Measuring the ANZACs

Crowdsourcing a complete transcription of WWI soldiers through Zooniverse

Evan Roberts, Minnesota Population Center
University of Minnesota

Joint research with
Kris Inwood, Economics and History,
University of Guelph

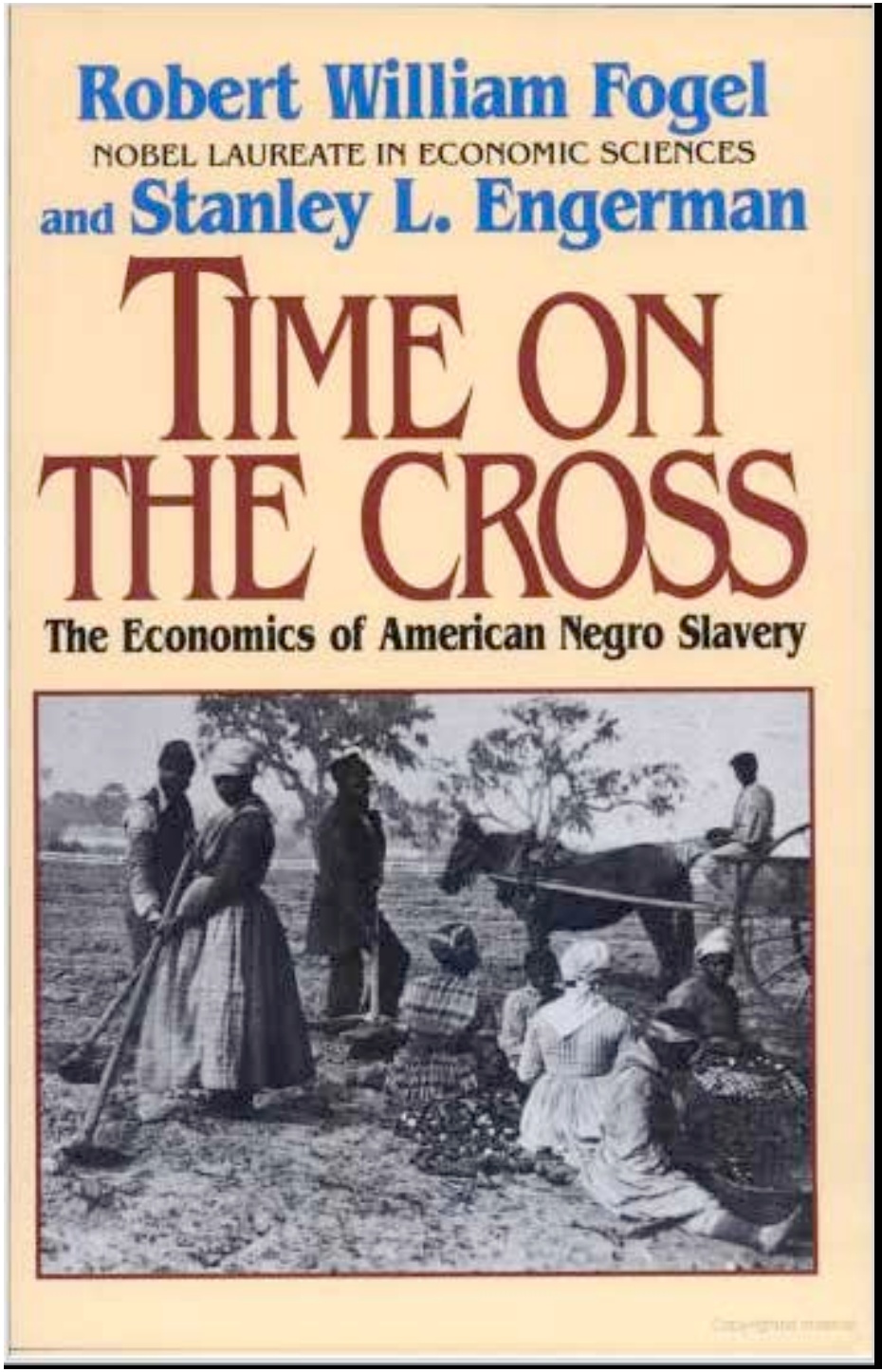
Les Oxley
Economics
University of Waikato

Collaborating institutions
Archives New Zealand
Auckland War Memorial Museum



http://www.measuringtheanzacs.org/

Background and motivation: Using large databases of heights to measure well-being



Anthropometric history

- Controversial 1974 book claimed material well-being of slaves similar to free people
- How do you measure material well-being without information on earnings?

Stature as a measure of living standards

- Changes in average stature of a population reflects net nutrition during growth period (0-20 years)
- Reductions (improvements) in net nutrition can come from
  - Lower (greater) food consumption
    e.g. price of food rises (falls)
  - Increased (reduced) physical exertion
    e.g. Children working more (less) on farms during adolescence
  - Increased (reduced) incidence of disease (particularly infectious disease)

Nature of Research on Stature

- Height ‘potential’ of most large populations is comparable
- Need individual-level evidence but in the end address the dominant tendencies of a population
- Organize by birth cohorts
- Large samples needed because genetic variation is strong but largely uninteresting.
- Also some effects stronger than others: eg cohort change is glacial but class differences huge implying methodological challenge
- Rely heavily on military and prison sources

Key findings in anthropometric history

- “Antebellum puzzle”: Despite rising incomes, the average stature of American men shrank between 1820 and 1840
- “Industrialization is hazardous to your health”: Despite rising incomes average stature stagnated in 1880s and 1890s in many European countries.
- “Peculiar population”: American slaves severely stunted as children, but stature recovered in adolescence and adulthood
- “Tallest in the world”: Plains Indians were vey tall despite population decline

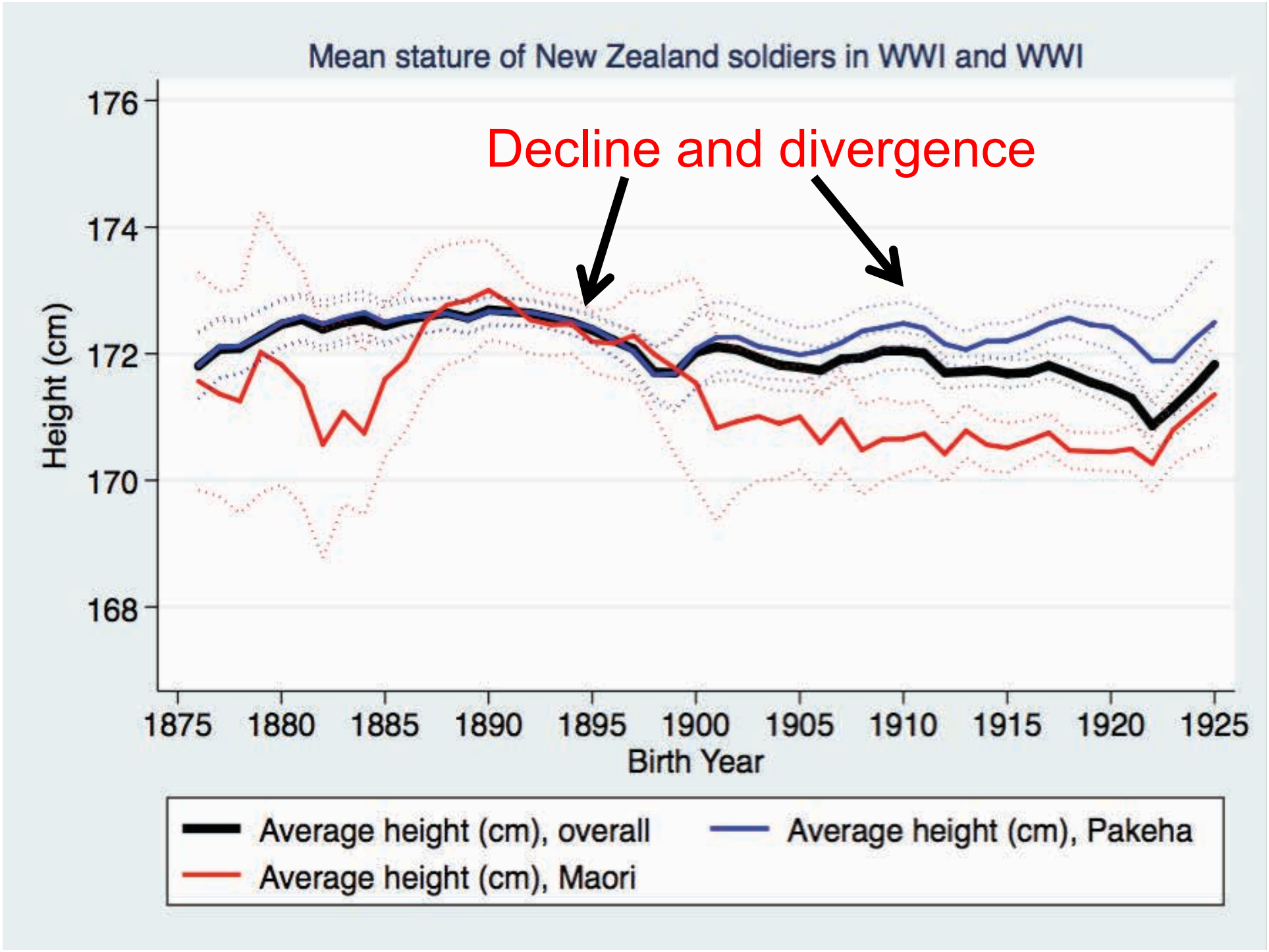
Background and motivation

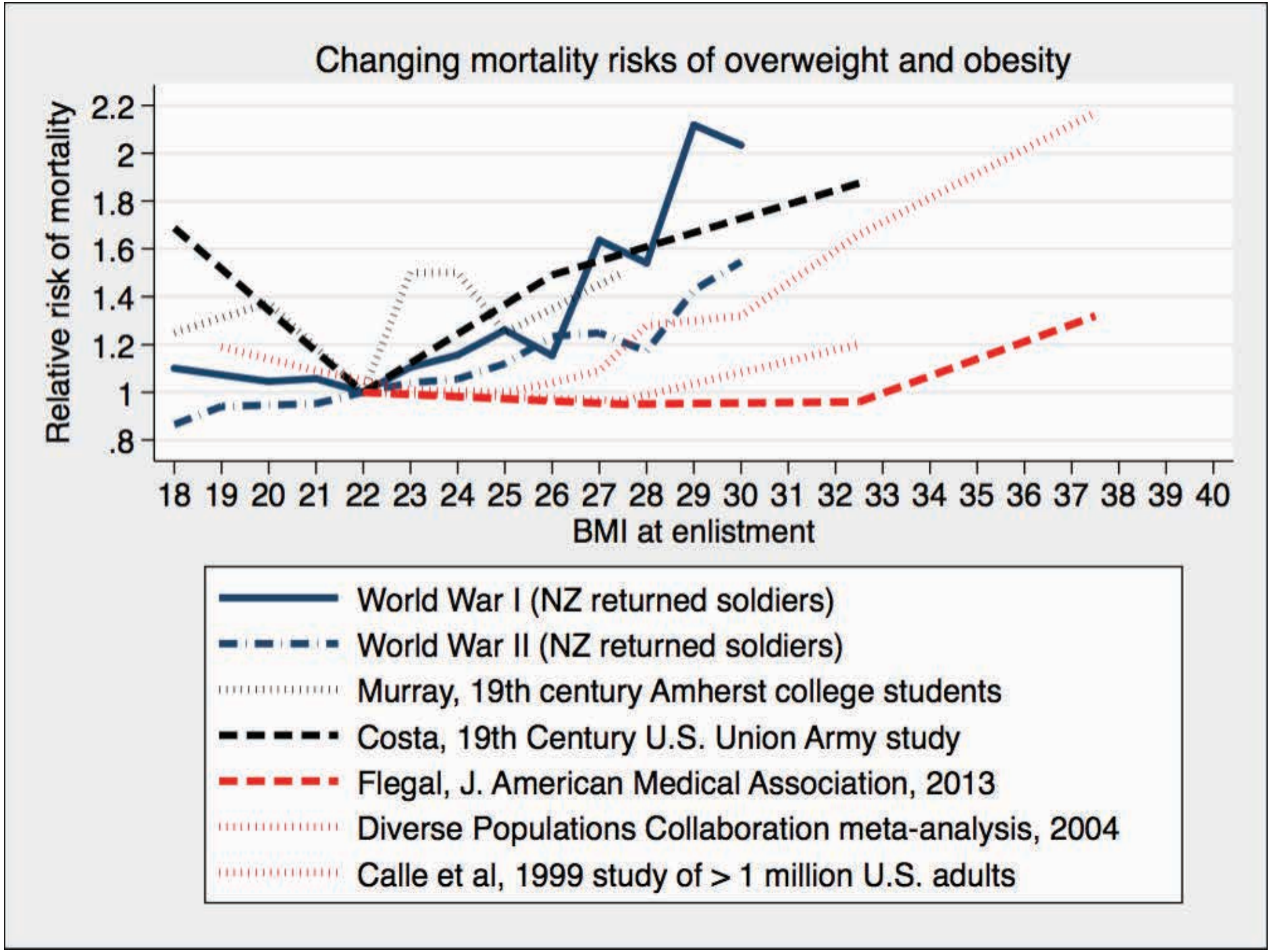
NZ anthropometric history

Our long-term research agenda

- How have health and biological standards of living changed in New Zealand from c.1850 to the present?
- How have differences in health between the indigenous Maori population and European settlers changed from c. 1850 to the present?
- How do biological measures of living standards confirm or refute measures of New Zealand economic growth based on income data?

3. Trends in stature in New Zealand


Mean stature of New Zealand soldiers in WWI and WWI
Decline and divergence
Height (cm)
176
174
172
170
168
1875
1880
1885
1890
1895
1900
1905
1910
1915
1920
1925
Birth Year
Average height (cm), overall
Average height (cm), Pakeha
Average height (cm), Maori

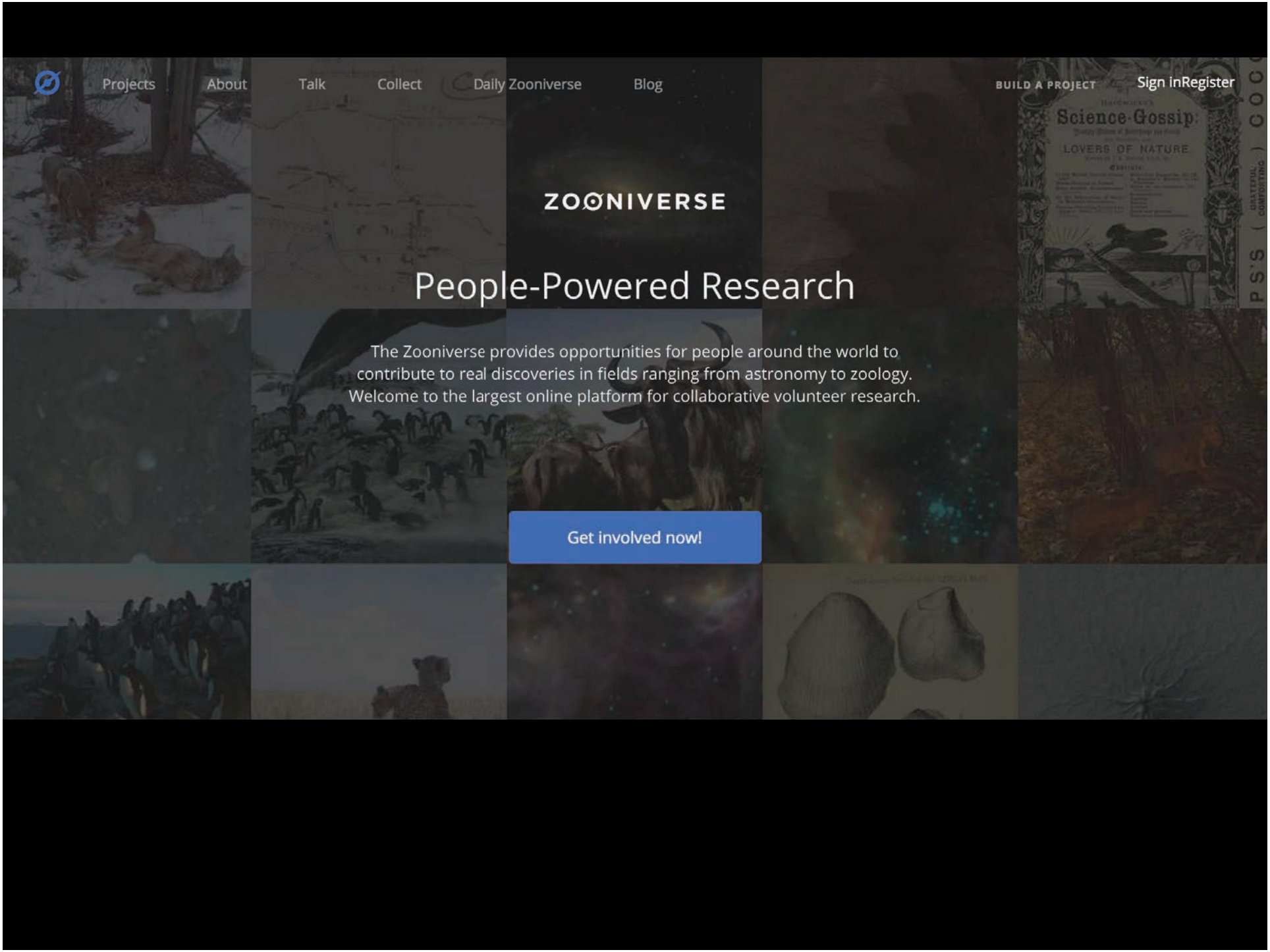

Changing mortality risks of overweight and obesity
Relative risk of mortality
2.2
2
1.8
1.6
1.4
1.2
1
.8
18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
BMI at enlistment
World War I (NZ returned soldiers)
World War II (NZ returned soldiers)
Murray, 19th century Amherst college students
Costa, 19th Century U.S. Union Army study
Flegal, J. American Medical Association, 2013
Diverse Populations Collaboration meta-analysis, 2004
Calle et al, 1999 study of > 1 million U.S. adults

What we've found so far

- Health disparities for Māori emerged several generations after colonial settlement
- Similar findings to US historical studies that mortality risk rises for BMI above 28 (higher end of overweight and obese)
- New Zealand data for World War II suggest a significant reduction in mortality risk between cohorts just 25 years apart. Being overweight is less risky than it used to be [at least for mortality]

Limitations of our current sample

- ~20,000 records [reasonable size in this literature]
- Can't do too many interactions in our models
  - Did Māori do better or worse in particular birth cohorts?
  - Did class differentials vary across time or ethnicity?
- We did not collect combat exposure information
  - (service history sheets)
- Nearly a thousand women served [as nurses]
  - It would be great to include them


Projects
About
Talk
Collect
Daily Zooniverse
Blog
BUILD A PROJECT
Sign inRegister
ZOONIVERSE
People-Powered Research
The Zooniverse provides opportunities for people around the world to contribute to real discoveries in fields ranging from astronomy to zoology. Welcome to the largest online platform for collaborative volunteer research.
Get involved now!

What are these?



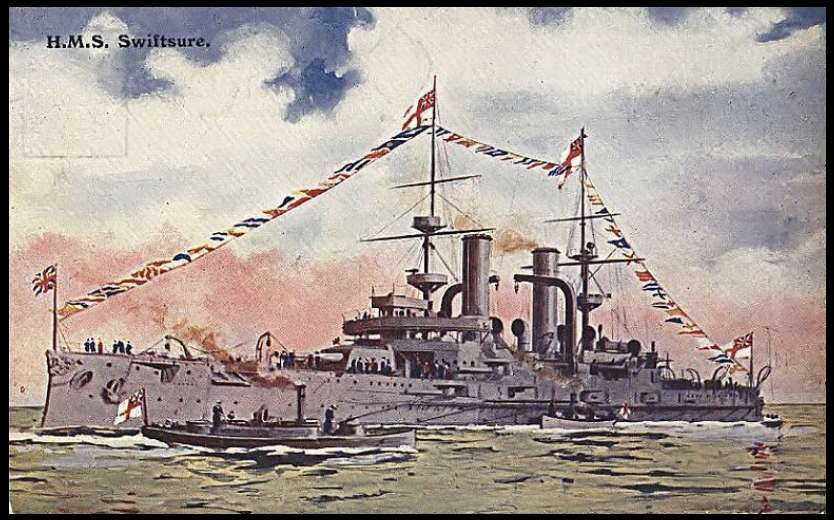


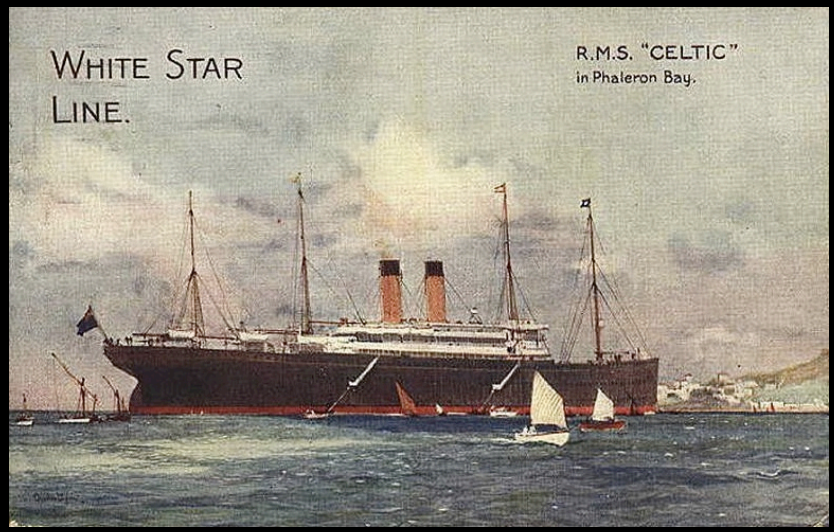
And what do they have to do with these?



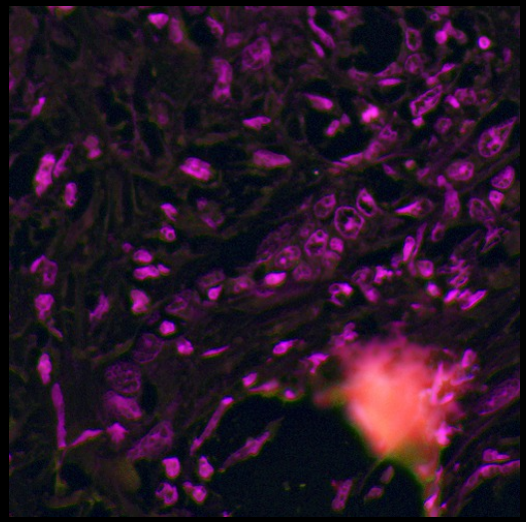


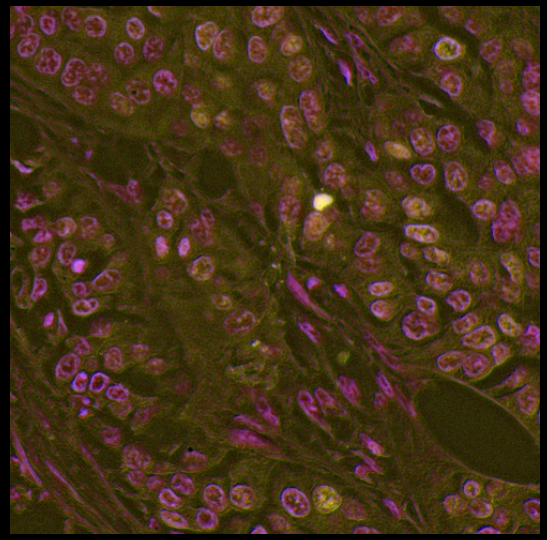
Or in fact these?





Or these?



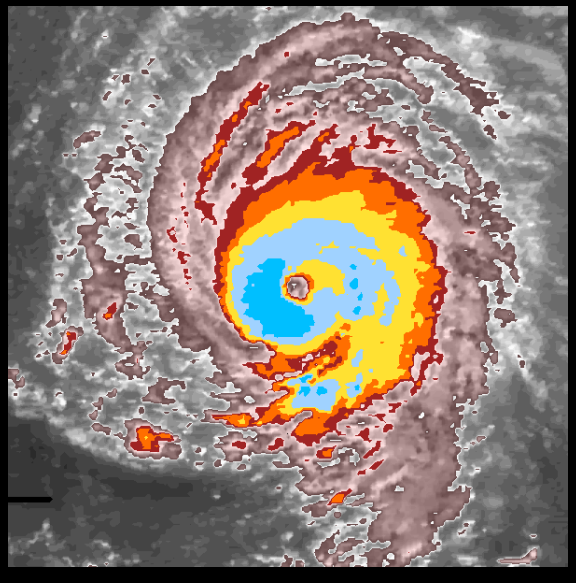


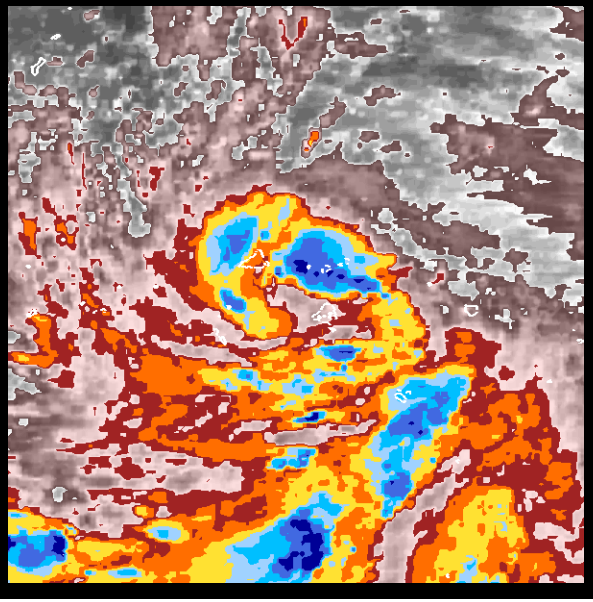
Or even these?





Or these?





These are all images from research projects that rely on analysis of complex images and pattern recognition capabilities that are currently beyond computer algorithms!

Which is better at spotting the tiger?







WildCam Gorongosa

Identify animals in trail camera images from Gorongosa National Park!

Get Started



Fossil Finder

Join us in the search and discovery of fossils at Lake Turkana, Kenya.

Get Started



Galaxy Zoo: Bar Lengths

Measure the engines of evolution in disk galaxies.

Get Started



Whales as Individuals

Help us identify individual Humpback Whales by clueing our computer algorithms in to patterns on their tails

Get Started



Season Spotter Image Marking

Help keep an eye on changing seasons by marking images!

Get Started



Season Spotter Questions

Help keep an eye on changing seasons by answering questions!

Get Started



Chimp & See

Discover the secret life of chimpanzees. We need your help to study, explore, and learn from thousands of videos.

Get Started



AnnoTate

Help transcribe documents from the Tate collection, and reveal the secret lives of artists.

Get Started



Science Gossip

Uncover the history of citizen science. Help us to classify their drawings and map the origins of citizen science.

Get Started



Wildebeest Watch

Explore collective intelligence in wildebeest!

Get Started

Classifying: another human evolutionary tic

Amazingly, only two basic galaxy shapes – but very complex, no two are the same.



Spirals

Lots of star formation so mostly (but not all!) blue-ish.



Ellipticals

Older star populations so mostly (but not all!) red-ish.

Historically, astronomers would classify galaxies "by eye" even when 10,000 images! But what about a million galaxies??

Astronomy as example of data flood



2.0 meter Sloan Digital Sky Survey Telescope
Apache Point, New Mexico

1980 – Palomar Sky Survey:
10,000 galaxies
One expert can classify galaxies visually.

2000s - Sloan Digital Sky Survey:

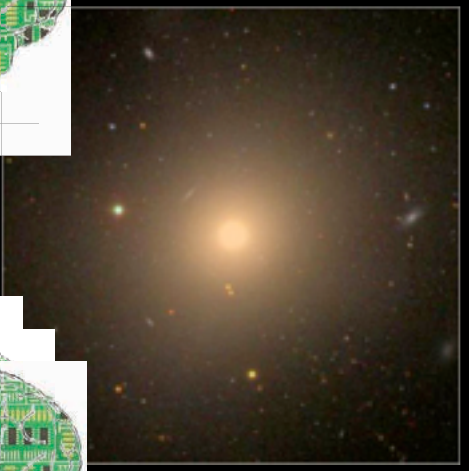
~1000000 galaxies

Need computers to process that much data!

But how can we “computerize” the best algorithm that can process complex data?







Galaxy Zoo launched in 2007 inviting public to classify galaxies



In 1.5 years, 35 million classifications by ~150,000 users

Roughly 3.3 continuous person-years!

Tutorial: Simple, divisible tasks

Part 1B ... More Tricky Spiral or Elliptical Galaxies

Some galaxies are a bit more tricky. As you noticed in the previous section, some spiral galaxies can look like ellipticals when viewed edge-on. Also, in some faint spiral galaxies, you have to look hard to see the spiral arms. Now, see if you can separate the genuine ellipticals from the spirals.

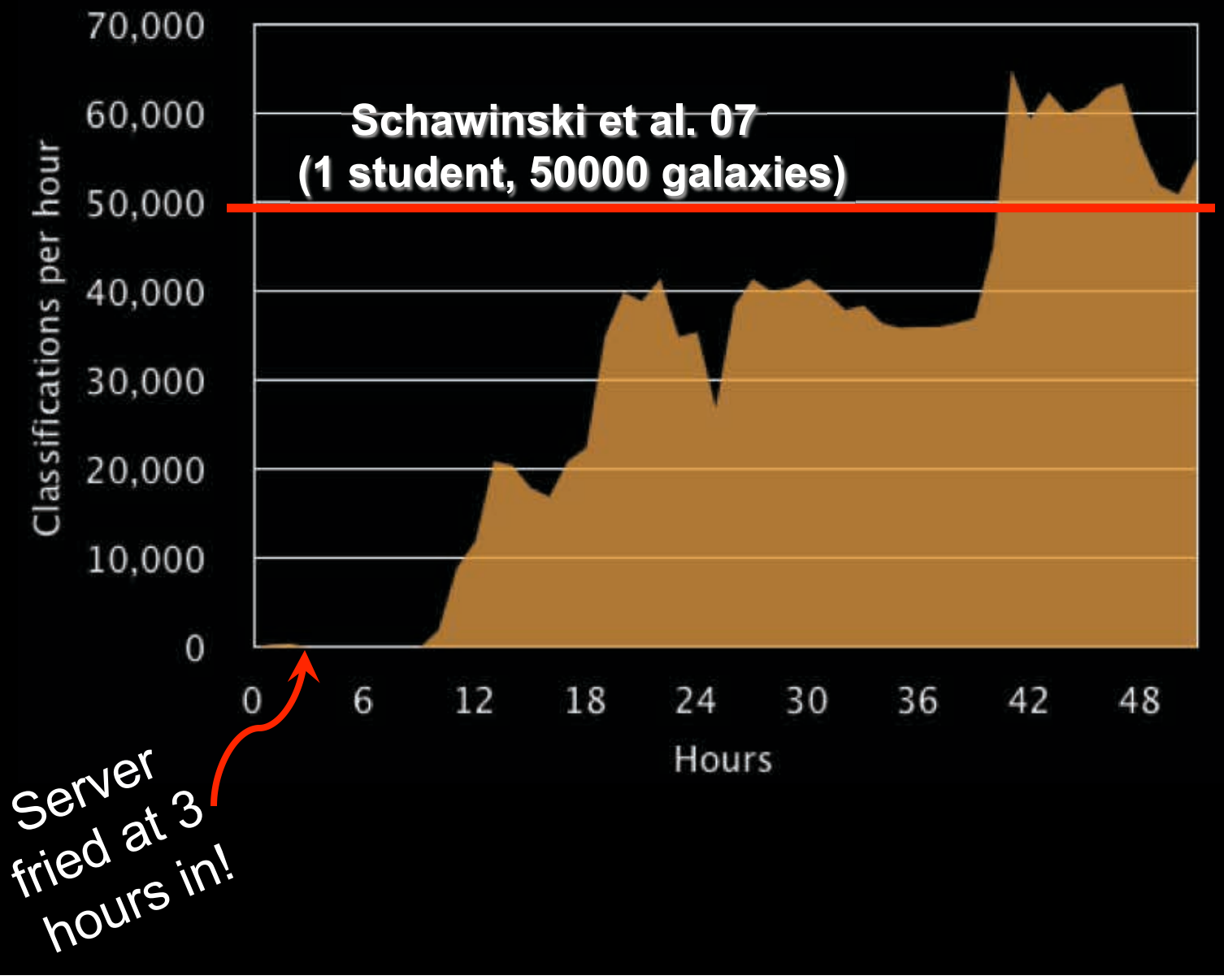
Try your hands at some!
Click the image to see if you're right.



Part 1C ... Merging Galaxies

Sometimes, galaxies crash into each other, or come close. These are called merging galaxies. Merging galaxies are very interesting to astronomers because we think that large galaxies are built from mergers of small galaxies – if we see merging galaxies, we can see a snapshot of how that process happens. When you look for mergers, look for places where two galaxies appear to be merging into one. The galaxies should be close together, and you should be able to see some connection between them. In the trial or in your galaxy analysis, whenever you see this, click the button that says "Mergers".

The first two days…

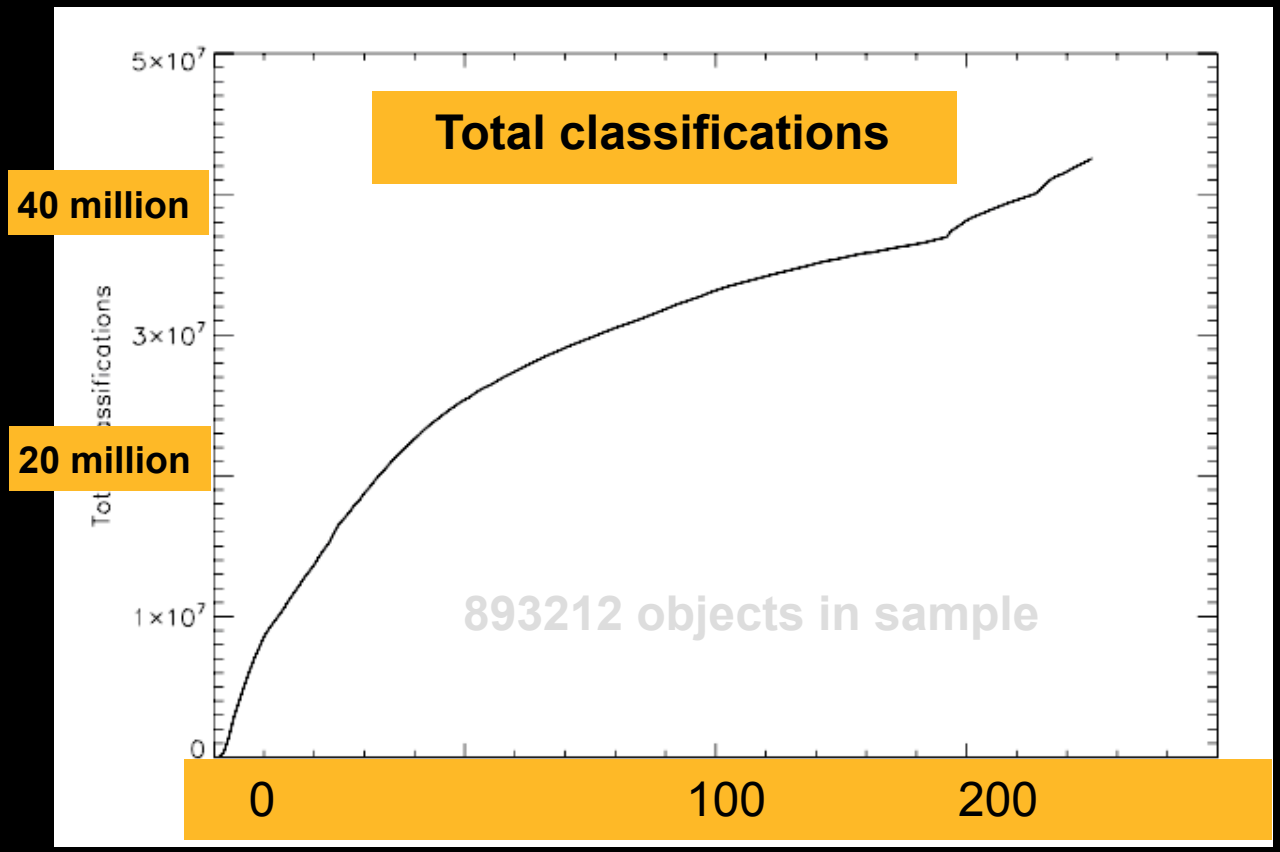






Public outreach and media presence
to engage broad and large audience

First six months



After ‘cleaning’ raw clicks:

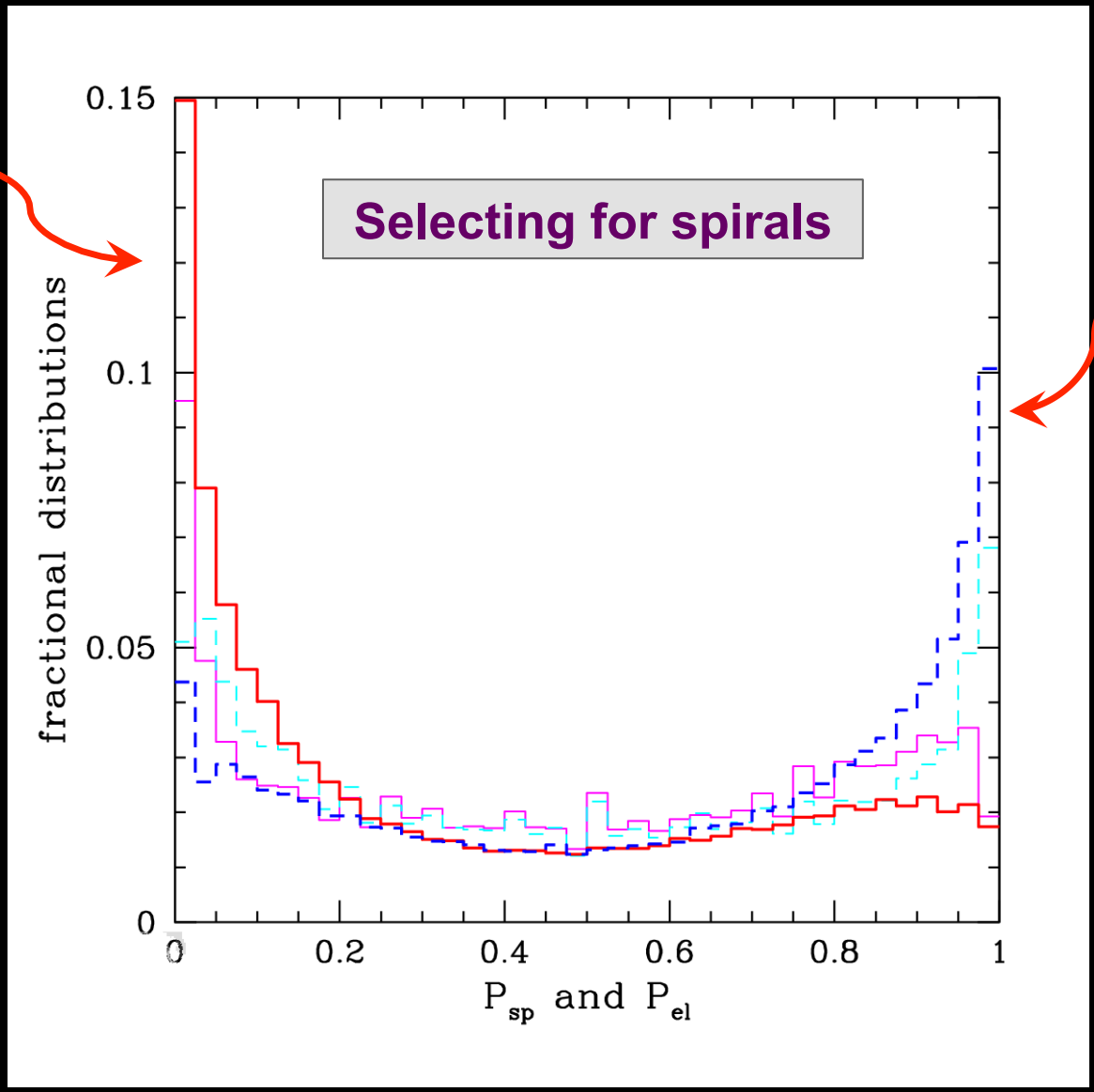
34,617,406 classifications by 82,931 users

Roughly 3.3 continuous person-years!

Combining clicks

"Pure" elliptical sample

"Pure" spiral sample



Galaxy Classification – now as probabilities!

The case for crowdsourcing your research!

The scale of the problem – what do we do with 40 Tbytes a day?

- Largest professional classification is only ~5% of SDSS
- Galaxy Zoo provided 3.3 person years in first 6 months

(Measurable) Accuracy – wisdom of the crowds

- Multiple independent classifications give us an estimate of error

- For a set of noisy data, enough inexpert classifiers will produce more accurate classifications than an expert classifier (where 'enough' depends on how noisy the data, and how inexpert and expert the classifiers).

OK – BUT!!

1. Can we apply this to research beyond galaxy zoo?
2. Can we get the attention of the crowd?

The Data Deluge



BIG DATA!!

Not just Astronomy suffering from Big Data!

Moving Beyond Galaxy Zoo

- **"Information age" = Data-Intensive Research**

Researchers overwhelmed with data sets from many different sources

- Data captured by instruments
- Data generated by simulations
- Data generated by sensor networks
- Data generated through digital recovery methods

- eScience is the set of tools and technologies to support data federation and collaboration
  - For analysis and data mining
  - For data visualization and exploration
  - For scholarly communication and dissemination

eScience is also the technological backbone that enables today's Citizen Science

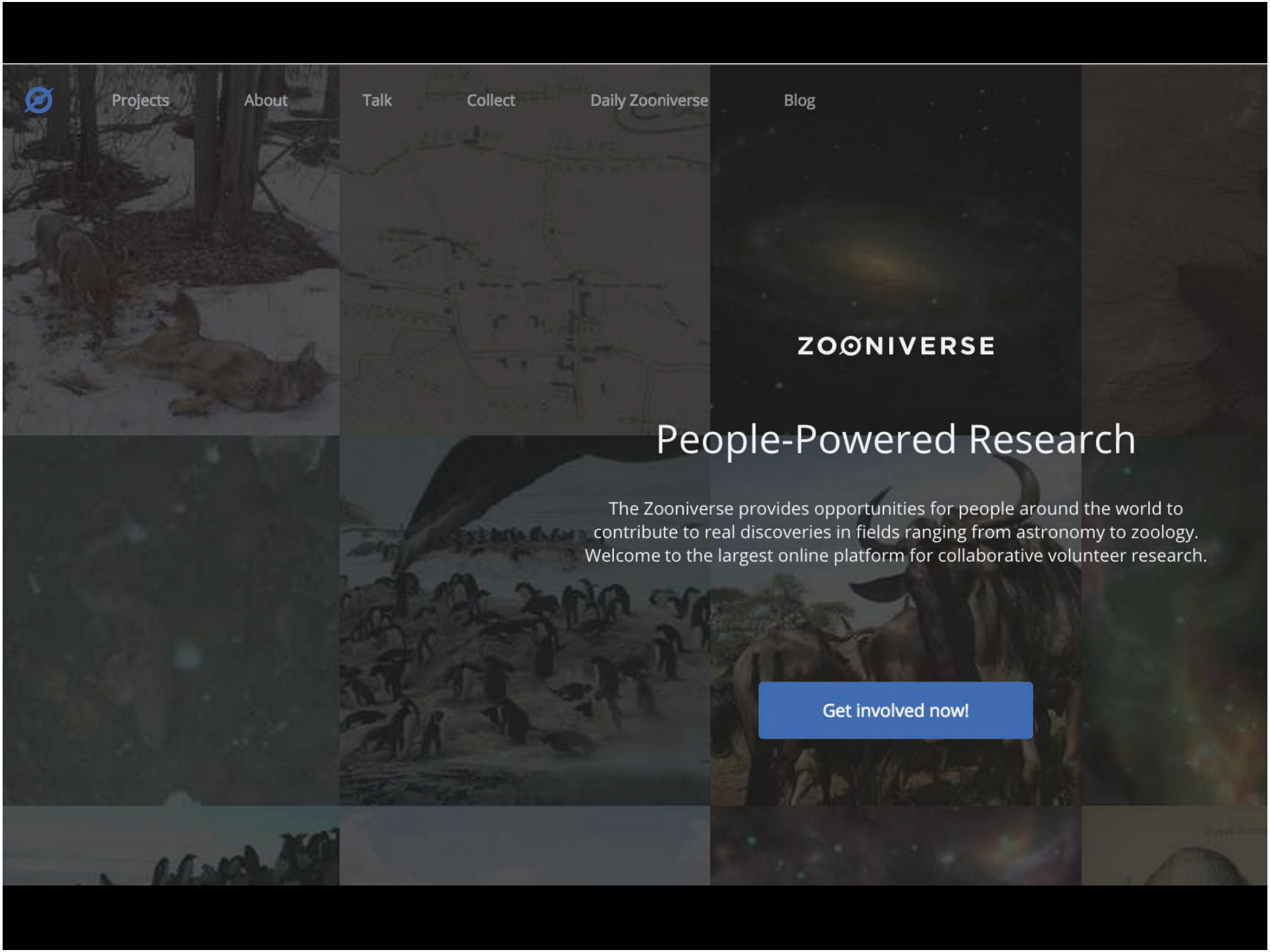
Moving Beyond Galaxy Zoo

1. **Passive** – distributed computing so no involvement from volunteer other than their computer (SETI@Home, Einstein@Home)

2. **Data Collection** – distributed sample collection/ observation (**bird counts**, variable star observations, weather data, **community-based documentation**)
   - "Classic" citizen science solving the need for distributed sensors as well as current internet-enabled data collection projects.



3. **Data Analysis** - distributed data analysis: enabled by the internet (**Zooniverse**, FoldIt!)
   - Solution to large amounts of complex data where volunteers are part of data-processing pipeline for complex systems e.g. pattern matching


Projects
About
Talk
Collect
Daily Zooniverse
Blog
ZOONIVERSE
People-Powered Research
The Zooniverse provides opportunities for people around the world to contribute to real discoveries in fields ranging from astronomy to zoology. Welcome to the largest online platform for collaborative volunteer research.
Get involved now!



10,000
volunteers

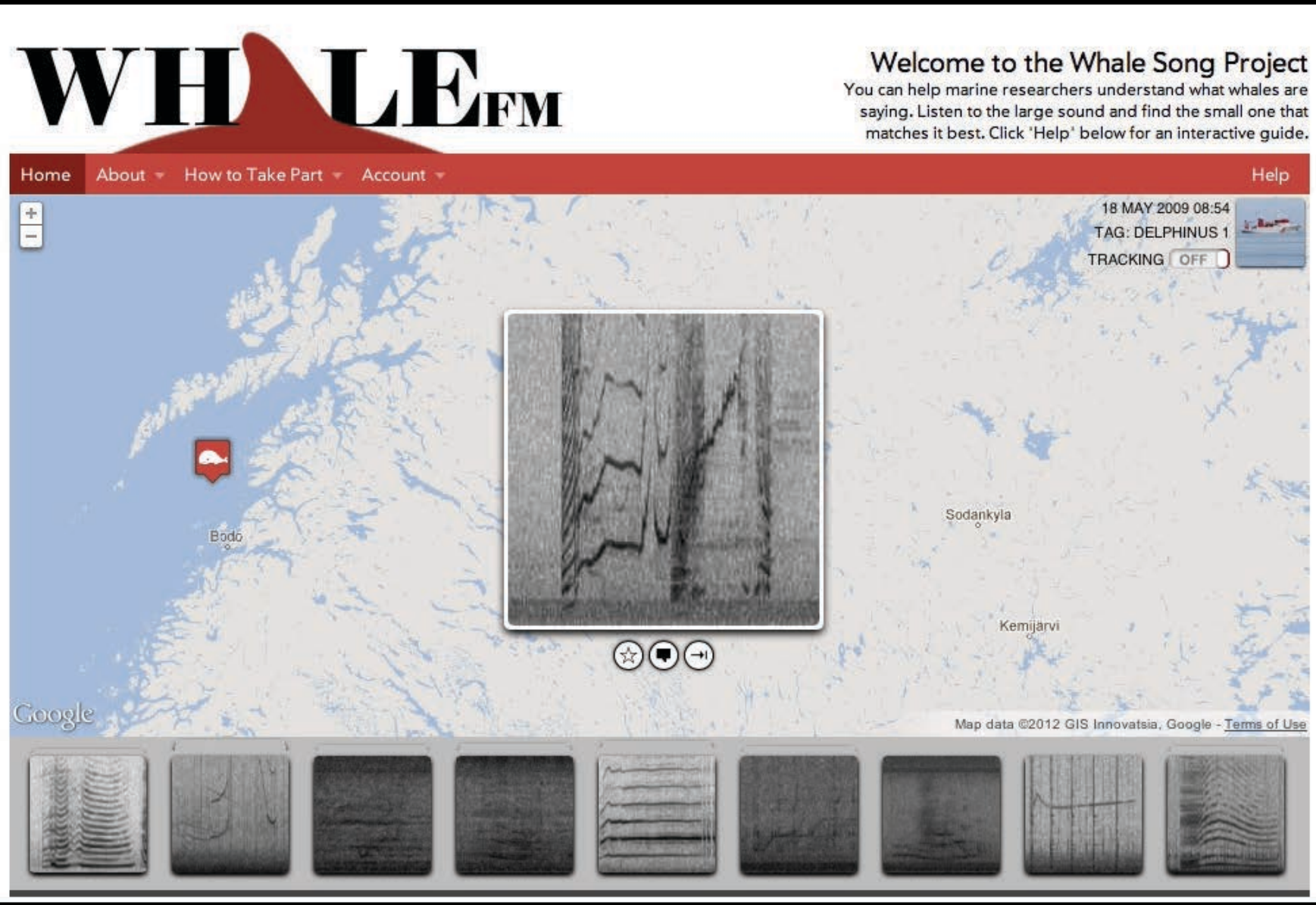
classifications
150,000



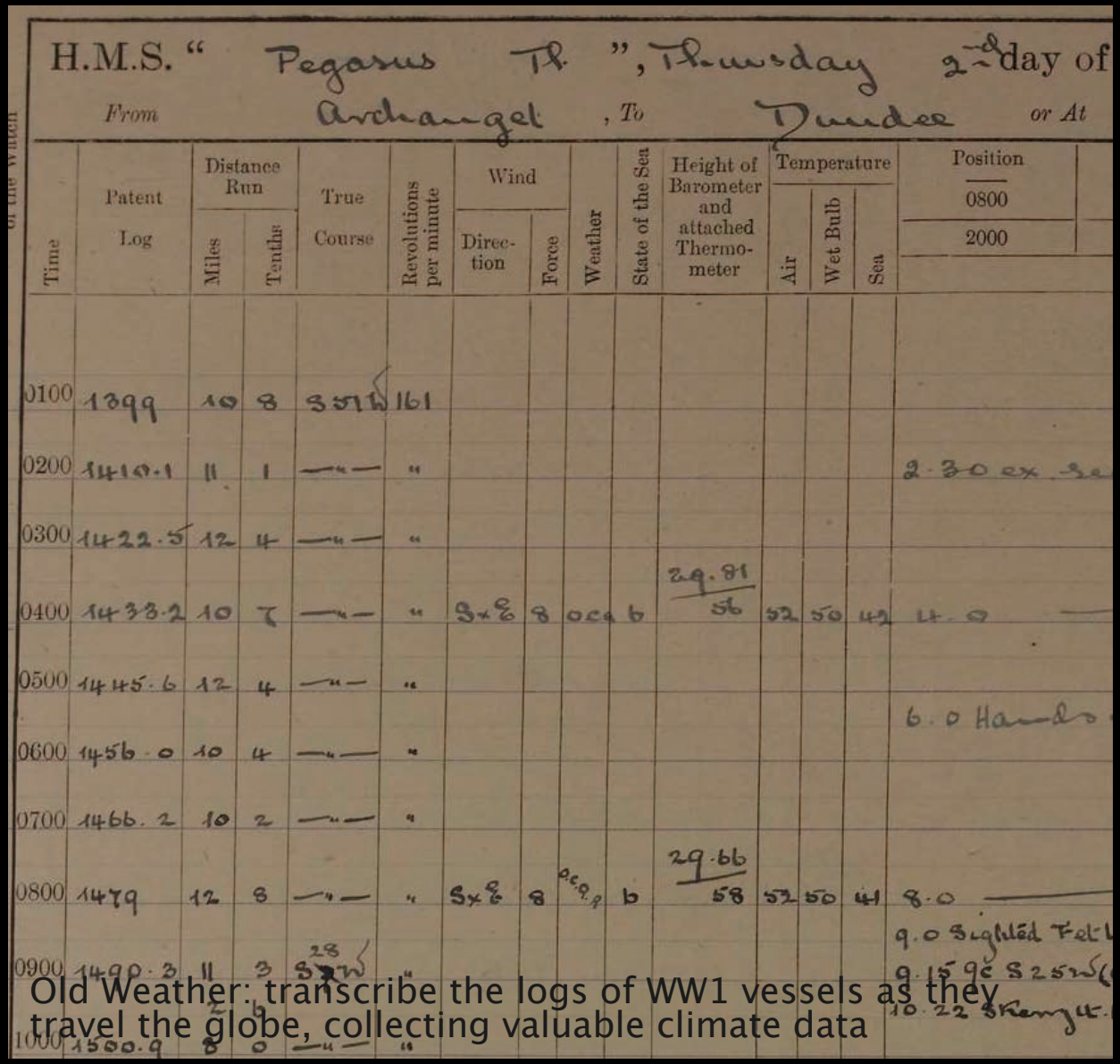


- Whale FM: Listen to whale calls to decode their language

whale.fm


WHALE FM
Welcome to the Whale Song Project
You can help marine researchers understand what whales are saying. Listen to the large sound and find the small one that matches it best. Click 'Help' below for an interactive guide.
Home
About
How to Take Part
Account
Help
18 MAY 2009 08:54
TAG: DELPHINUS 1
TRACKING OFF
Bodo
Sodankyla
Kemijarvi
Google
Map data ©2012 GIS Innovatsia, Google - Terms of Use

Old Weather: Ships logs have a similar data structure and transcription issues to military personnel files



H.M.S. " Pegasus Th. ", Thursday 2nd day of

From Archangel , To Dundee or At

Time	Patent Log	Distance Run: Miles	Distance Run: Tenths	True Course	Revolutions per minute	Wind: Direction	Wind: Force	Weather	State of the Sea	Height of Barometer and attached Thermometer	Temperature: Air	Temperature: Wet Bulb	Temperature: Sea	Position 0800 / 2000
0100	1399	10	8	S 57½ W	161									
0200	1410·1	11	1	—"—	"									2·30 ex
0300	1422·5	12	4	—"—	"									
0400	1433·2	10	7	—"—	"	SxE	8	Occ	b	29·81 / 56	52	50	42	4·0
0500	1445·6	12	4	—"—	"									
0600	1456·0	10	4	—"—	"									6·0 Hands
0700	1466·2	10	2	—"—	"									
0800	1479	12	8	—"—	"	SxE	8	Occ	b	29·66 / 58	52	50	41	8·0
0900	1490·3	11	3	S 28 W	"									9·0 Sighted Tel 9·15 9c S 25 W
1000	1500·9	8	0	—"—	"									10·22 Steering

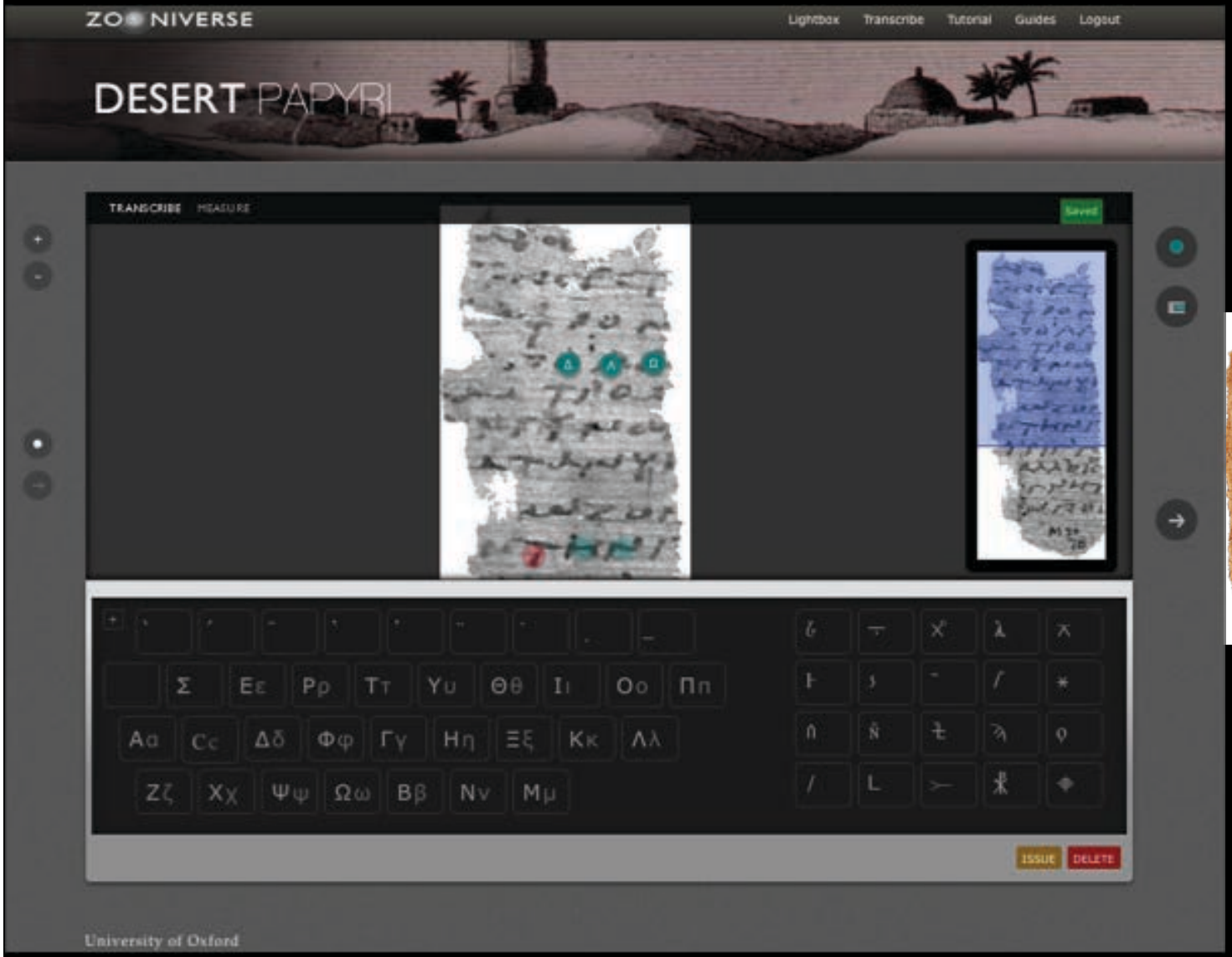
- Old Weather: transcribe the logs of WW1 vessels as they travel the globe, collecting valuable climate data

25,000 volunteers
1,000,000 pages





Ancient Lives Project: transcribing 2000 year old papyri fragments from Oxyrhynchus, Egypt. Partnership between UMN and Oxford, UK.



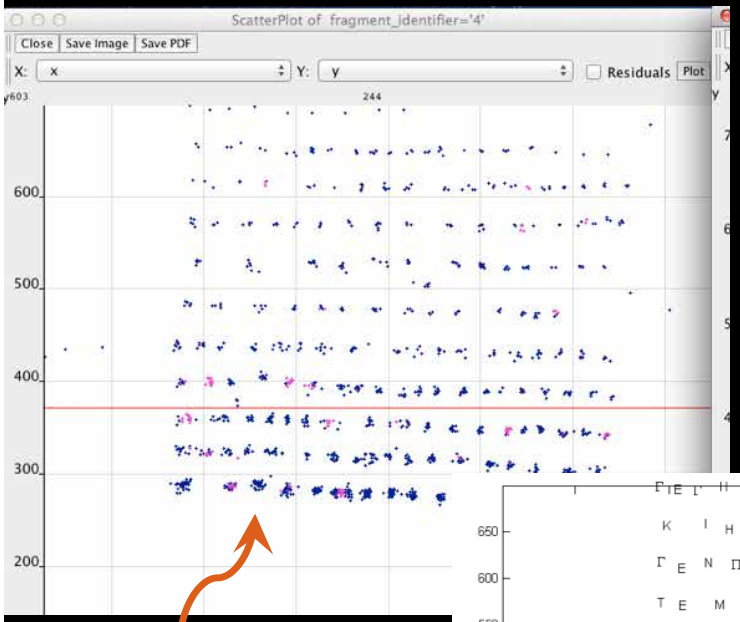
300,000
volunteers

classifications
100m

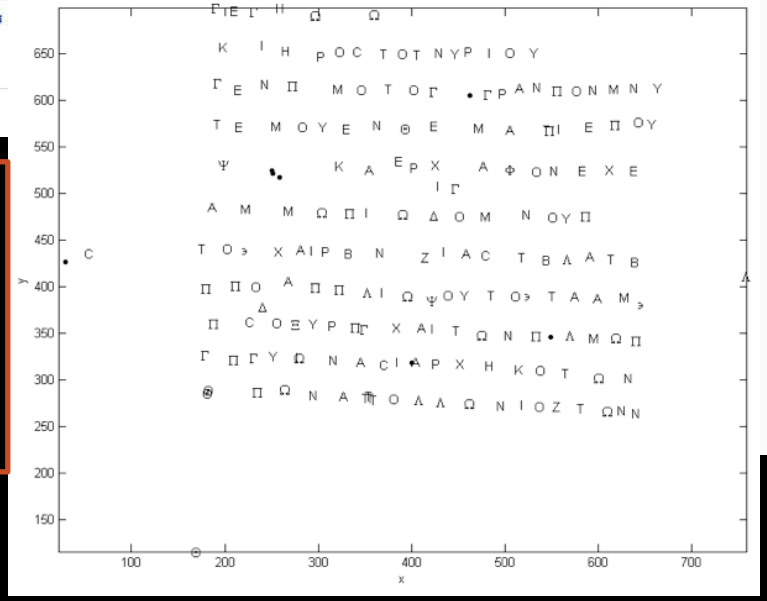




Ancient Lives Project: UMN developed the pipeline to turn clicks into consensus into transcribed text that can be edited & curated.

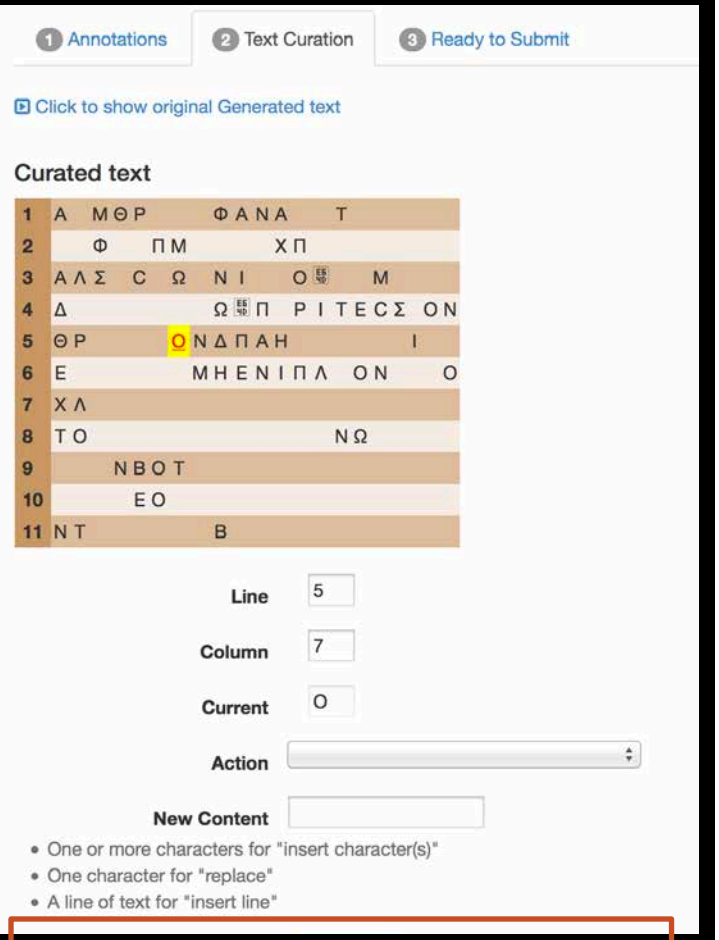


Centroiding raw clicks and aligning centroids to factor out scribe's slant.



Identifying most probable letter with each centroid.





Annotation and curation tool displaying original image and consensus transcription that can be edited or augmented with metadata.

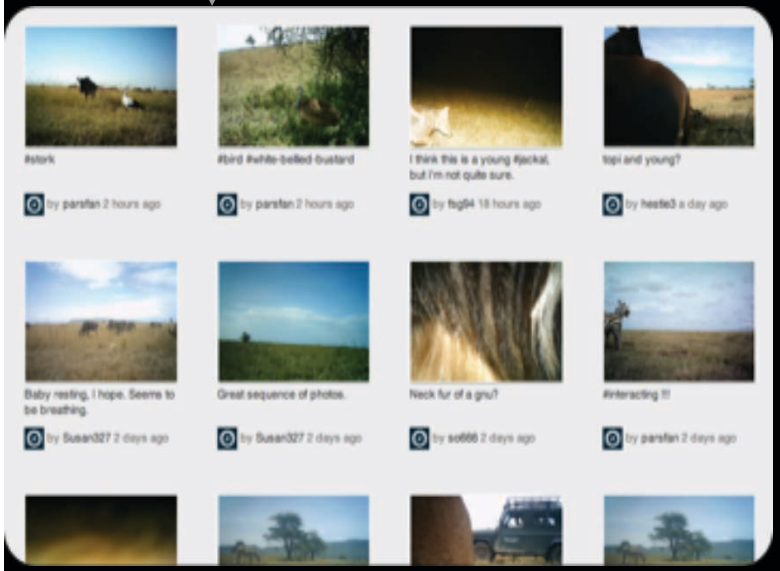
Anatomy of a Zooniverse Project



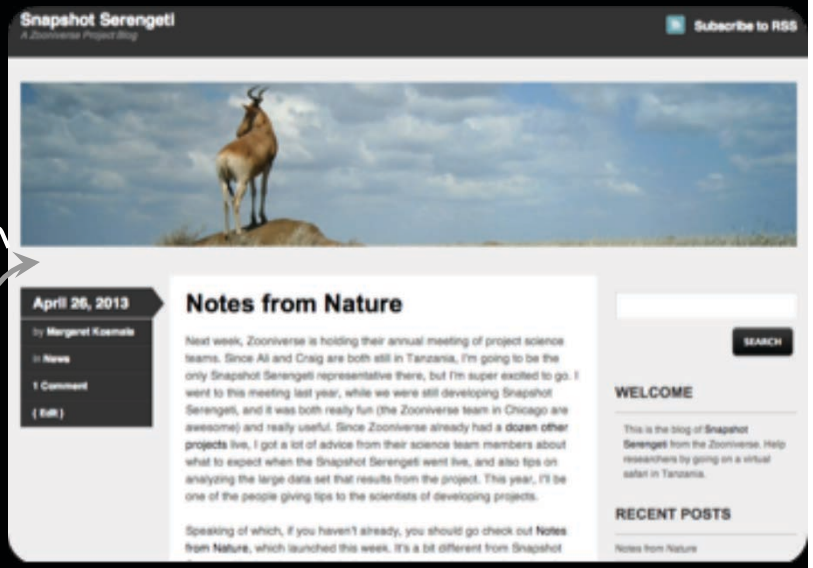


Primary task interface

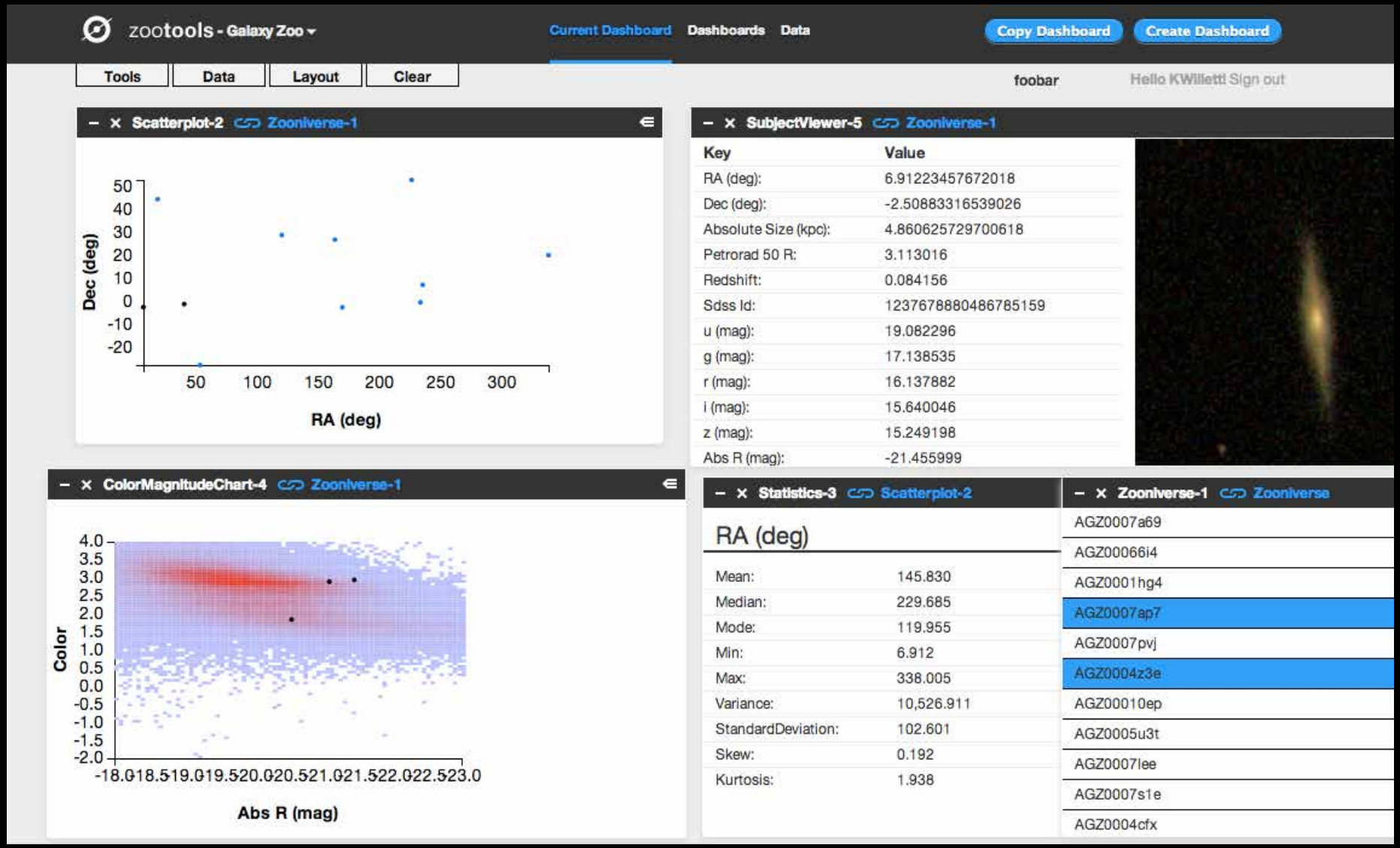
"Talk" – where volunteers discuss their findings and ask for help.



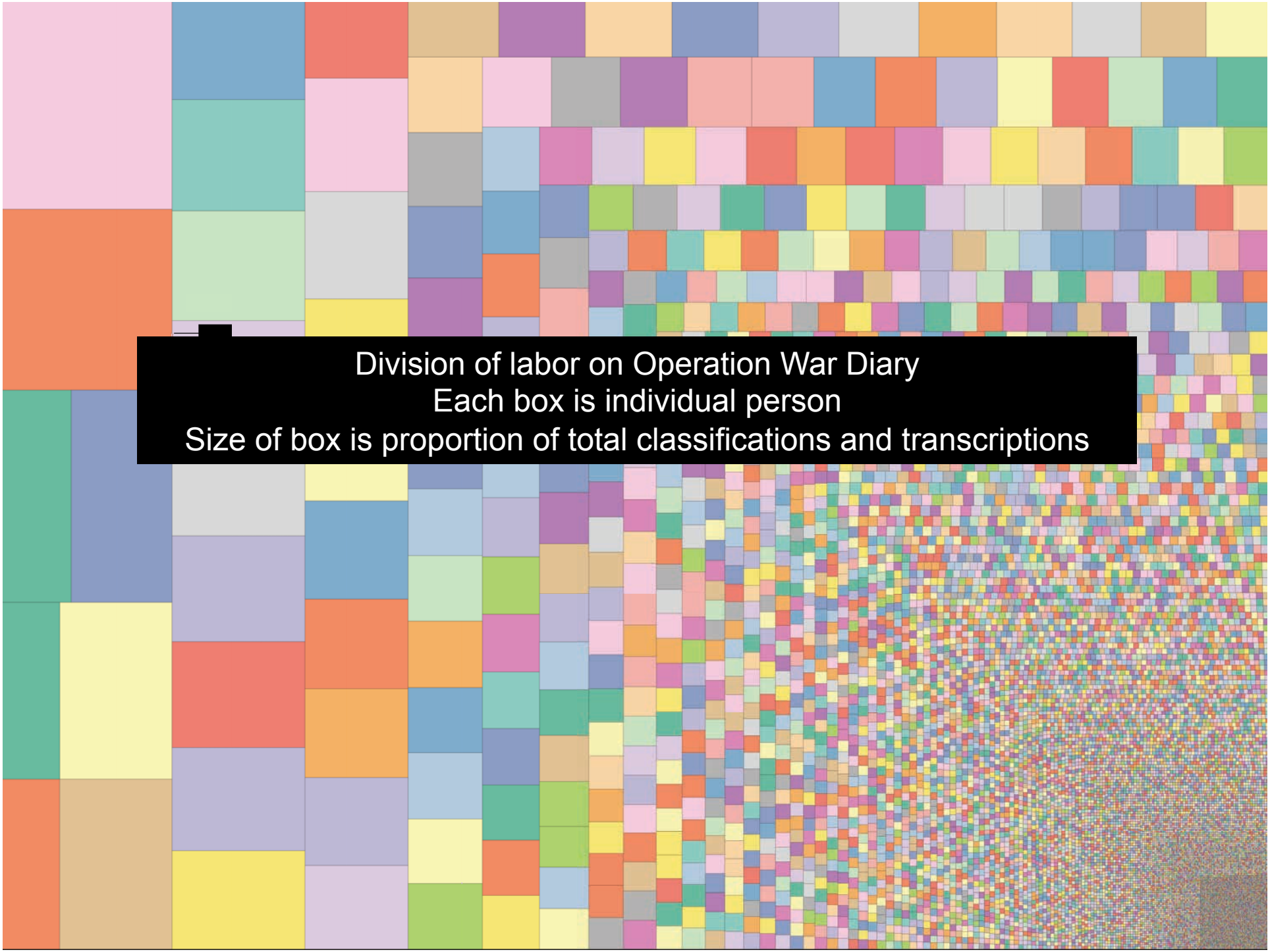
"Blogs" – where volunteers hear from the research team.



Anatomy of a Zooniverse Project



Data analysis tools - enabling volunteers to explore the data:
tools.zooniverse.org


Division of labor on Operation War Diary
Each box is individual person
Size of box is proportion of total classifications and transcriptions

Most of the work being done by relatively small number of citizen scientists!

How do their insights and knowledge reach the research teams?

Classes of Knowledge Discovery



“There are known knowns. These are things we know that we know. There are known unknowns. That is to say, there are things that we know we don't know. But there are also unknown unknowns. There are things we don't know we don't know.”

Donald Rumsfeld

Classes of Knowledge Discovery:

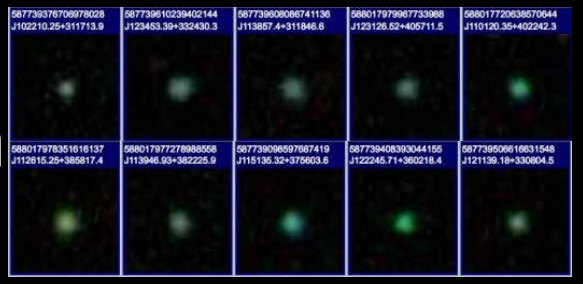
Known knowns : Primary task. Data reduction by science team.



Known unknowns : Related to primary task. Results funneled to specific researchers.



Unknown unknowns : Serendipity. Currently rely on forum moderators to filter.



Accelerating Knowledge Discovery:

- Old-fashioned “linear” forum tool couldn’t support 600,000 people over multiple projects with different levels of engagement discussing their finds
    - Scientists could not easily find serendipitous discoveries posted on the forum
- The social media (communication activities) not totally integrated with data filtering activities.
    - Lack of “easy-transfer” between primary task and forum ended up losing people on both sides

Want to make the most out of user’s desire to conduct their own research and accelerate knowledge discovery.

Key is in communication with and between volunteers:



- Develop "object oriented" discussion tool linked directly from classification task
  - More engaging: integrates social media aspects including tagging, creating and sharing collections, discussing observations with peers and team researchers

"talk.galaxyzoo.org"

- Develop a Citizen Science Journal allowing users to publish their work

"letters.zooniverse.org"

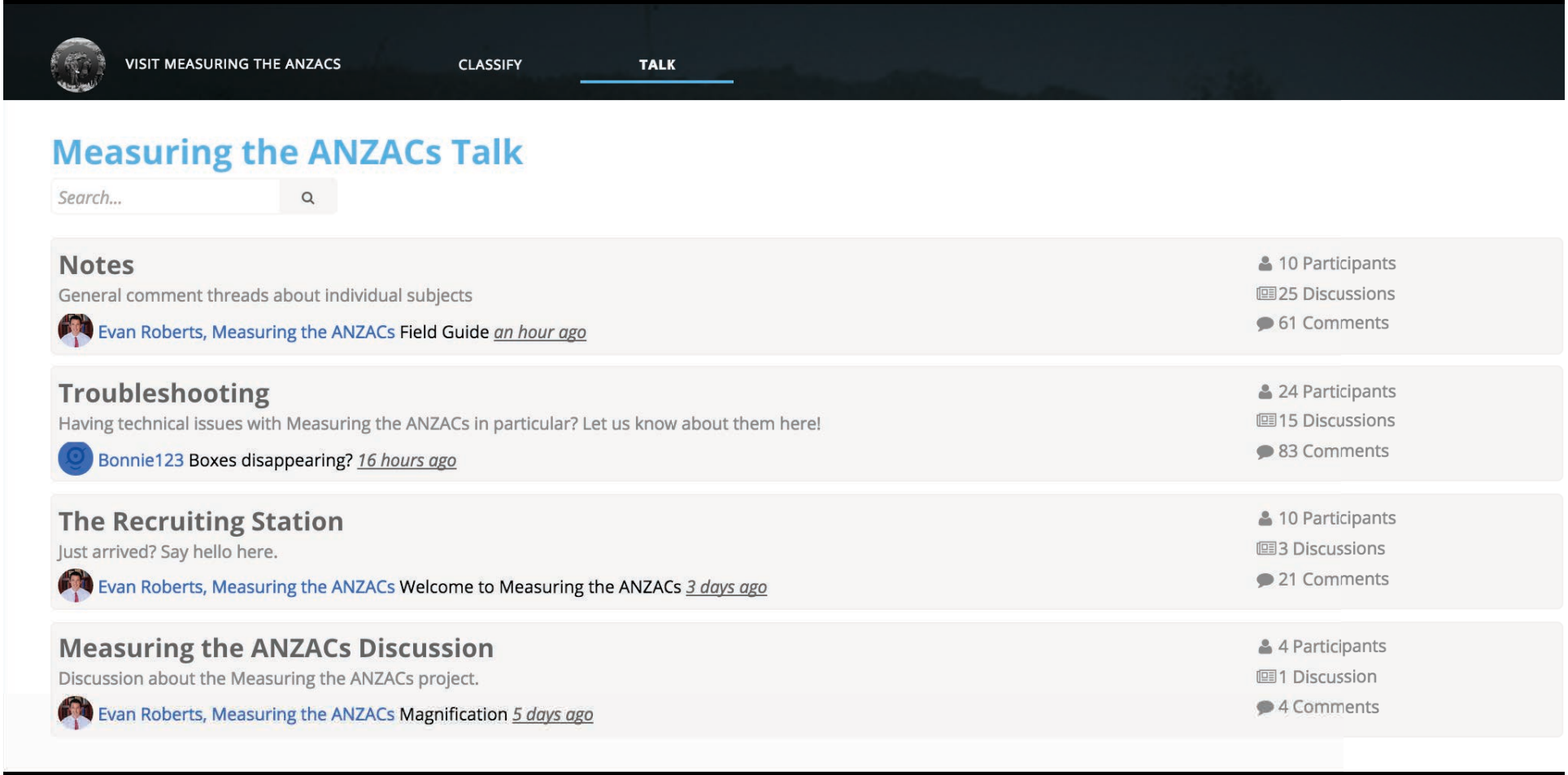
Beta version!

- Develop more advanced volunteer-based research enabling tools

"tools.zooniverse.org"

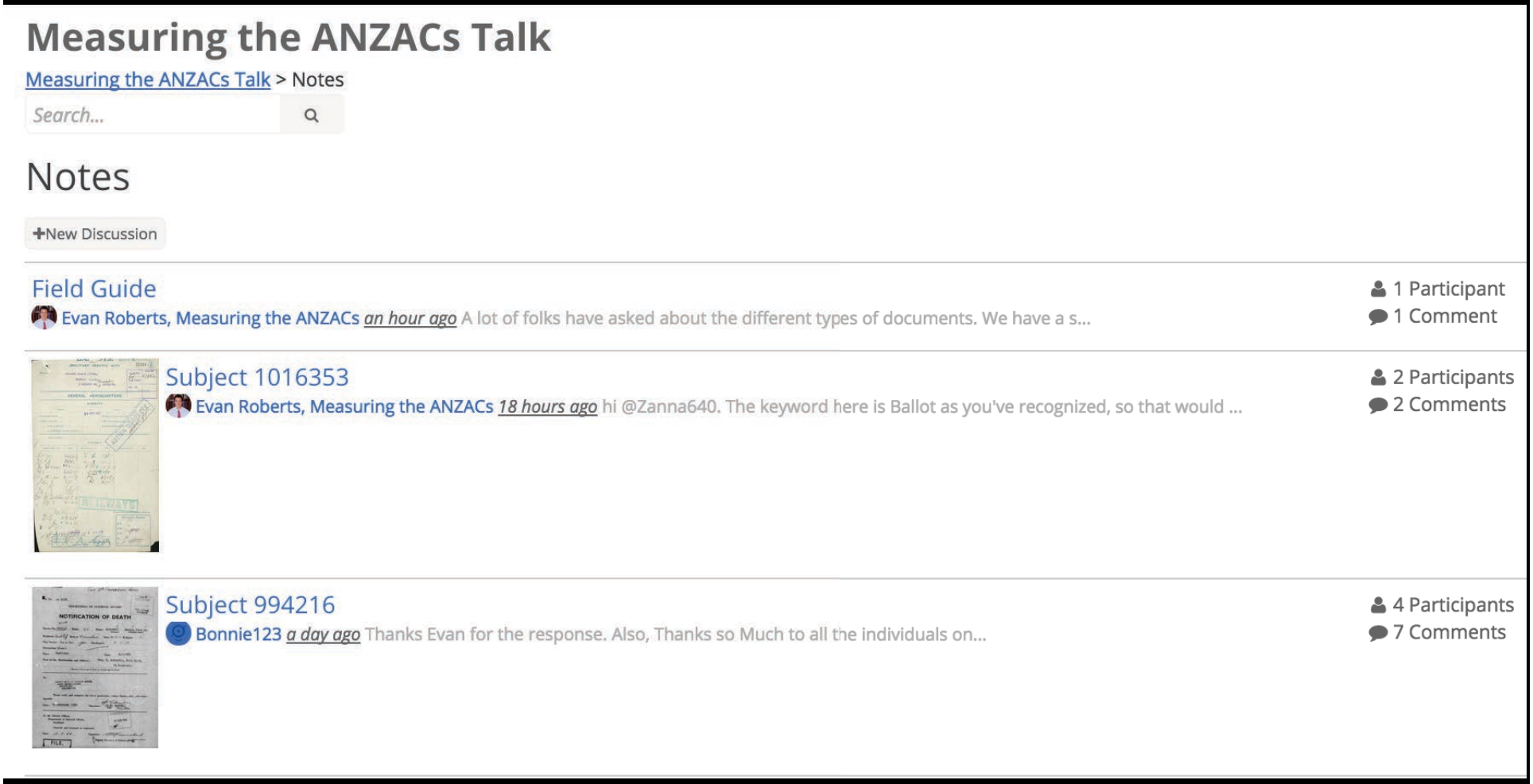
Beta version!

Talking in the Zooniverse



"Talk" entered directly through "talk" website shows boards, collections...

Talking in the Zooniverse



Citizen scientists have asked a lot of questions about the “objects” (pages in personnel files). We’ve been responding!

“Talk” entered through classification interface provides “microblog” or “tweet”-like comment box.

What Zooniverse does and will do

- Large volume of complex images – big data challenge
  - Crowdsourcing - Human processing in combination with machines – is one solution
- Zooniverse: largest crowdsourcing research platform
  - At the same time we do research on the science of crowdsourcing (how to optimize human-machine partnership -> optimizing human attention and cycle of machine learning)
  - Provides education opportunity for millions of members of the public as well as in formal education
- "Talk": tool to engage volunteers in collaborative discovery

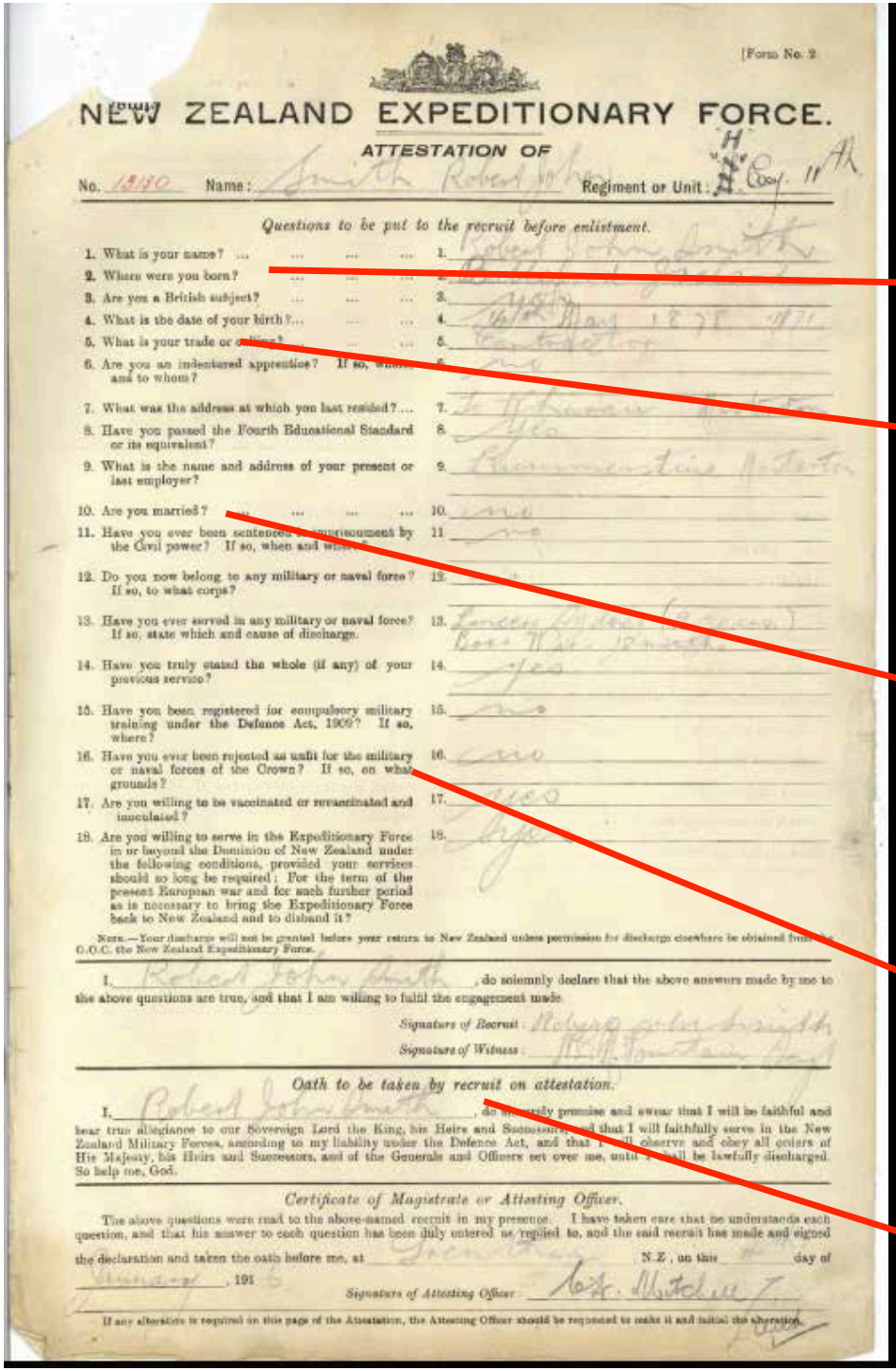
- **Future: the Zooniverse has developed a "Do it Yourself" API that will allow anyone with a good chunk of data to put forth a Zooniverse project as easy as putting out a blog [on a webserver you manage yourself].**

  https://github.com/zooniverse/scribeAPI


ANZAC
MARK
TRANSCRIBE
ABOUT
FIELD GUIDE
BLOG
DISCUSS
LOG IN:
Measuring the ANZACs
Do your part to help transcribe first-hand accounts of New Zealanders from the Australian and New Zealand Army Corps circa World War I.
START MARKING
or
START TRANSCRIBING
Search Records by Name

What Measuring the ANZACs will do!

- Partial transcription of all 140,000 files
  - Millions of images!
  - Multiple pages and forms per file
- Classify page types
- Capture data from attestations (enlistment), service histories, and death notifications
- Timely public history engagement with community
  - Aim: Finish by end of the war's centenary



[Form No. 2

NEW ZEALAND EXPEDITIONARY FORCE.

ATTESTATION OF

No. 1990 Name: Robert John Smith Regiment or Unit: H Coy. 11th

Questions to be put to the recruit before enlistment.

1. What is your name? ... 1. Robert John Smith
2. Where were you born? ... 2. [illegible]
3. Are you a British subject? ... 3. Yes
4. What is the date of your birth? ... 4. 16th May 1892
5. What is your trade or calling? ... 5. Carpenter
6. Are you an indentured apprentice? If so, when and to whom? 6. No
7. What was the address at which you last resided? ... 7. [illegible]
8. Have you passed the Fourth Educational Standard or its equivalent? 8. Yes
9. What is the name and address of your present or last employer? 9. [illegible]
10. Are you married? ... 10. No
11. Have you ever been sentenced to imprisonment by the Civil power? If so, when and where? 11. No
12. Do you now belong to any military or naval force? If so, to what corps? 12. [illegible]
13. Have you ever served in any military or naval force? If so, state which and cause of discharge. 13. [illegible]
14. Have you truly stated the whole (if any) of your previous service? 14. Yes
15. Have you been registered for compulsory military training under the Defence Act, 1909? If so, where? 15. No
16. Have you ever been rejected as unfit for the military or naval forces of the Crown? If so, on what grounds? 16. No
17. Are you willing to be vaccinated or revaccinated and inoculated? 17. Yes
18. Are you willing to serve in the Expeditionary Force in or beyond the Dominion of New Zealand under the following conditions, provided your services should so long be required: For the term of the present European war and for such further period as is necessary to bring the Expeditionary Force back to New Zealand and to disband it? 18. Yes

NOTE.—Your discharge will not be granted before your return to New Zealand unless permission for discharge elsewhere be obtained from the G.O.C. the New Zealand Expeditionary Force.

I, Robert John Smith, do solemnly declare that the above answers made by me to the above questions are true, and that I am willing to fulfil the engagement made.

Signature of Recruit Robert John Smith

Signature of Witness [illegible]

Oath to be taken by recruit on attestation.

I, Robert John Smith, do solemnly promise and swear that I will be faithful and bear true allegiance to our Sovereign Lord the King, his Heirs and Successors, and that I will faithfully serve in the New Zealand Military Forces, according to my liability under the Defence Act, and that I will observe and obey all orders of His Majesty, his Heirs and Successors, and of the Generals and Officers set over me, until I shall be lawfully discharged. So help me, God.

Certificate of Magistrate or Attesting Officer.

The above questions were read to the above-named recruit in my presence. I have taken care that he understands each question, and that his answer to each question has been duly entered as required by, and the said recruit has made and signed the declaration and taken the oath before me, at [illegible] N.Z., on this [illegible] day of January, 191

Signature of Attesting Officer [illegible]

If any alteration is required in this page of the Attestation, the Attesting Officer should be requested to make it and initial the alteration.

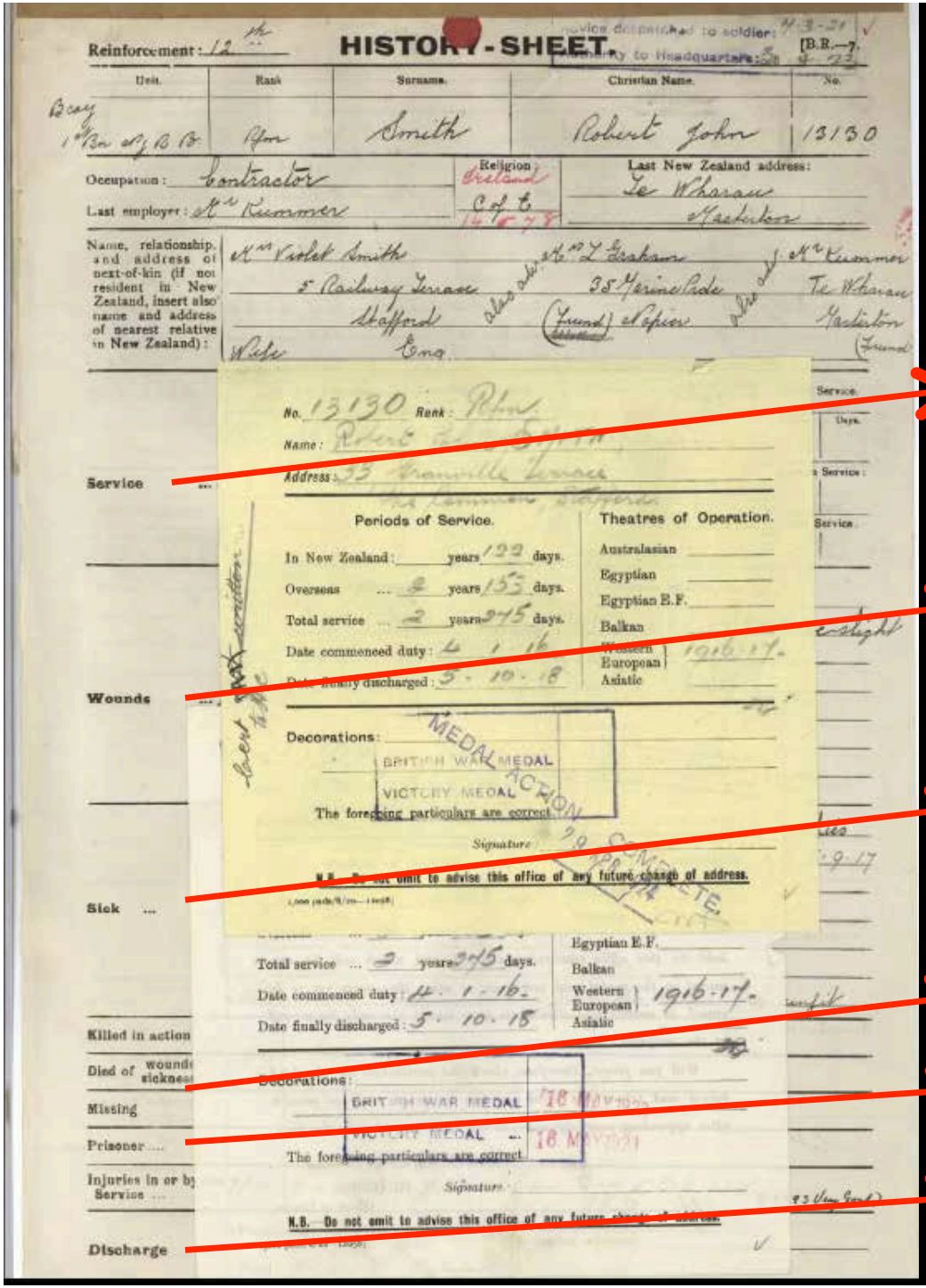
Who are you and where are you from?

What's your job?

How can we contact Mum & Dad or your Missus when you die?

Ever served before?

Do you swear to obey orders?



Service (Where did he go? What did he do?

Wounds

Sickness

Death

Prisoner

Discharge

Measuring the ANZACs data structure

- 147 variables/fields/columns from source material
  - Some variables describe categories of events that can occur multiple times
    - Sick, wounded, transfers between units, entering new combat zone
    - You only die once
- 6 different document types to identify
- Paradata : Data about the transcription of each element
  - 17 elements of paradata about each transcription or classification! Includes who, when, X/Y co-ordinates on page
- Significant automatically generated metadata

STATEMENT of the SERVICES of Tremain (Surname.) William Hebdite (Christian Name.) No. 8/4234

Regiment or Corps.	Promotions, Reductions, Casualties, &c.	Rank	Date.	Authority for Entry.
D Coy. 11th Reinfts. Infty.	Booked to	Pte	24/2/15	R/O 380
1st Res Coy	Transf to	"	11/3/16	" 171
D Coy. 11th Reinfts. Infty.	"	"	14/3/16	" 174
1st Bltn. Otago Reg.	Joined Draft Posted 4th Coy	Pte	18.7.16	P.C. 17 Dn No 23. Rouen 1/8/16 B.R. 27/10
N.Z. Command depot Codford	Taken on strength depot from 2.2. Con Hosp Wdcht & posted V.D. Coy.	Pte	10.4.17	Pro Dec C.D. Cod. 10.4.17 A 48.
1st Bn. Otago Regt. 4th Coy.	Emb. for N.Z. per 17. S.S. Marama	Pte.	14.7.17.	D.O. 30. Rouen Trans 2.8.17

CONDUCT-SHEET.

Regiment or Corps.	Place.	Date.	Offence.	Punishment.	Authority for Entry.

Home Service.	Years.	Days.
		192
Foreign Service:	1	146
Total Service:	1	338

Reinforcem
4th Coy 1st Ba
D Coy. 11th Re
Occupatio
Last empl
Name, re and ad next-of-ki resident Zealand, name an of neares in New Z
DIS
SINK
DISCH
Service
No.
Wounds ...
Sick ...
4th Con 1st Otago N.Z.R.
Killed in action ...
Died of wounds* sickness*
Missing ...
Prisoner ...
Injuries in or by the Service ...
illness
Discharge ...
Provisio
Final :
Pension ...
Please note if you have permanency
* Strike out words not required.

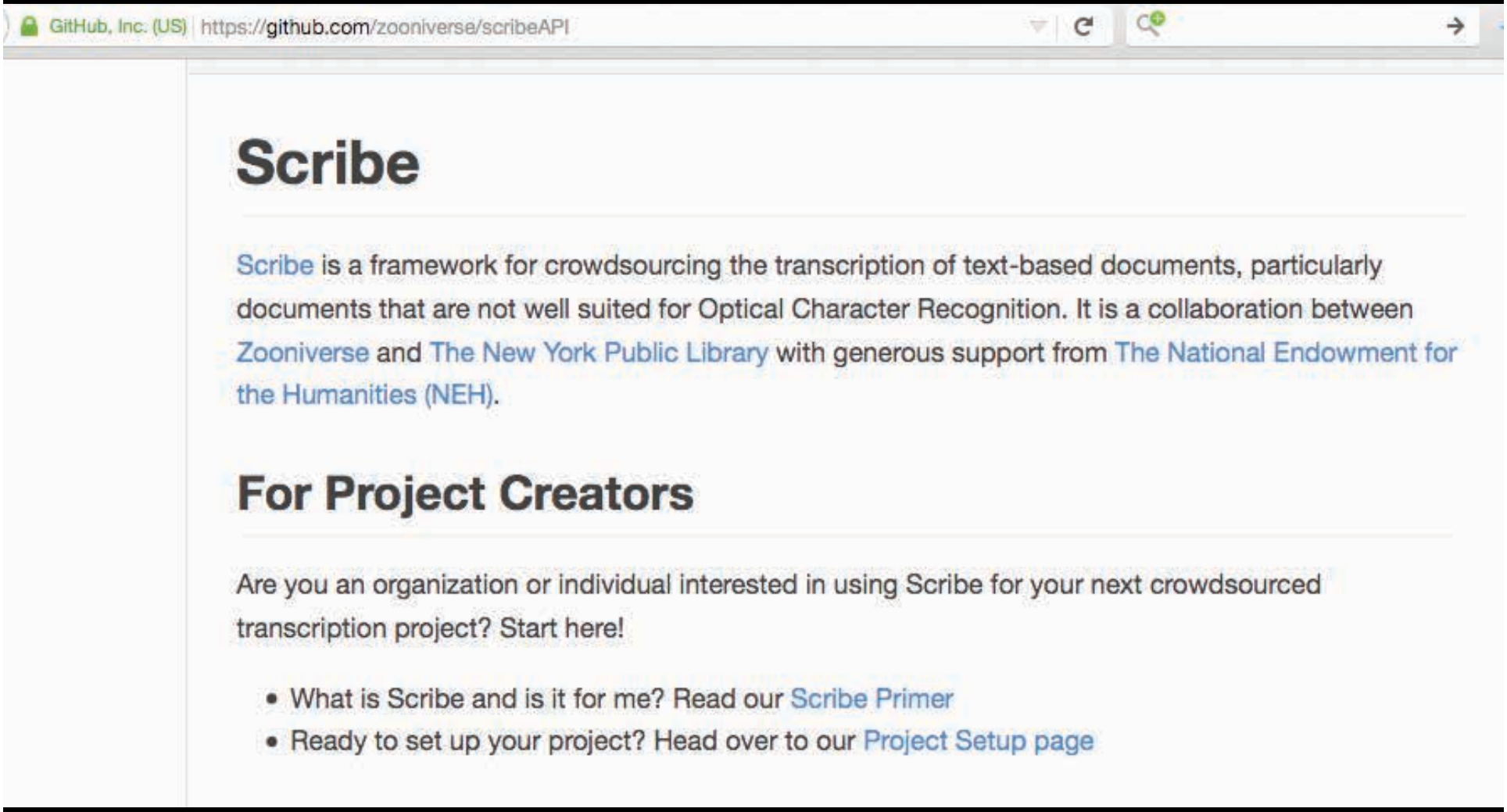
Future questions for our research team

- Effect of short-run events on long-term health
  - Russian flu epidemic (1889)
  - 1893 depression
  - Did the 1893 depression have different effects in different places
- Expanded sample linked to vital records
  - With history sheets, effect of wounds on later life health
- Link British-born troops to existing collections of complete British census records
- Better identification of Māori in sample with names of next of kin

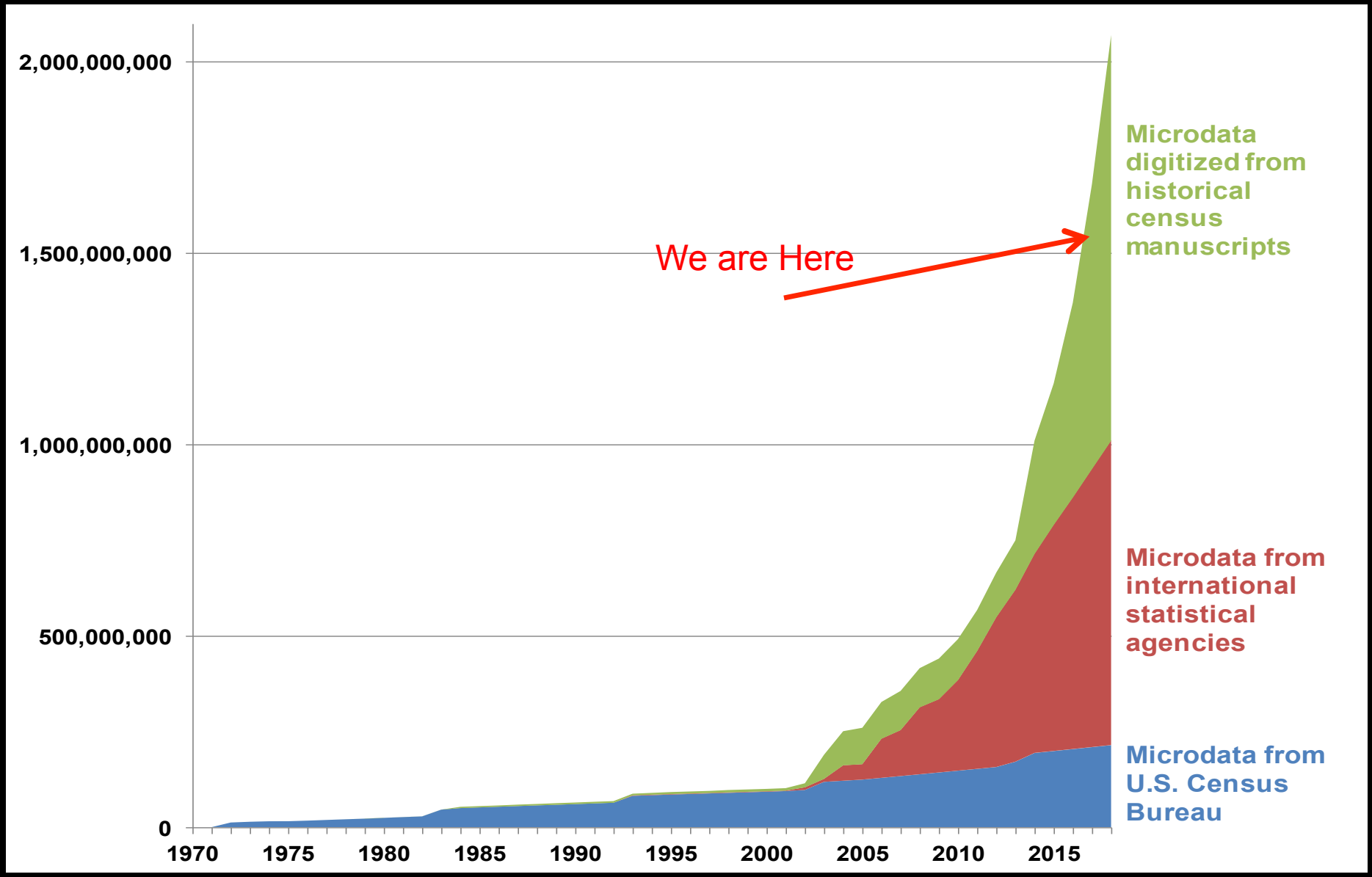
Potential questions for other researchers

- Migration - where were people born, enlist, where were their parents?
  - Imperial and local connections
- Social networks. Who served with whom and how were they connected?
- Women in the war: 626 nurses + other auxiliary health care aides
- Māori / Pākehā inter-marriage – limited marital registration of Māori before 1913.
- Descriptions of the body: Tattoos, wounds
- Deviance and discipline from History sheets
- Structure of military careers

Potential for more demographic research



Growth of public-use census microdata



DEPARTMENT OF COMMERCE—BUREAU OF THE CENSUS

SIXTEENTH CENSUS OF THE UNITED STATES: 1940

POPULATION SCHEDULE

State: New York — County: Bronx — Incorporated place: Bronx Borough (~~New York~~) — Ward of city: — — Township or other division of county: A.D. 1 — Block Nos. A B C D F — Unincorporated place: — Institution: —

S.D. No. 22 — E.D. No. 3-1462 — Sheet No. 1 A — 1414

Enumerated by me on April 2, 1940. Vincent Moynihan, Enumerator

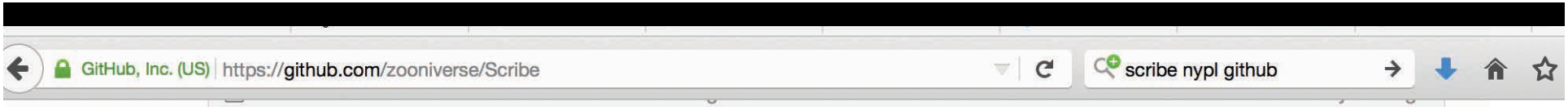
Line	Street	House No.	HH No.	Own/Rent	Value/Rent	Farm	Name	Relation	Sex	Race	Age	Marital	School	Grade	Birthplace	Citizenship	1935 City	1935 County	1935 State	On farm	21	22	23	24	25	Hours	Weeks unempl.	Occupation	Industry	Class	Weeks worked	Income	Other income
1	Walnut Ave	103	1	R	38	No	Granfeldt, Carl	Head	M	W	51	S	No	H-4	New York		Same House			No	Yes	—	—	—	—	40		Timekeeper	Cut Stone Yard	PW	52	2280	No
2							—, Arvid	Brother	M	W	48	S	No	8	New York		Same House			No	Yes	—	—	—	—	40	25	Helper	Cut Stone Yard	PW	46	1,165	No
3							—, Sigrid	Mother	F	W	83	Wd	No	8	Sweden	Na	Same House			No	No	No	No	No	H						0	0	No
4							Here ends Block A and B and begins Block C. → "No dwelling on block B"																										
5	East 134th St	825	2	R	40	No	Falco, Anthony	Head	M	W	50	M	No	H4	Italy	Na	Same House				Yes	—	—	—	—	48		Superintendent Loft Building	Phillip Jones Shirt Company	PW	52	2400	No
6							—, Jennie	Wife	F	W	48	M	No	H4	Italy	Na	Same House				No	No	No	No	H								
7							—, Nicholas	Son	M	W	22	S	No	H4	New York		Same House				Yes	—	—	—	—	40		Stock Clerk	Shirt Mfng Company	PW	52	1040	No
8							—, Theresa	Daughter	F	W	19	S	No	H4	New York		Same House				No	No	Yes	—	—		20	Switchboard Operator	Shirt Mfng Company	PW	45	600	No
9							Here ends Block C-D- and E and begins Block F → "No dwelling on Block D and E"																										
10		743	3	O	1350	No	Ferrari, Joseph	Head	M	W	64	M	No	2	Italy	Na	Same House				No	No	No	No	U						0	0	Yes
11							—, Elizabeth (X)	Wife	F	W	68	M	No	2	Italy	Na	Same House				No	No	No	No	H						0	0	No
12		743	4	R	28	No	Imor, Augustus	Head	M	W	33	M	No	8	New York		Same House				Yes	—	—	—	—	44		Truck Driver	Wholesale Hardware Co	PW	52	1400	No
13							—, Ida	Wife	F	W	31	M	No	8	New York		Same House				No	No	Yes	—	H		39	Stenographer	Office Work	PW	26	1352	No
14							—, Augustus	Son	M	W	7/12	S	No	—	New York																		
15		743	5	R	28	No	Barbiere, Fiorenzo	Head	M	W	50	M	No	8	Italy	Na	Same House				No	No	Yes	—	—		13	Butcher	Cafe & Restaurant	PW	52	1456	No
16							—, Mary	Wife	F	W	41	M	No	8	Italy	Na	Same House				No	No	No	No	H						0	0	No
17							—, William	Son	M	W	17	S	Yes	C1	New York		Same House				No	No	No	No	S						0	0	No
18							—, Fred	Son	M	W	16	S	Yes	H3	New York		Same House				No	No	No	No	S						0	0	No
19		743	6	R	28	No	Riguiri, Ralph	Head	M	W	28	S	No	8	New York		Same House				Yes	—	—	—	—	40	39	Chauffeur	Main Tenance Dept. N.Y.C.	GW	26	1000	No
20		743	7	R	28	No	Anderson, Edward (X)	Head	M	W	29	M	No	8	New York		Waterford	Sullivan	New York		Yes	—	—	—	—	48		Fireman	N.Y. City Fire Dept.	GW	52	2000	No
21							—, Celeste	Wife	F	W	29	M	No	H-1	New York		Same House				No	—	Yes	—	—	2	38	Receptionist	Beauty School	PW	27	700	No
22		743	8	R	28	No	Berry, Mathew (X)	Head	M	W	40	M	No	8	New York		Same Place	Bronx	New York		Yes	—	—	—	—	40		Plumber	New York City	GW	52	1860	No
23							—, Margaret	Wife	F	W	30	M	No	8	New York		Same Place	Bronx	New York		No	No	No	No	H						0	0	No
24							—, Margaret	Daughter	F	W	18	S	No	8	New York		Same Place	Bronx	New York		Yes	—	—	—	—	44	32	Assembler	Zipper Mfng Co	PW	33	308	No
25	East 136th St	739	9	O	6000	No	Tenke, Emory (X)	Head	M	W	58	M	No	8	Yugoslavia		Same Place	Bronx	New York		Yes	—	—	—	—	44		Baker	Restaurant & Cafe	PW	52	1050	No
26							—, Gazella	Wife	F	W	46	M	No	8	Germany		Same Place	Bronx	New York		No	No	No	No	H						0	0	No
27							—, Emory Jr	Son	M	W	20	S	No	H-2	New York		Same Place	Bronx	New York		Yes	—	—	—	—	42	32	Delivery Boy	Underwear Factory	PW	26	240	No
28							—, Dazia	Son	M	W	18	S	No	H-2	New York		Same Place	Bronx	New York		No	—	Yes	—	—	65		Filler	Soda Water Factory	PW	0	0	0
29							—, Vilma	Daughter	F	W	16	S	No	8	New York		Same House				No	No	No	No	U						0	0	No
30							Seifert, Yolanda	Daughter	F	W	22	M	No	8	New York		Same House				Yes	—	—	—	—	40		Assembler	Zipper Factory	PW	0	0	No
31							—, William (X)	Son in Law	M	W	31	M	No	8	New York		Same Place	Bronx	New York		No	—	Yes	—	—	43	13	Elevator Opr	Loft Building	PW	52	1040	No
32							—, Priscilla	Grand Daughter	F	W	7/12	S	No	0	New York		Same House														0	0	No
33		737	10	R	32	No	Martin, Thomas (X)	Head	M	W	30	M	No	H3	Scotland	PA	Same Place	Bronx	New York	No	Yes	—	—	—	—	40		Oiler	Edison Electric Co.	PW	52	2200	No
34							—, Francis	Wife	F	W	30	M	No	H3	Scotland	AL	Bloomfield	Hudson	New Jersey	No	No	No	No	No	H						0	0	No
35							—, Janet	Sister	F	W	22	S	No	H3	Scotland	AL	New York	Bronx	New York	No	Yes	—	—	—	—	48		Salesgirl	Dept Store	PW	52	700	No
36							—, Mathew	Brother	M	W	19	S	No	H4	Scotland	AL	Same Place	Bronx	New York	No	No	—	Yes	—	—		64	New Worker			0	0	No
37							—, Isabel	Mother	F	W	57	Wd	No	8	Italy	Na	Same House				No	No	No	No	H						0	0	No
38		735	11	O	4,000	No	Maurione, Carlos	Head	M	W	65	M	No	8	Italy	Na	Same House				No	No	No	No	U						0	0	No
39							—, Mary	Wife	F	W	52	M	No	0	Italy	AL	Same House				No	No	No	No	H						0	—	No
40							—, Joseph (X)	Son	M	W	29	M	No	10	New York		Same House				Yes	—	—	—	—	40	32	Truck Driver	Excavating Concern	PW	32	520	No

Margin note (beside lines 15–18): Formerly finished Mrs. Ferrari. Other.

SUPPLEMENTARY QUESTIONS

For Persons Enumerated on Lines 14 and 29

Line	Name	Birthplace of Father	Birthplace of Mother	Code	Mother tongue	Veteran	Wife/widow of veteran		Social Security No.	Deductions		Usual occupation	Usual industry	Class of worker
14	Imor, Augustus	New York	New York		English	No	No	No	No	—	—			
29	Tenke, Vilma	Yugoslavia	Germany	17	English	No	No	No	No	—	—	None	Student	S.



Rakefile	Demo	4 years ago
config.ru	first commit	5 years ago
license.txt	license.txt	4 years ago

README.md

Welcome to Scribe

Scribe is a framework for generating crowd sources transcriptions of image based documents. It provides a system for generating templates which combined with a magnification tool guide a user through the process of transcribing an asset (an image).

Getting Started

- We first need to install a mongodb server. This is then specified in config/mongodb.yml (see config/mongodb.hudson.yml for an example). Since databases are created lazily in MongoDB just specify the database name you want to use there.
- Site settings (config/site_settings.hudson.yml) contains the application name and other detail about the project. You should rename site_settings.hudson.yml to site_settings.yml
- To generate the templates for the project look at the lib/tasks/sample_weather_bootstrap.rake file. You need to specify each entity type you wish transcribed and its fields along with help text for the user for each.