Crowdsourcing 101: Fundamentals and Case Studies
October 29, 2014

Presented by:
Crowdsourcing Consortium for Libraries and Museums (CCLA)

crowdconsortium.org
@crowdconsortium
Today’s Presenters

Mia Ridge
Chair of the Museums Computer Group at Open University and a member of the Executive Council of the Association for Computers and the Humanities (ACH)

Ben Vershbow
Director of NYPL Digital Library + Labs

Victoria Van Hyning
Digital Humanities Postdoctoral Fellow, Zooniverse
Crowdsourcing 101: Fundamentals and Case Studies

29 October 2014

Mia Ridge, Open University/Trinity College Dublin
http://miaridge.com/
@mia_out
Overview

• Defining crowdsourcing
• Example projects, typical tasks and input/output content
• Participants, motivations and levels of engagement
• Design tips
What is crowdsourcing?

'the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call’ (Jeff Howe and Mark Robinson for Wired, 2006)

Or, using cognitive surplus: 'the spare processing power of millions of human brains’ (Clay Shirky)
Cultural heritage crowdsourcing is...

...asking the public to undertake meaningful tasks related to cultural heritage collections in an environment where the activities and/or goals provide inherent rewards for participation. The project should contribute to a shared, significant goal or research interest.
...vs 'user-generated content'

- UGC: comments, creative responses or items contributed by the public
- Unlike crowdsourcing, the act of contributing is more valuable than the contribution - individual, not mutual value
Basically, cultural heritage crowdsourcing is...

Transforming input content into output content...

...via a powerful purpose and/or enjoyable tasks that people want to help you with
MISSING FRIENDS.

Circulating as it does throughout the

Southern Hemisphere. "The World's

News" is unique as a medium for the
discovery of missing friends. The

charge for these advertisements is 14

words $1 cash, and thereafter is per

line.

"DUTTENWORTH, WILLIAM, of Hull, England, who
was last heard of in 1882, and was last heard of in
1891, is requested to write to his nephew, J. W. Butten-
worth, 24 Lockwood-terrace, Hull.

GIPIN, JOSEPH, last heard of at Melbourne. Com-
municate with cousin, Mary Jane Killen (now Mrs.
Cooper), 90 Watercroft, Almondbury, Yorkshire, Eng-
land;
2012 Statistics
Total records indexed: 534,108,416
Total records arbitrated: 263,254,447
Total volunteers contributing: 348,796
Total estimated hours contributed: 12,764,859

On “5 Million Name Fame” event day, July 2012:
Indexed Records: 7,258,151
Arbitrated Records: 3,082,728
Total Records Worked: 10,340,879
Volunteers participating: 46,091.
Transcribe Bentham

http://www.transcribe-bentham.da.ulcc.ac.uk/
Metadata Games
Micropasts

http://micropasts.org/
British Library Georeferencer

http://www.bl.uk/maps/
Record Number: 16878

Reading Experience:

Evidence:
'Sunday 3 October. I am reading "A Room of One's Own". Most delightful and profound - if I had the time I would write an essay about life in the WRNS'.

Century: 1900-1945
Date: 3 Oct 1943
Country: England
Time: n/a
Place: n/a
Type of Experience (Reader): silent aloud unknown solitary in company unknown single serial unknown
Type of Experience (Listener): solitary in company unknown single serial unknown

Reader / Listener / Reading Group:

Reader: Barbara Pym
Age: Adult (18-100+)
Gender: Female
Date of Birth: 2 Jun 1913
Socio-Economic Group: Professional / academic / merchant / farmer
Occupation: Writer
Religion: Church of England
Country of Origin: England
Country of Experience: England
Listeners present if any: e.g. family, servants, friends: n/a
Typical tasks in crowdsourcing

- Task granularity: 'microtasks' to long-term, complex tasks
- Tagging (subjective, factual; free-text, vocabularies; mark-up, tags)
- Transcribing (including OCR correction)
- House-keeping (moderating forums, flagging content for review)
- Sharing knowledge (personal or researched)
- Creating links, relationships, categorising
- Stating preferences, opinions
- Crowdfunding
Who participates in crowdsourcing?
'16,400 little boxes – one for each person who’s contributed to oldWeather. The area of each box is proportional to the number of pages transcribed, between us all we’ve done 1,090,745 pages.'
Motivations for participation
One task...

Mr. Lucas seconded the motion. When a work
of such magnitude is brought forward, he thought the Go-
vernment fully justified m\(l\)orrowing money for its com-
pletion. As to the opponents, of, borrowing, he should
like to ask some of them if they knew what three months'
bills meant—(a laugh)—or three months’ bills two or
three times renewed; of their overseers’ drawing orders
upon them from their stations, which was another
form of borrowing, as his pocket could bear testimony;
(He is, hear, and a laugh.) The railway might probably
cost more than was estimated, but at the same time the
railway returns were very much underrated. The produce
from the north was daily in increasing—land was rapidly
growing into cultivation—and it would increase still more
with the construction and extension of a railway. He

The Australian Register, Wednesday, September 12.

SUPREME COURT—CIVIL SIDE.

TUESDAY, SEPTEMBER 12.
[Before Mr. Justice Cooper.]

AYLWIN V. DAY.

Mr. Wearing for the plaintiff, Mr. Gwynne for the de-
defendant.

This was a new trial of an action for £250, obtained, as
alleged, by false pretences, which was decided in favour
of the plaintiff last term. The plaintiff, a new arrival in
the colony when the cause of action occurred, was a wine
merchant in Adelaide, but desired to become a stockholder
and owner of a cattle or sheep-run. He signified his de-
sire to a person named Tyrrell, who undertook to lend
the defendant, who stated that he had a run in the neigh-
bourhood of Mount Allen to dispose of. It was ar-
ranged that Tyrrell should visit the run, and on his
approval the plaintiff was to purchase it for the agreed
price of £250. Tyrrell left in company with the de-
defendant, who was to accompany him to Gawler Town,
from which place he was to proceed to Clare, and there, in
the defendant’s brother’s company, and show them the
run. It appeared, however, that, instead of acting on
that arrangement, Tyrrell accompanied the defendant
to his house on Gawler Plains, where he remained for about
a week, and then returned to Adelaide and reported to
the plaintiff in the presence of the defendant that he had
seen the run and approved of it, stating that the plaintiff
should keep from 10,000 to 15,000 sheep. Upon the faith of
that promise the defendant paid over £100 on account. That was on the 25th De-

The remaining £100 was paid a week after.

The defendant handed over a paper to plaintiff, as follows:

“December 28th, 1853.

“New run applied for by Mr. John Day on the 11th November, 1853. — W. James, pro Commissioner of Crown
Lands.”

The defendant stated at the same time that on presenting
that paper the Crown Lands Commissioner would have no
obligation.
...many motivations for participation

• Altruistic
  – helping to provide an accurate record of local history

• Intrinsic
  – reading 18thC handwriting is an enjoyable puzzle

• Extrinsic
  – an academic collecting a quote from a primary source
Extrinsic motivations

http://gwap.com
Altruism

http://helpfromhome.org/

What is Microvolunteering?

Volunteer your time in bite sized chunks, from your own home and when you want to. Benefit worthy causes to suit your lifestyle at your own convenience. Dip in, dip out with absolutely no commitment. No specialised skills required. It’s all FREE!

And, oh yeah, you can do them in your pyjamas!
Intrinsic motivations

- fun
- the pleasure in doing hobbies
- the enjoyment in learning
- mastering new skills, practicing existing skills
- recognition
- community
- passion for the subject

State Library of Queensland, Australia
https://secure.flickr.com/photos/statelibraryqueensland/3198305152/
Motivations as opportunities

People crave:

- satisfying work to do
- the experience of being good at something
- time spent with people we like
- the chance to be a part of something bigger

(Jane McGonigal, 2009)
Motivations for organisations

• Fix the 'semantic gap' and enhance discoverability
• Digitisation backlog: collections are big, resources are small
• Create engaging, meaningful experiences for the public
• Access external knowledge and expertise
Going beyond 'microtasks'

• How can participants grow beyond 'classify this image' or 'type what you see'?
Participatory project models

Contributory
- the public contributes data to a project designed by the organisation

Collaborative
- both active partners, but lead by organisation

Co-creative
- all partners define goals together

(Center for Advancement of Informal Science Education (CAISE))
'Levels of Engagement' in citizen science

- Level 1: participating in simple classification tasks
- Level 2: participating in community discussion
- Level 3: 'working independently on self-identified research projects'

(Raddick et al, 2009)
FamilySearch ‘stepping stones’

• Indexing as ‘introductory, family history education’ including:
  – Knowledge about record types
  – Genealogical information
  – Handwriting practice
• From indexing, can move to ‘arbitration’
• Or onto your own family history research
The ethics of crowdsourcing

http://xkcd.com/1060
Design tips
Validate procrastination
Work with compelling content

Art and literature

Business

Children

City and town life

Country life

http://dh.tcd.ie/letters1916/diyhistory/
Other design solutions

• Understand and remove barriers to participation
• Emphasise importance of contribution
• Find interesting tasks, material
• Look to existing uses, communities
• Invest in community interaction, feedback
Design for cultural heritage crowdsourcing

• Understand your potential audiences’ interests and motivations
• Understand the quirks of your material
• Anticipate uses of the output content
• Understand barriers to participation

then...

• Tailor your design to suit
Future design challenges

• Integrating machine learning and human computation
  – What happens if we run out of meaningful tasks?
• Designing for mobile
• Participant retention
• Resources - crowdsourcing as 'free puppy'
Thank you!

Mia Ridge
Open University/Trinity College Dublin
http://miaridge.com/
@mia_out

Find out more:
Crowdsourcing our Cultural Heritage
http://www.ashgate.com/isbn/9781472410221
Crowdsourcing Iterations at NYPL Labs

Ben Vershbow - Director, Digital Library + Labs, New York Public Library

In the beginning...
Map Warper (2010 - present)

NYPL Map Warper

Maps of the city of New York / surveyed under directions of insurance companies of said city.

Depicts: 1854 Bibliographic records: CATNYP record
98 maps (98 maps rectified 100%)
(Visible) Download KML
(Visible) Export

Show Digitize Export Metadata Comments (0)

Transparency: 100

NATIONAL ENDOWMENT FOR THE Humanities
Map Warper (2010 - present)

Georectification task
Map Warper (2010 - present)
Map Warper (2010 - present)

NYPL Map Warper

Maps of the city of New York / surveyed under directions of insurance companies of said city.

Depicts: 1854 Bibliographic records: CATWHP record
98 maps (98 maps rectified 100%)
(Visible). Download KML

Show | Digitize | Export | Metadata | Comments (0)
Map Warper (2010 - present)

NYPL Map Warper

Atlas of the borough of Manhattan, city of New York.
Depicts: 1916
191 maps (183 maps rectified 96%)
(Visible) Download KML
Share | Download | Export

Building transcription task
Map Warper (2010 - present)

Open Data:
ESRI Shapefile, WMS, KML, CSV, GeoRSS
Progress:

> 5 thousand maps warped

> 120 thousand buildings transcribed
Challenges:

- transcription bottleneck

- steep learning curve
  (most transcription activity through onsite ‘citizen cartography’ workshops or classroom projects)
What's on the Menu? (2011-present)

Help The New York Public Library improve a unique collection!

We're transcribing our historical restaurant menus, dish by dish, so that they can be searched by what people were eating back in the day. It's a big job so we need your help! Learn more.

Connect: menus@nypl.org | Twitter | Facebook

Time to pay the taxman

Take the sting off April 15 with these financial menus!

Help review

Transcribed menus that need a second pair of eyes.

Explore

Browse the collection. 16,788 menus digitized and online.

Today's specials

An assortment of dishes from the past.

Map our Menus!

Need a break from transcribing? Help us locate menus with our new Geotagger tool.

**Help transcribe**

It's easy! No registration required!

So far: 1,242,566 dishes transcribed from 16,788 menus
What's on the Menu? (2011-present)
What’s on the Menu? (2011-present)
Quality assurance workflow
Quality assurance:

- 3-step workflow
- honor system
- very few instances of abuse
What's on the Menu? (2011-present)

Geolocation task
Progress:

> 1.3 million transcriptions

> 20 thousand geolocations
**What’s on the Menu?** (2011-present)

**Price**
Low: $0.30 → High: -

**Date appearing**
Earliest: 1901 → Latest: 1957

**Normalized frequency per year**
Appeared in 148 menus. Showing first 128: View all 148 menus

**Related dishes**
Click the + sign to add a dish to charts on this page or the eye to view its information page.

- Alligator Pear Rigare (1)
- Alligator Pear (half) (1)
- Alligator Pear [Salad] (1)
- Alligator Pear Florida (1)
- Salad, Alligator Pear (24)
- Alligator Pear salad (39)
- Alligator pear [salad] (1)
- Alligator pear (2)
- Alligator pear salad (1)
- Alligator Pear Smithfield (1)
- Alligator Pear, Half (1)
- Alligator pear (half) (1)
- Salad - Alligator Pear (1)
- Alligator Pear, half (5)
- Alligator Pear Salad (1, 59)
- Alligator Pear Waldorf (1)

**More information on “Alligator Pear”:**
- Images
- NYPL Digital Gallery, Flickr, Google Images.
- Recipes
- Epicurious, Google Recipe, Books, etc.

**Placement map:** Where this dish appears on the page.
What's on the Menu? (2011-present)

Data exports
There's a lot of data behind What's on the Menu?: a mix of simple bibliographic description of the menus (created by The New York Public Library) and the culinary and economic content of the menus themselves (transcribed by you). Now we're opening it up.

All data generated through What's on the Menu? is available in two ways:

Spreadsheet Exports
On the 1st and 16th of every month, we'll post a complete export of all menu and dish data collected so far (menus, dishes, prices, and more).

Download the latest data export in CSV format (09/16/12).

API
As the first project of NYPL Labs, we're happy to announce that Menus is also the first NYPL project to have a public API. In fact, we use this exact same API to build many of the features of this site.

You can learn how to use the API on our Github page, but you can get started now by sending us an email with the subject API ACCESS and a description of your project.

Also, feel free to add issues you've found using the API via our Github issues page.

What's an API?
No known copyright restrictions on this material. We ask that you credit The New York Public Library as source on any applications or publications.

Open Data / API
Curating Menus

We are scholars researching questions about food and culture using the historical menu collections from the New York Public Library.

Featured

When a Woman Collects Menus

The collection of historic menus at the New York Public Library (NYPL) was created by a woman named Frank E. Buttolph. We found ourselves compelled to tell a better and fuller story about her and her work as collector in the early twentieth century.

Read more

Digital Humanities: Data Curation
Challenges:

- digitization bottleneck (capacities + policies)
- staff time and role definitions
Success:

Small repeatable tasks
Map Warper (2010 - present)

NYPL Map Warper

Atlas of the borough of Manhattan, city of New York.
Depicts: 1916
191 maps (183 maps rectified 96%)

Building transcription task
Map Warper (2010 - present)

Atlas of the borough of Manhattan, city of New York.
Depicts: 1916
191 maps (183 maps rectified 96%)

Building transcription task
Map Warper (2010 - present)

NYPL Map Warper

Atlas of the borough of Manhattan, city of New York.
Depicts: 1916
191 maps (183 maps rectified 96%) (Visible) Download KML
Plus! Locating place on map to do work Consulting original map key (printout)

Building transcription task
Can we break this into smaller pieces?
Question:

Can we break this into smaller pieces?

And make it fun?
Map Warper (2010 - present)

Also:

Can a computer do any of this?
Map Vectorizer (2013)
github.com/NYPL/map-vectorizer
Map Vectorizer (2013)

github.com/NYPL/map-vectorizer
OCR for maps!
Quality control?
Building Inspector (2013-present)

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.

GET STARTED
Task 1: Check Footprints
Task 1: Check Footprints
Task 2: Fix Footprints
Task 3: Enter Addresses
Building Inspector (2013-present)

Task 4: Classify Colors
Building Inspector (2013-present)

Responsive design
Building Inspector (2013-present) 

Check

Consensus workflow
Consensus workflow

Check

YES

Address
Color

Building Inspector (2013-present)
bldinginspector.nypl.org
Building Inspector (2013-present)

Check

YES
Address

FIX
Color

Consensus workflow
Building Inspector (2013-present)

Check

- YES
  - Address
  - Color

- FIX
  - Fix

Consensus workflow

buildinginspector.nypl.org
Building Inspector (2013-present)

Check

YES*

Address

Color

* Consensus ‘NO’s go to polygon heaven

buildinginspector.nypl.org
Progress:

> 910 thousand tasks completed
Challenges:

• lack of promotion from NYPL properties

• lack of community architecture (beyond basic authentication/task tally)
Building Inspector (2013-present)

Next:
• more layers
• subway mode

Image: The New York Times
In progress:

**Scribe**
*Turn Documents into Data Sets*

**NYPL + Zooniverse**
Thank you!

@subsublibrary
@nypl_labs
labs@nypl.org

buildinginspector.nypl.org
Operation War Diary: 
Enriching Catalogues, Enhancing Research
Zooniverse

We make citizen science websites so that everyone can be part of real research online

daily.zooniverse.org

Something awesome from the Zooniverse every day

Space

How do galaxies form?
NASA's Hubble Space Telescope archive provides hundreds of thousands of galaxy images.

Explore the surface of the Moon
We hope to study the lunar surface in unprecedented detail.

Study explosions on the Sun
Explore interactive diagrams to learn about the Sun and the spacecraft monitoring it.

Find planets around stars
Lightcurve changes from the Kepler spacecraft can indicate transiting planets.

GALAXY Z@O

MOON ZOO

SOLAR STORMWATCH

planethunters.org
Few have witnessed what you're about to see
Experience a privileged glimpse of the distant universe as observed by the SDSS, the Hubble Space Telescope, and UKIRT

Classify Galaxies
To understand how galaxies formed we need your help to classify them according to their shapes. If you're quick, you may even be the first person to see the galaxies you're asked to classify.

Begin Classifying

How Do Galaxies Form?  History of Galaxy Zoo
Transcription Projects

• Existing projects:
  - ‘Old Weather’: www.oldweather.org/
  - ‘Notes from Nature’: www.notesfromnature.org/
  - ‘Operation War Diary’: www.operationwardiary.org/

• Upcoming Projects:
  - ‘Secret Lives of Artists’ with Tate Britain
  - ‘Shakespeare’s World’ with Folger Shakespeare Library
Ancient Lives

Study the lives of ancient Greeks

The data gathered by Ancient Lives helps scholars study the Oxyrhynchus collection. Transcriptions collected digitally are combined with human and computer logic to identify known texts and documents.
The story of the British Army on the Western Front during the First World War is waiting to be discovered in 1.5 million pages of unit war diaries. We need your help to reveal the stories of those who fought in the global conflict that shaped the world we live in today.

GET STARTED!

Working together with Citizen Historians during the First World War Centenary

Zooniverse

IWM

The National Archives

Zooniverse privacy policy
OWD by the numbers:

• Over 1.2 million page views since January
• 10,790 registered users
• 55,050 comments on Talk, the project discussion area
• 1,884 commenting users
• Approximately 61,000 pages ‘completed’ i.e. tagged and transcribed by 7 users.
3 DIVISION: DIVISIONAL AMMUNITION COLUMN (1 AUG 1914 - 30 SEP 1919)

WAR DIARY

INTELLIGENCE SUMMARY.

(Enemy hostilities not required.)

Date, Time, Place

Summary of Events and Information

6th Sep. LES BROSSES

Left St. Pierre at 6 p.m. and arrived near

LES BROSSES at 8 p.m.

9th Sep. REBAIS

Left 8.30 p.m. arrived at REBAIS at 11 p.m.

10th Sep. MONTREUIL

Left at 10 a.m. marched by ORLY, BUSIERES, and

arrived MONTREUIL at 1.30 p.m.

11th Sep. VEUVILLY

MONTREUIL - MERVY and arrived at 5 p.m

12th Sep. VEUVILLY

DAMMARD

Left at 1 p.m. and marched via VEUVILLY - PULCHY

LA VILLE - GRAND ROZOT arriving 5 p.m.
<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Summary of Events and Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Jun 1917</td>
<td>4</td>
<td>198th Infantry Brigade, 198th Infantry Brigade (2)</td>
</tr>
<tr>
<td>5 Jun 1917</td>
<td>1</td>
<td>Le Quesnoy (7)</td>
</tr>
<tr>
<td>5 Jun 1917</td>
<td>3</td>
<td>Estaires (16)</td>
</tr>
<tr>
<td>6 Jun 1917</td>
<td>1</td>
<td>Le Quesnoy (5)</td>
</tr>
<tr>
<td>11 Jun 1917</td>
<td>2</td>
<td>Le Quesnoy (6)</td>
</tr>
<tr>
<td>11 Jun 1917</td>
<td>1</td>
<td>Le Quesnoy (5)</td>
</tr>
<tr>
<td>11 Jun 1917</td>
<td>1</td>
<td>Le Quesnoy (5)</td>
</tr>
<tr>
<td>11 Jun 1917</td>
<td>3</td>
<td>Le Quesnoy (6)</td>
</tr>
<tr>
<td>21 Jun 1917</td>
<td>1</td>
<td>R 63 (1)</td>
</tr>
<tr>
<td>21 Jun 1917</td>
<td>2</td>
<td>Battery left and returned to Waggon Line at NO 5 and 2/6 East Lancs.</td>
</tr>
<tr>
<td>21 Jun 1917</td>
<td>2</td>
<td>D. D. L.</td>
</tr>
<tr>
<td>21 Jun 1917</td>
<td>2</td>
<td>D. D. L.</td>
</tr>
</tbody>
</table>

http://wd3.herokuapp.com/pages/AWD0000h3c
OWD Goals

• To provide evidence about the experiences of named individuals in the field diaries and add this to the Imperial War Museum’s ‘Lives of the First World War’ project: https://livesofthefirstworldwar.org/

• To enrich the National Archives’ catalogues

• To gather data for academic research
Are we on target?
Like Operation War Diary on Facebook and follow on Twitter to stay up to date on new discoveries.

@OpWarDiary  
https://www.facebook.com/OperationWarDiary?ref=hl
Get Involved:

Look for upcoming tweets for further details!

Comments, questions? Let us know!

Join the Crowd (Consortium)!

@crowdconsortium

contact@crowdconsortium.org