human-computer collaboration

mauricio giraldo arteaga
@mgiraldo
@nypl_labs

IPAM Culture Analytics and User Experience Design, April 2016
hello
not a real library scientist
digitization  metadata  access

public

traditional digital library program
access
digitization
metadata
access

digitization
metadata
access

r+d

public

public
engagement

engagement
Public Domain Collections: Free to Share & Reuse

Did you know that more than 180,600 of the items in our Digital Collections are in the public domain?

That means everyone has the freedom to enjoy and use these materials in almost limitless ways. The Library now makes it possible to download each item in the highest resolution available directly from the Digital Collections website.

Search Digital Collections

No permission required. No restrictions on use.

Below you’ll find tools, projects, and explorations designed to inspire your own creations—go forth and respect!

Visualize the Public Domain

An experiment by NYPL Labs to help patrons understand and explore what is contained in this release. Browse by century, genre, collection, or color.

Discover the Collections

Learn more about our public domain releases. Our blog post series will talk about what it means, and where you can go with these materials.

Apply for the Remix Residency

To promote transformative, interesting, and creative new uses of our Digital Collections and data.

Use Our Data and Utilities

Our digitized collections are available as machine-readable data over one million records for you to search.
Explore 679,265 items digitized from The New York Public Library's collections.

This site is a living database with new materials added every day, featuring prints, photographs, maps, manuscripts, streaming video, and more.

Search keyword(s)

Featured Collections:
- Jerome Robbins Dance Division Moving Image Archive
- American Jewish Committee Oral History Collection

Featured: Messenger boy, Street view of New York City.
what happens after digitization?
human-computer collaboration
embrace imperfection

corollary of “perfect is the enemy of good”
A designer’s definition of ‘perfect’ is different for computational designers.  

John Maeda

because it is not achievable
human-computer collaboration
computers are good at some things...
people overestimate OCR quality
Jan 17 1900
Gerard Hotel
Luncheon

Cham Brthk. 20.50
Cream of Oysters 20-30
Beef Tea 20-30

Mincey Habitut au Gratin et Croquaille 80.09

Beef Steak Pie Family Style 30.38
Figa Beet Fried in Batter 80.40

Cold Roast Beef 80-80 Cold Canada Mutton 80-80
Havel Corned Beef 80-80

Mixed Baba 25 Bread Tomatoes 25

Yima Beans 25 Buttered Beets 25

Baked Pataces 25

Current Buns 25 Baked Apple Cream 25

Pudding Cake 25 Preserved Apricots 25

Baked Lemon Brown Sugar 15 Grape Fruit 20

Gumdrop 25 Orange 25

Pudding Cake 15 Bananas 25

Citron Cake 15 French Vanilla Ice Cream 25

Clam Cake 15 French Biscuit Tortini 25

"ma vii"

Q<...... V ," 4
~ 5
r,
muck,
Q '<<'
1 :Maw
"ma vii"
okay... so maybe computers are not that good
people are good at other things
human-computer collaboration

I avoid the term "crowdsourcing"
two examples
like Google Maps for the 19th century

but Google Maps cannot answer questions about the 19th century
*this is a simulation. actual process is intensive. consult your mathematician before trying
and now you start tracing those buildings by hand
1852-1854
1852-1854
can we automate this?
computers are good at some things...
Historical map polygon and feature extractor

Mauricio Giraldo-Arango
New York Public Library
mgiraldo@nypl.org

ABSTRACT

Polygon map analysis data extraction from historical maps such as US Census maps from the 1800s can readily remain as data on the New York Public Library (NYPL) online viewing site and therefore need to be efficiently extracted. New tools designed for polygon map analysis on the NYPL's library site were required to provide these polygon data. Due to the requirements of polygon analysis, the approach taken was one of enhancing map portions to provide the polygons needed in a usable format for polygons.

On average, a single US site area is 1,100 acres and a total of 32 areas of mapping across the US in 1760 of the publicly available site areas total 31,692 acres. The polygons are extracted from the NYPL's online library site to help identify the coverage of the land poly in a usable format for polygons.

From the maps, a map to site polygons was the primary approach. The next implementation was the creation of polygons from the site areas and this was done to help identify the coverage of the land polygons in the site areas.

Categories and Subject Descriptors

Theory of Computation: Computational Geometry and Object

Keywords: map polygons, feature polygons, data polygons, polygon analysis

1. INTRODUCTION

Maps provide data for analysis of polygonal maps such as US Census maps from the 1800s can readily remain as data on the New York Public Library (NYPL) online viewing site and therefore need to be efficiently extracted. Due to the requirements of polygon analysis, the approach taken was one of enhancing map portions to provide the polygons needed in a usable format for polygons.
yay footprints!

60,000+ of those!
like OCR for maps!™

(not really trademarked)
but OCR is pretty bad
people are good at other things!
people don’t choose to complete these
we have over 60,000 footprints to check!

will people want to do this?
what is the minimum contribution we need?

we want the lowest friction possible so people will want to contribute
this was 2013, touch-screen mobile had taken off
what about malicious users?

or even well-meaning ones who make mistakes
75% or more agreement between 3 or more people

arbitrary numbers that have worked for us
YES is on the right side because most people are right-handed and the algorithm is right most of the time.
will people participate?

remember that little tweet button?
3499 buildings checked! Data mining old Maps with the Building Inspector from @NYPLMaps @nypl_labs
27702 buildings checked! Data mining old maps with the Building Inspector from @NYPLMaps @nypl_labs
buildinginspector.nypl.org
Mark my words, “Building Inspector” will be the next Flappy Bird.
buildinginspector.nypl.org openvisconf
@mgiraldo
*Footprints marked as “NO” go to polygon heaven*
address

had to use full keyboard on mobile because fractions
classify
Building Inspector

The New York Public Library presents

BUILDING INSPECTOR
Kill Time. Make History.

Welcome citizen cartographers! Help unlock New York City’s past by identifying buildings and other details on beautiful old maps.

NEW MAPS!
GET STARTED

HOW YOU CAN HELP
The Library is training computers to recognize building shapes and other data on digitized insurance atlases. Via these easy, bite-sized tasks, you can help check the computers’ work and capture other valuable information.
we add new maps as old ones are completed

the bottleneck now became geo-rectifying those maps ـ_(ツ)_/~
this is actually version 2
HELP TRACE NEW YORK CITY'S PAST

The New York Public Library is training computers how to recognize building shapes and other information from old city maps. Help us clean up the data so that it can be used in research, teaching and civic hacking.


 Privacy Policy - Rules and Regulations - Policy on Patron-Generated Web Content - Terms and Conditions 

A project of NYPL Labs and the Lionel Pincus and Princess Firyal Map Division of The New York Public Library

The New York Public Library

© 2013 NYPL
good tutorials are hard
Here’s how the app works:

We'll show you one computer-generated building outline at a time, overlaid over the original map.

- **NO**: Reject the outline
- **FIX**: Correct the outline
- **YES**: Accept the outline

View progress at the top right.
Super Mario Bros. (Nintendo, 1985)
we have too many edge cases

or: how i learned to stop worrying and embrace imperfection
people skip them anyway
NYPL Community Oral History Project

oralhistory.nypl.org
Visible Lives

Oral Histories of the Disability Experience

Visible Lives: Oral Histories of the Disability Experience is an oral history project that works to both preserve and document a thematic history through personal recollections. This project will collect stories of people who are living with a disability and the Library will train community members to conduct these interviews. Interviews will be shared in a preservation archive at The Milstein Division and on the New York Public Library website. Public programs will also connect neighborhood residents and project participants.

Visible Lives is a project of Andrew Heinsel Braille and Talking Book Library in Manhattan. A public archive will be kept at this local branch for future generations to listen to and research.

For more information about this project or to share YOUR story: Please contact Alexandra Kelly at Outreach Services and Adult Programming, AlexandraKelly@nypl.org or (212) 623-0552.
make these stories more accessible
mark

transcribe
allows for basic text search

but it’s not a proper transcript
Visible Lives
Oral Histories of the Disability Experience

Visible Lives: Oral Histories of the Disability Experience is an oral history project that works to both preserve and document a thematic history through personal recollections. This project will collect stories of people who are living with a disability and the Library will train community members to conduct these interviews. Interviews will be shared in a preservation archive at The Milstein Division and on the New York Public Library website. Public programs will also connect neighborhood residents and project participants.

Visible Lives is a project of Andrew Heiskell Braille and Talking Book Library in Manhattan. A public archive will be kept at this local branch for future generations to listen to and research.

For more information about this project or to share YOUR story: Please contact Alexandra Kelly at Outreach Services and Adult Programming, AlexandraKelly@nypl.org or (212) 621-0552.
we felt we needed something different
computers are good at some things...
like OCR for audio!™

(not sure if they trademarked that)
we get transcription “snippets” from 1 to about 6 seconds long in varying levels of quality
<table>
<thead>
<tr>
<th>start_time</th>
<th>sequence</th>
<th>original_text</th>
</tr>
</thead>
<tbody>
<tr>
<td>4077</td>
<td>0</td>
<td>So many Muslims out and going to be having a conversation with the ones of you know</td>
</tr>
<tr>
<td>9008</td>
<td>1</td>
<td>for visible lives and history project at in the library</td>
</tr>
<tr>
<td>13107</td>
<td>2</td>
<td>. Today is December second twenty fourteen.</td>
</tr>
<tr>
<td>17948</td>
<td>3</td>
<td>And. This is being recorded at into high school library</td>
</tr>
<tr>
<td>27957</td>
<td>5</td>
<td>Tell me something about growing up. Sure where you come from.</td>
</tr>
<tr>
<td>31267</td>
<td>6</td>
<td>Sure so I was born in Plainfield New Jersey</td>
</tr>
<tr>
<td>36538</td>
<td>7</td>
<td>which is about twenty miles west of New York City</td>
</tr>
<tr>
<td>41148</td>
<td>8</td>
<td>almost New Jerseyites or New Jerseyans say that there are two</td>
</tr>
<tr>
<td>46127</td>
<td>9</td>
<td>parts of Jersey north and south. But. Some</td>
</tr>
<tr>
<td>50367</td>
<td>10</td>
<td>. Some folks will argue there's something called Central</td>
</tr>
<tr>
<td>55307</td>
<td>11</td>
<td>Jersey but usually those people that don't come from New Jersey. So</td>
</tr>
<tr>
<td>60987</td>
<td>12</td>
<td>North Jersey where I grew up when I was first where I was</td>
</tr>
<tr>
<td>65227</td>
<td>13</td>
<td>born we lived actually in playfields in a</td>
</tr>
<tr>
<td>70138</td>
<td>14</td>
<td>little apartment my mother and father and my first our first dog to a little</td>
</tr>
<tr>
<td>74897</td>
<td>15</td>
<td>poodle who I swear I remember my mother Oh he said I only remember it from</td>
</tr>
<tr>
<td>79608</td>
<td>16</td>
<td>pictures because I was so little I couldn't possibly remember the dog.</td>
</tr>
<tr>
<td>85347</td>
<td>17</td>
<td>But after you know after a while we move to a bigger house in a town called</td>
</tr>
<tr>
<td>89998</td>
<td>18</td>
<td>Middlesex this is all around the same sort of part of New Jersey sort of Union</td>
</tr>
<tr>
<td>94737</td>
<td>19</td>
<td>County Somerset County. And then for a while</td>
</tr>
<tr>
<td>99498</td>
<td>20</td>
<td>after that we lived there till I was about two and a half or three because my my</td>
</tr>
<tr>
<td>104267</td>
<td>21</td>
<td>brother Steven was born at that time. When was this when I was born in one thousand</td>
</tr>
<tr>
<td>108918</td>
<td>22</td>
<td>sixty two March one thousand nine hundred sixty two and my brother was born in November of</td>
</tr>
<tr>
<td>113877</td>
<td>23</td>
<td>one thousand sixty four. We moved in with my mother's</td>
</tr>
<tr>
<td>118877</td>
<td>24</td>
<td>mother for a little while while we were while my parents were searching for a</td>
</tr>
<tr>
<td>123388</td>
<td>25</td>
<td>house where at a town where they really wanted to</td>
</tr>
<tr>
<td>127938</td>
<td>26</td>
<td>settle. On and that was in North Plainfield New Jersey where I</td>
</tr>
<tr>
<td>132787</td>
<td>27</td>
<td>lived for I'd say. Really a year he lived with my grandmother</td>
</tr>
</tbody>
</table>
people are good at other things...
You know mining is for me and I'm going to have a conversation one that we saw here that sort of visualize an oral history project and you are public library.

He's the twenty second of May two thousand and fifteen. And this is being recorded at Stuyvesant Town
we conducted a few usability studies
You know mining is for me and I'm going to have a conversation one that we saw here that sort of visualize an oral history project and you are public library.

He's the twenty second of May two thousand and fifteen. And this is being recorded at Stuyvesant Town
and, this is being recorded at the Andrew Heiskell Library.
and this is being recorded at the Andrew Heiskell library.
And this is being recorded at the Andrew Haskell library.
and this is being recorded at the ? High School library.
and this is being recorded at the [?] High School library.
And, this is being recorded at Andrew Highskill library.
And, this is being recorded at the Andrew Highskill Library.
And... this is being recorded at the Andrew Heiskell Library
and, this is being recorded at the high school library.
And. This is being recorded at the [inaudible] library
And. This is being recorded at the Andrew Heiskell library
ugh, Some folks will argue there's something called Central
folks will argue there's something called Central
some folks will argue there's something called Central
%uh, uh some folks will argue there's something called Central Jersey
uh, uh, some folks will argue there's something called Central Jersey.
some folks will argue there's something called Central
uh, some folks will argue there's something called Central Jersey
Some folks will argue there's something called Central
Some folks will argue there's something called Central Jersey
Uh, some folks will argue there's something called Central
%uh, some folks will argue there's something called Central
it’s hard to reach consensus
embrace imperfection
The New York Public Library
Community Oral History Project

Make hundreds of NYC stories accessible -- one word at a time.

Help The New York Public Library fix computer-generated transcripts from hundreds of stories from the library's Community Oral History Project.

An example of how the transcript editor works (click for sound)

Select an interview to get started.

Filter by Collection: All Collections ▼  Sort by: Completeness (most to least) ▼  Search Title/Description ▼
made with customizability in mind
Make hundreds of Moth stories accessible.
One word at a time.

Help The Moth correct computer-generated transcripts from our library of broadcasted stories.
Click on a story to get started!
this is one week after launch
it is still being improved
two of several projects we’ve worked on so far
of human-computer collaboration
it’s a collaborative process

Willa Armstrong, Shawn Averkamp, Paul Beaudoin, Brian Foo, Josh Hadro, Elizabeth Hummer, Ara Kim, Shana Kimball, Tom Listanti, Matthew Miller, Eric Shows, Bert Spaan, and more at NYPL...
one more thing...
"What Big Media Can Learn From the New York Public Library" (article here)

Alexis Madrigal, a senior editor at The Atlantic, mentioned my ongoing Reaching for the Out of Reach project & linked to cursivebuildings.com yesterday in a very interesting article on all the good things NYPL is doing with archives. Rad! So happy to play a part in these important discussions. Back when this project first gained notice on a large scale, the library contacted me with the idea, & to my surprise, were very supportive of the endeavor (& I had exactly asked permission to re-imagine their archived stereographic. It’s great to see them get positive press in response to their initiative on this matter.

ps. Interesting topic! A previous, related post.

Reaching for the Out of Reach 60:
A young man rests at Peru Fork, Colorado, circa 1875.
(mag from this project: mag peru)
how to decode the 3D data?
in the browser
1. Move the line so that it divides the stereograph in two.
2. Press CORRECT ROTATION.
3. Move the rectangles so that they cover each side of the photo.
4. Press ANIMATE.
thank you!

mauricio giraldo arteaga
@mgiraldo
@nypl_labs