. Marty said he'd hung out with Kerouac, Boston poet Bob Donlin, disc jockey Pat Donovan, Kaufman, Neal Cassady, a guy named Nick Panopolous, and others at some woman's apartment up on Telegraph Hill. According to the notes I took, her name was "Viasta," and her friend, a pianist named Baker, used to stop by and play a lot. Marty said Jack would smoke pot and drink with them—"he'd partake but he wouldn't participate." Jack seemed square to Marty, was always putting them and himself down—had a gloomy outlook on what they were doing that Marty found disturbing. Jack didn't really seem to be enjoying himself, Marty said, and would always read them the *Diamond Sutra* or his own *Scripture of the Golden Eternity* when what they really wanted was just to get high. Neal was the only one there Marty was really close to at that time (though later he became quite close to Kaufman). Marty said he appreciated and harmonized with Neal's spontaneity—the living and digging and having kicks just for their own sake, not for the sake of some future book about them.

But Marty said something else that night that struck me as a little strange. Maybe with his eye on Kaufman, always the center of everyone's admiration (if not adulation), Marty remarked: "I'm the only great poet from that period who didn't get famous—all because I ran off to Mexico in 1957, just as the news media were discovering the Beats." I detected a little hint of envy and disappointment in Marty's voice, and I realized his life was not quite as simple as he portrayed it—that he shared a lot of the same conflicts the rest of us who emerged from that Beat scene did. Marty did love and live for the moment, just as his friend Neal Cassady had; but

Growing content



But still chained



Texts, corpora, scale

*“Literary scholars and historians have in the past been limited in their analyses of print culture by the constraints of physical archives and human capacity. **A lone scholar cannot read, much less make sense of, millions of newspaper pages.** With the aid of computational linguistics tools and digitized corpora, however, we are working toward a large-scale, systemic understanding of how texts were valued and transmitted during this period”*

David A. Smith, Ryan Cordell, and Elizabeth Maddock Dillon, ‘Infectious Texts: Modeling Text Reuse in Nineteenth-Century Newspapers’ (2013)
<http://www.ccs.neu.edu/home/dasmith/infect-bighum-2013.pdf>

“Reading individual works is as irrelevant as describing the architecture of a building from a single brick, or the layout of a city from a single church”

Franco Moretti, Stanford

Transformations in research and scholarship

Digital Scholarship

The production, use and integration of digital content, services and tools to facilitate scholarship and research

Allows research questions to be investigated in new ways

Using new tools and methods

Leading to new discoveries

Generating new understanding

Across many disciplines

Needed change

1. Digital content
2. Tools and services
3. Methods
4. Engagement
5. Capacity

So what now?

- Libraries and archives have spent the last two decades creating digital assets through digitisation and preserving born-digital objects.
- We can use technology to do more than simply discover these digital objects
- We can embrace the opportunities afforded by analysing these digital collections at scale.
- If scholars view our archives as an infinite pool of multiple layers of loosely held data from which new research questions can be wrung then so must we.
- The Library needs to provide services beyond simple resource discovery - more than helping a single user find a single item or object via a catalogue

BL Response

- A cross-disciplinary mix of curators, librarians and programmers
- The Digital Research Team explores how digital technologies are re/shaping research and how this informs how the library does its business.
- Encourage and support scholars of all disciplines & developers to do innovative research with and across the library's diverse digital content.



Aquiles
Alencar-Brayner



James Baker



Mahendra Mahey



Nora McGregor



Ben O'Steen



Stella Wisdom

Digital Scholarship Training Programme

It is about helping the British Library acclimatise to the idea that the Library is becoming a place full of data as much as it is a place full of physical stuff, and that there is a growing community of users who see it that way.



British Library Labs

Experiment with our digital collections

<http://labs.bl.uk>

British Library Labs 2014 Competition

Calling all researchers, scholars, trailblazers and software developers!

We want you to propose an innovative and transformative project idea by **22 April** that uses the British Library's [digital collections](#).

Two entries will be announced on **23 May**

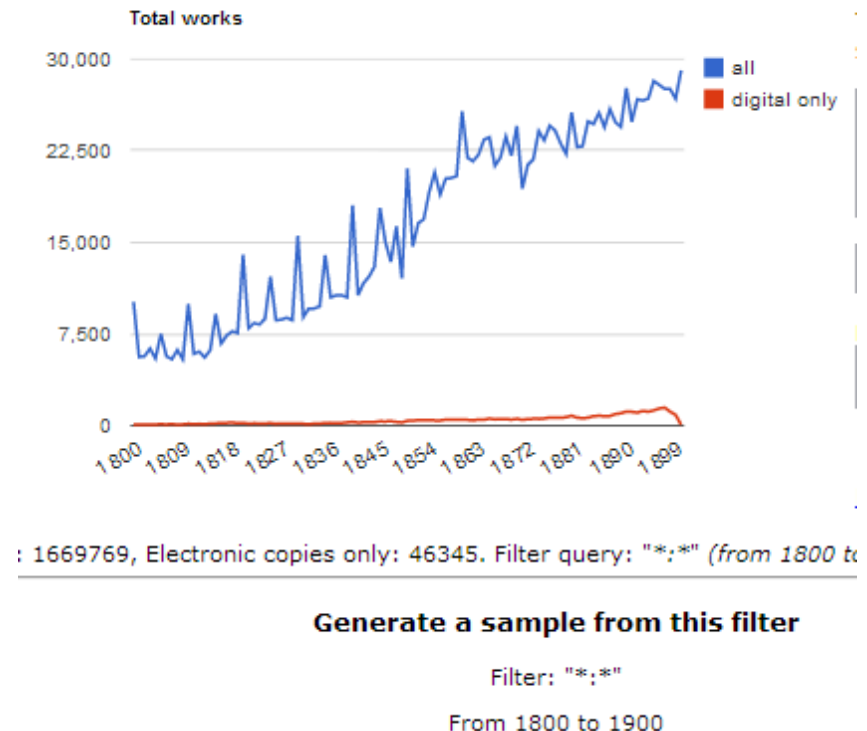
Winners have the chance to work in 'residence' (expenses paid) on their project with the Labs team between **April and Nov.**

The work will be showcased at an event in November / December, with a **first prize of £3,000 and second prize of £1000.**



Labs 2013 winner

- Pieter Francois - postdoctoral researcher at the University of Oxford
- Focus on European travel in the 19th Century
- Uses statistical methods to support text analysis
- Tool produces representative samples of texts based on search criteria



<http://samplegenerator.cloudapp.net>

Off the Map Competition



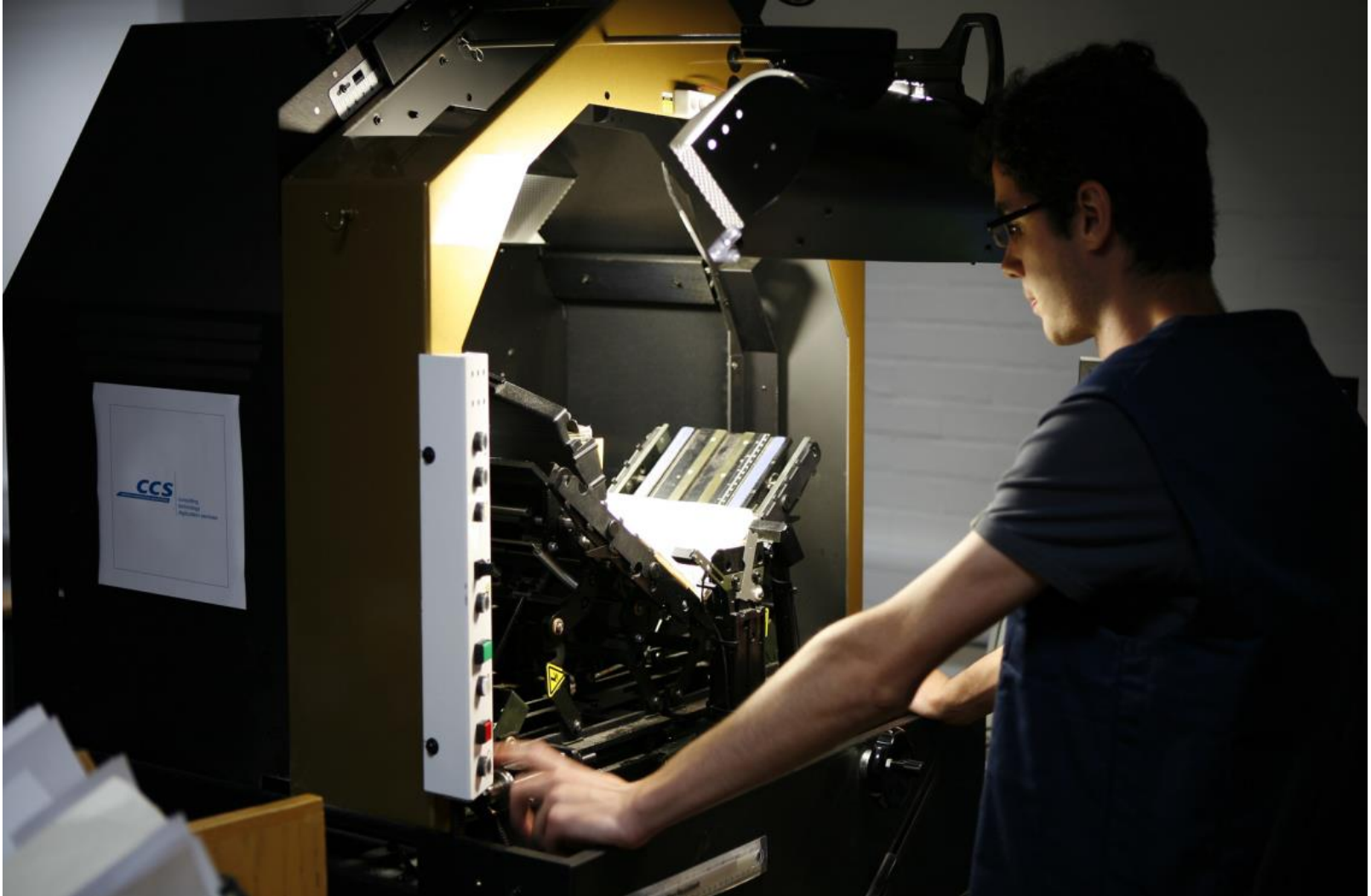
A challenge for videogame design students to turn historic maps and engravings from the British Library collections into a 3D environment using Crytek's CRYENGINE software.

Pudding Lane Productions, a team of six second-year students from De Montfort University, Leicester, won first prize.

Their work was showcased at GameCity, an annual festival of videogame culture.

<http://youtu.be/SPY-hr-8-M0> (Flythrough starts at 0:50)

The Mechanical Curator



Extracting Images from OCR

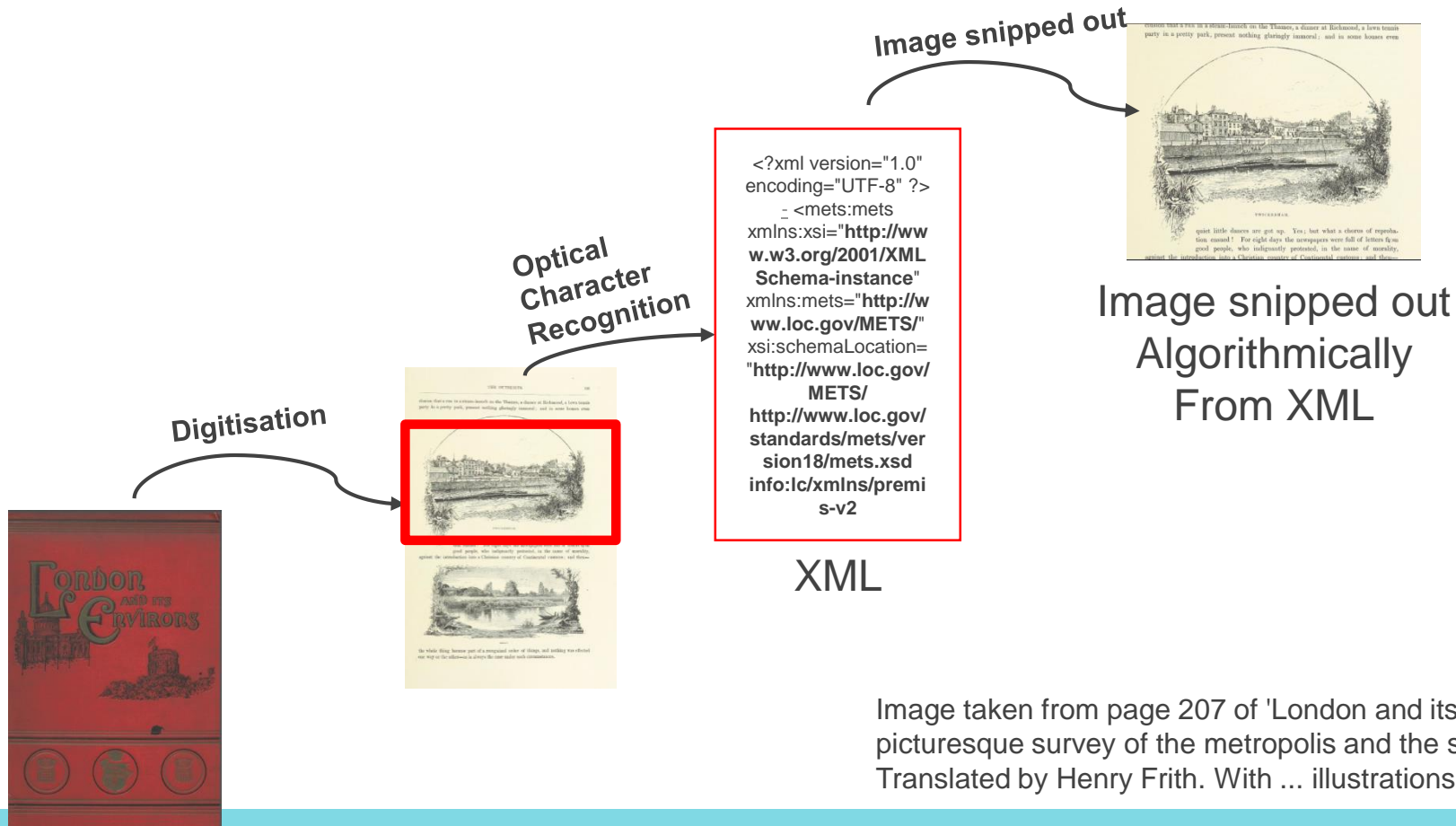
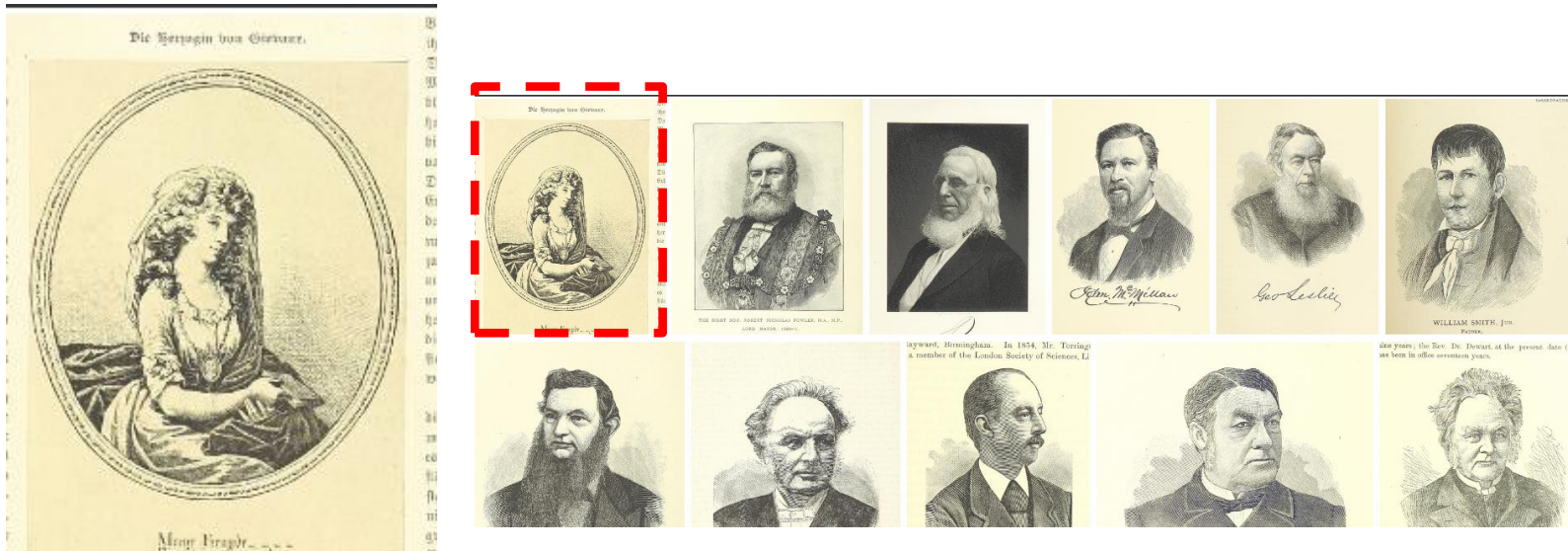


Image taken from page 207 of 'London and its Environs. A picturesque survey of the metropolis and the suburbs ... Translated by Henry Frith. With ... illustrations'

Face Recognition of 19th Century Faces



The face-recognition algorithm worked better for female faces than men's

The Mechanical Curator

Randomly selected small illustrations and ornamentations, posted on the hour.

Rediscovered artwork from the pages of 17th, 18th and 19th Century books.

[@MechCuratorBot](#) - [British Library Digital Scholarship Blog](#)

[Archive](#) [About](#) [Links](#)



Image from 'Rose, Shamrock, and Thistle. A story of two Border towers', [001957329](#)

- **Author:** KETTLE, Mary Rosa Stuart.
- **Page:** 190
- **Year:** 1893
- **Place:** London
- **Publisher:** T. Fisher Unwin

The British Library

16
SetsAugust 200
Member Sir

Photostream

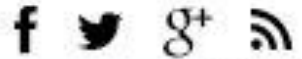
Sets

Favorites

Creations

NEW





[Login/Register](#) | [Newsletter](#) | [Subscribe](#)

GO

[GALLERIES](#) /// [VIDEOS](#) /// [COLUMNS](#)

[POPULAR](#) ▶ [2013 Invention Awards](#) | [Magazine](#)

Automated Program Culls Inexplicable Illustrations From Antique Books

Say hello to the British Library's Mechanical Curator.

By Francie Diep Posted 10.09.2013 at 2:30 pm



quarter, though Justice Goe and Miller were able to contribute one dollar, eighty-seven and one-half cents. For a hi-tory of the depreci-

"Tl
Virgin



the Appendix

The Appendix
Blog

[Current Issue](#) [Archives](#) [Blog](#) [Contributors](#) [Subscribe](#) [Shop](#) [About](#)



The British Library Just Made One Million Images Public Domain, and It's a Big Deal

Posted by [Benjamin Breen](#) on December 15, 2013

The good people at the British Library have been killing it lately.

First, their Medieval Manuscripts blog brought us the phenomenon of [knights jousting with snails](#), which was deservedly picked up by [the Guardian](#), among

[Benjamin Breen](#) is the editor-in-chief of *The Appendix* and a PhD candidate in history at the University of Texas at Austin.

Dora's lost data



"Hi, my name is Dora, and I'm a junior curator. It's my first day and I've made a big mistake – I accidentally deleted all the information we were going to add to our collections online. I need to re-label them, and quickly...

Can you help? **Add words about the thing in the picture that would help someone find it on**

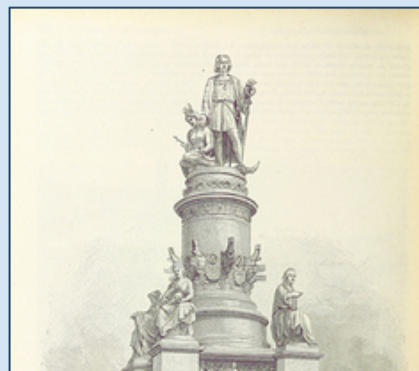
Google – how it looks, what it does, who might have used it – anything you can think of."

MARINELLI, Giova

La Terra, trattato popolare di geografia universale per G. Marinelli ed altri scienziati italiani, etc. [With illustrati

View [object on the British Library on Flickr Commons site](#) (opens in new window).

Date: 1898 Place: Milan (Accession num: 2383118)

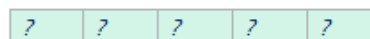


Your score for this game:

175 points.

Login or register to save your points

Objects you've played in this game



So far players like you have improved **594 records** for **3 museums and libraries** through games on this site.

Top registered players (all games)

Han (5620 points)
Josie (5000 points)
IrnC (4400 points)
mia (3895 points)
Hepburn (2190 points)
ad (1785 points)
archiwicz (1225 points)
miatepink (1115 points)
JoK (890 points)
jennifuchs (870 points)

Register or login to save your points

Username

Password

Mechanical Curator In Numbers

>120 million image views
since launch on 13th
December 2013

>50,000 tags added

>30,000 images favourited

Users have uncovered:

>2000 maps

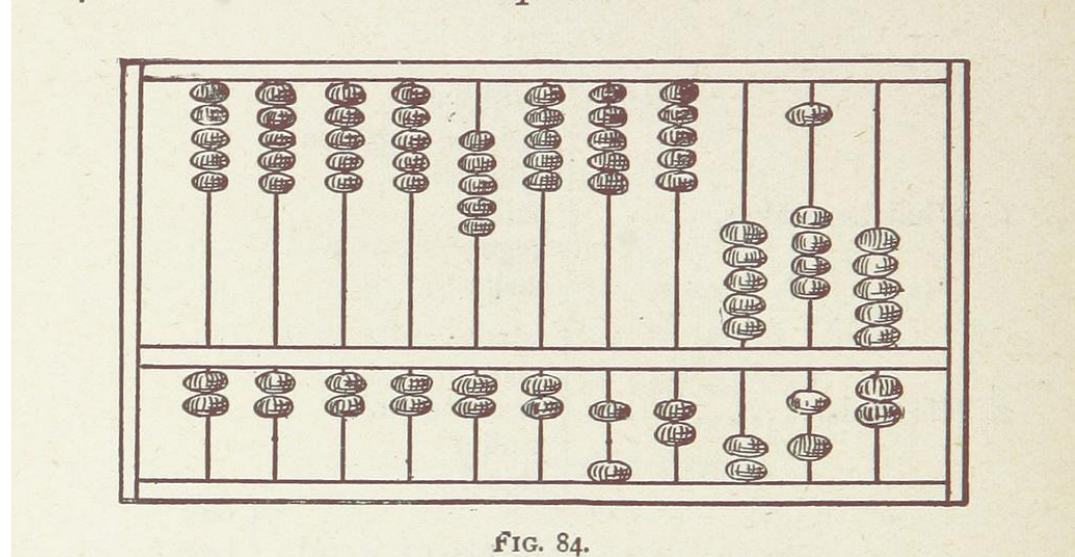
>4000 portraits

>250 ships

>750 illustrated book covers

4 cats

and 2 libraries



Top Favourites:

56 <http://flickr.com/photos/britishlibrary/11139197046>

55 <http://flickr.com/photos/britishlibrary/11307051005>

52 <http://flickr.com/photos/britishlibrary/11307158676>

44 <http://flickr.com/photos/britishlibrary/11307140236>

42 <http://flickr.com/photos/britishlibrary/11307137526>



Creative Uses

<http://blpublicdomain.wikispaces.com/Creative+Projects>



Benefits

- Robust commercial-grade hardware and networking
- Scalable storage and computation
- Simple configuration and use
- Isolation from in-house IT infrastructure
- Simpler data transfer
- Ease of experimentation
- Pathway to production

Fits to five things researchers want

- Focus on my research
 - Feel at home
 - Bring my own tools
 - Be creative
 - Start small - but finish big
-
- There is still work for infrastructure providers to do!

Challenges

- User identification, authentication and authorisation
- Managing a library of virtual machines
- Managing access to content
- Hybrid models for complex licensing, ownership, copyright
- Pricing and business models
- Packaging content for researchers
- Dataset identification and citation

More than a fad...

“The emergence of the new digital humanities isn’t an isolated academic phenomenon. The institutional and disciplinary changes are part of a larger cultural shift, inside and outside the academy, a rapid cycle of emergence and convergence in technology and culture”

Steven E Jones, *Emergence of the Digital Humanities* (2013)

<http://emergenceofdhbook.tumblr.com/>

And the cloud is key to providing scholars, researchers, and the public with the resources they need

Thanks

Discussion

Remember: Labs 2014
submissions by **22 April**

 #bldigital

<http://britishlibrary.typepad.co.uk/digital-scholarship>

Thank you!

Your are free to:

- Copy, share, adapt, or re-mix
- Photograph, film, or broadcast
- Blog, live-blog, or post video of;

this presentation provided that:

- You attribute the work to its author and respect the rights and licences associated with its components



Text attribution Greg Wilson, Two Solitudes, SPLASH 2013 (29 October 2013)
<http://www.slideshare.net/gvwilson/splash-2013>



This work is licensed under a
Creative Commons Attribution 3.0
Unported License unless stated
otherwise.