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



MARK TRANSCRIBE

DISCUSS LOG IN:

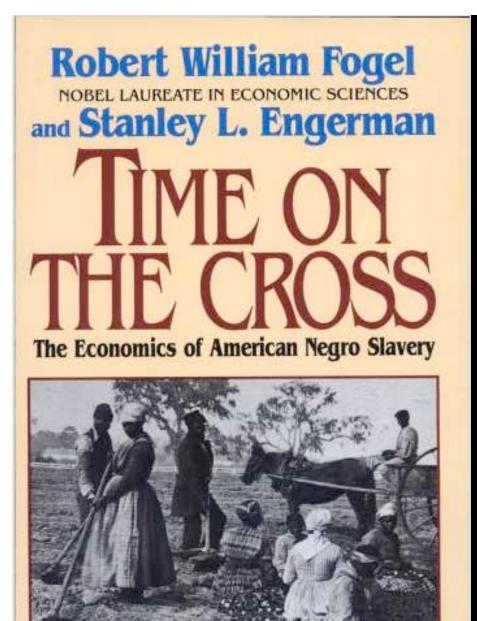
Measuring the ANZACs

Do your part to help transcribe first-hand accounts of New Zealanders from the Australian and New Zealand Army Corps circa World War I.



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

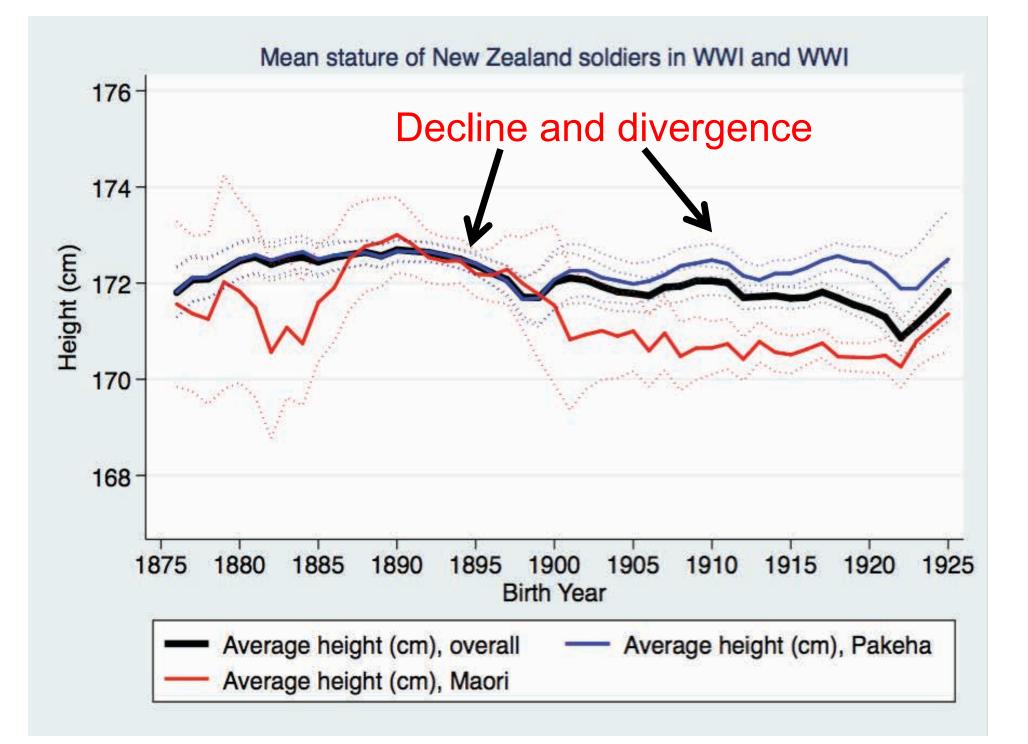
Background and motivation

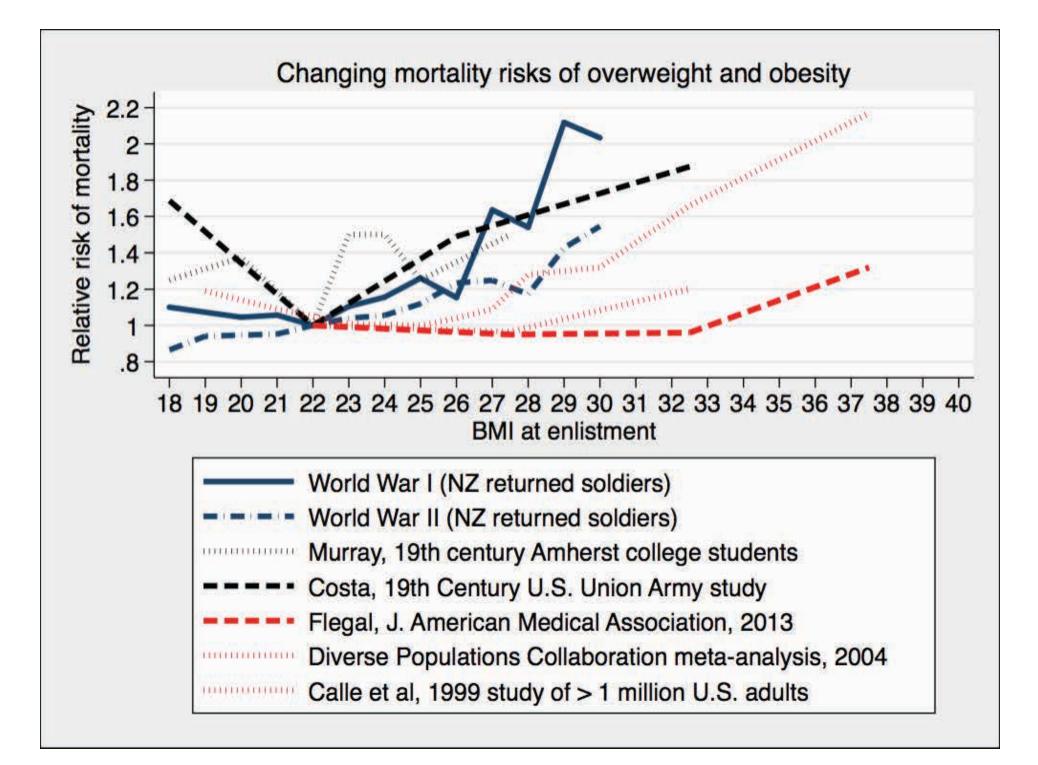
NZ anthropometric history

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





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

ZOONIVERSE

Blog

Daily Zooniverse

Sign inRegister

BUILD A PROJECT

Ø

Projects

A Com

About

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?



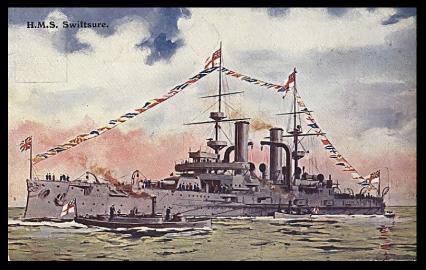


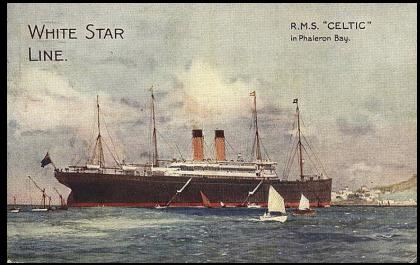
And what do they have to do with these?



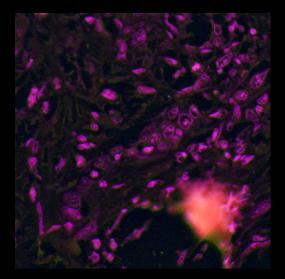


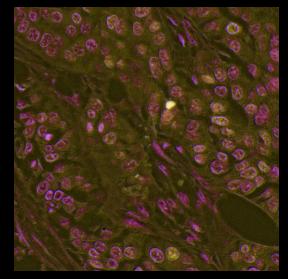
Or in fact these?





Or these?





Or even these?

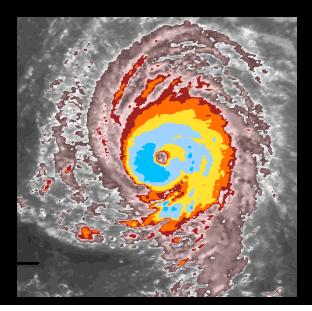


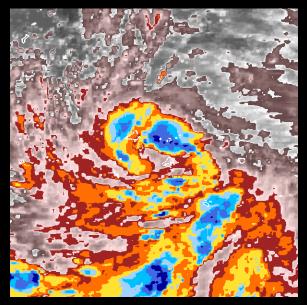
MUNTER MUTTE CTU OT WORK WORTCOM CNXCHICANTACINO ownerst Antone was for CONTOTICATION ornersanie Tom ລັດຂອງຄະຫາງທີ່ແຜ່ YROTON DIKASTON TINXOCNTOICICA CONTROLUTINGS THE CYALCONOT CTIMINOY אינטורידו אישאאי CONTRACTORES SOMTONTIAS ST. MOUNT CHUTOCTOWNA - MUNI אמות הואטנטריין CTTU ANCTOTING 1324 TOTOMT POW TUN CXCLONDITIO -5770 TIONE OBO 125 63 TREPOCINETT FOR TOWNERS CTG בואל בואון

TOYLICTTOLOY, YOUNG JUNCON ANKACTAN ON TROCTLO PROSTING or more control and Some Come Comon 28 TOURSON CONVISION BIONCINIMTOMITES NONKNTHINTTIN doutherrownie whe איזירנאנים, האיזיראוים CANKADTINA PTCONCO-TIONCTINCIANTONIAS THANYCECEMINEN TOIBLOWNNKANNOIS ANANTHOTONOKIC-NALLICHUNDINOS CO. KNUTHER CARONEAL THOMADINCKCIUG TUNHALTONTHIAT NUMBER TOUNTROOMON CA THENTECTONS JUON KALINACUM ONTECTORER; FORTIN WAR CTOCHON THEO

LUKTADOWTIACION RNOT TOTAKON MOS OCTICCICTOTIACAXIN MAN TO MAN CALCOLAL KATCHNCHMCHAR יייריוט באפורטוטיוויי דישווומדמונות אמצט ANTOINCTOTONCOLO TAXYONONWOORNING THIMONORALITICOLO NOISC TWINTIONING Antreaganethe LOUGTCAL UNICOUNT ON TON TEOTON AN 24TOINCPRATILION 20001012. CONTROPORT φλαιτοιπονατισμασμ AROTIANCATTUTEA TEMACECHARDNICCI OT MORAL PHOP RAN TORTHERMONT CON TICEOPARCEN EINDONEYCH TOY TONCUCHONNIQUO OTALMONCOTIRSAN CNITATIATIATIAT are summer and

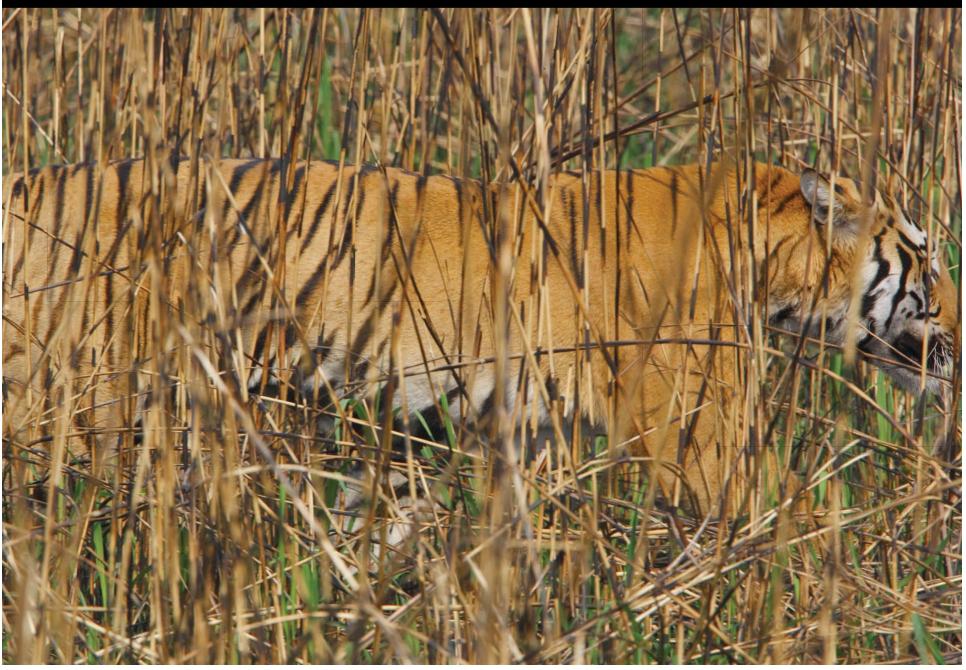
Or these?

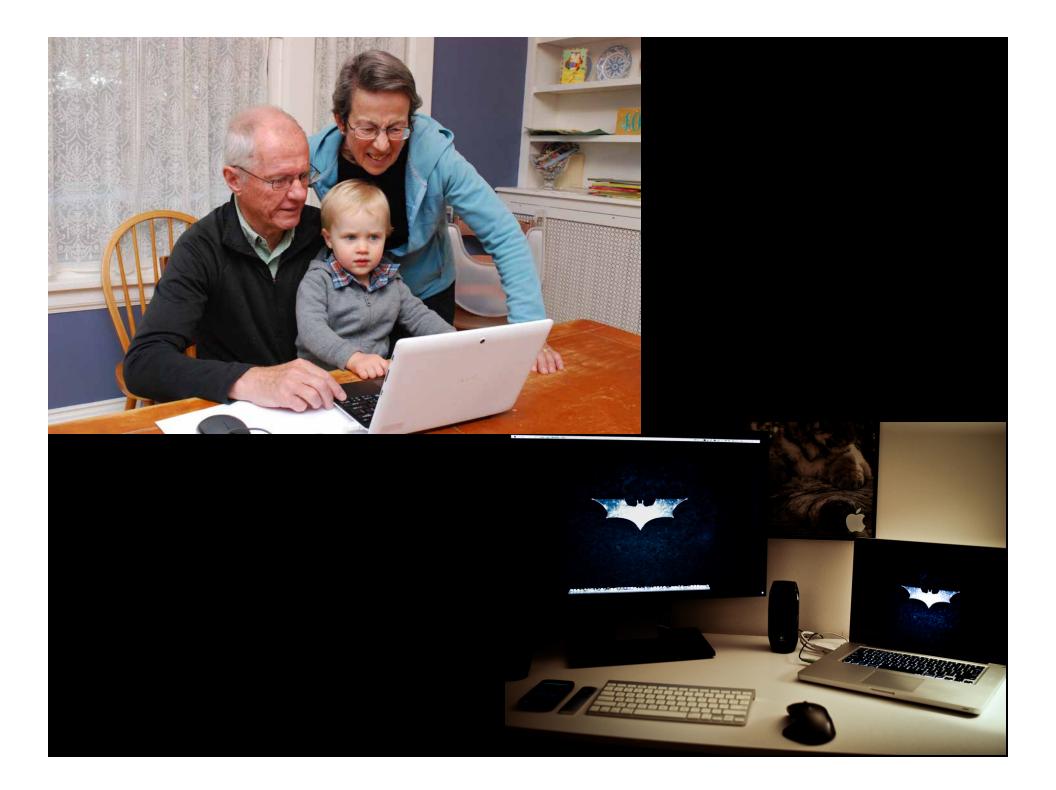




These are all images from research projects that rely on analysis of complex images and pattern recognition capabilities that are currently beyond computer algorithms!

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

6 🧳 🤔 🍠

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

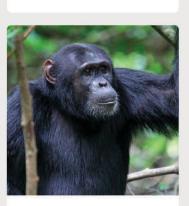
Help keep an eye on changing seasons by marking images!

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.





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



Ellipticals

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

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?

1,1050

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.

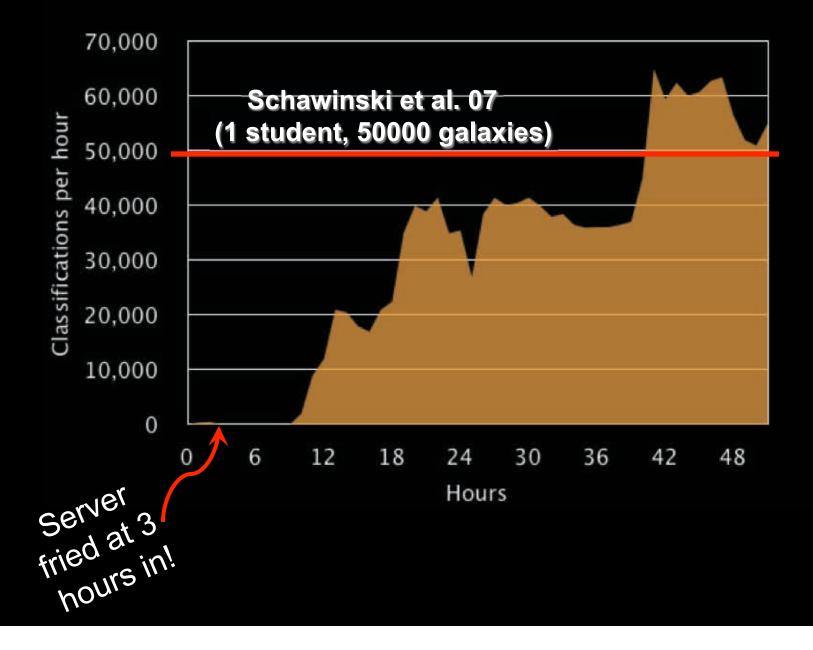
Try your hands at some! Click the image to see if you're right.

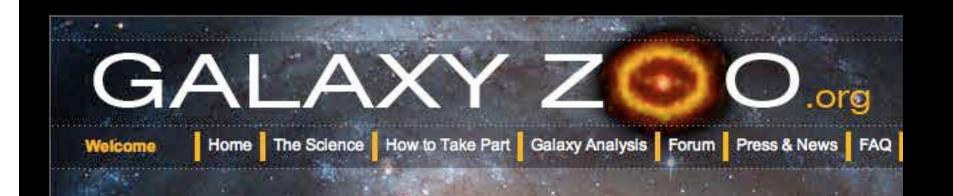


Part 1C ... Merging Galaxies

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...

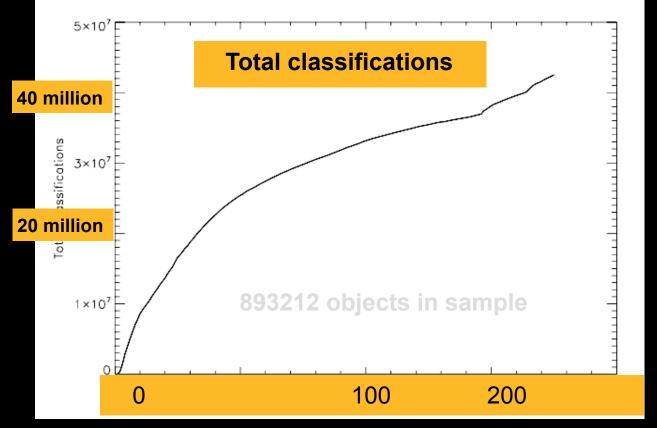






Public outreach and media presence to engage broad and large audience

First six months



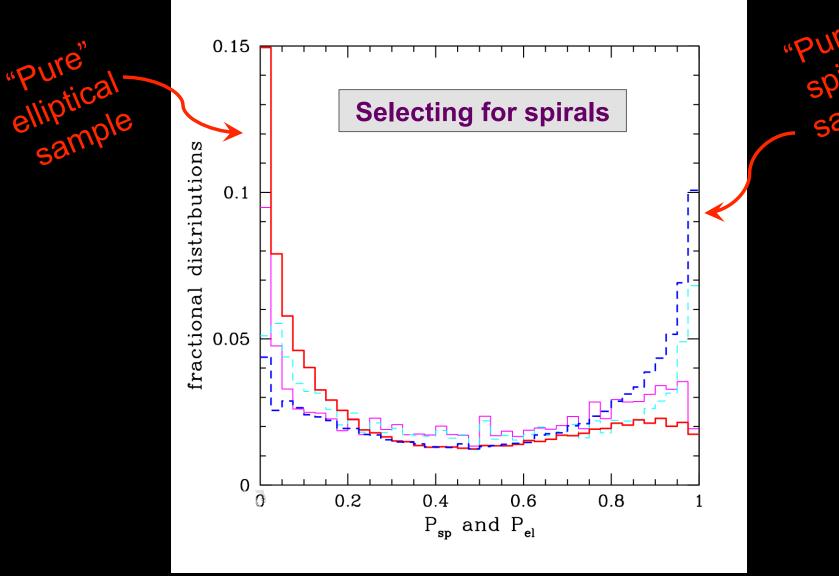
Days since launch

After 'cleaning' raw clicks:

34,617,406 classifications by 82,931 users

Roughly 3.3 continuous person-years!

Combining clicks



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?

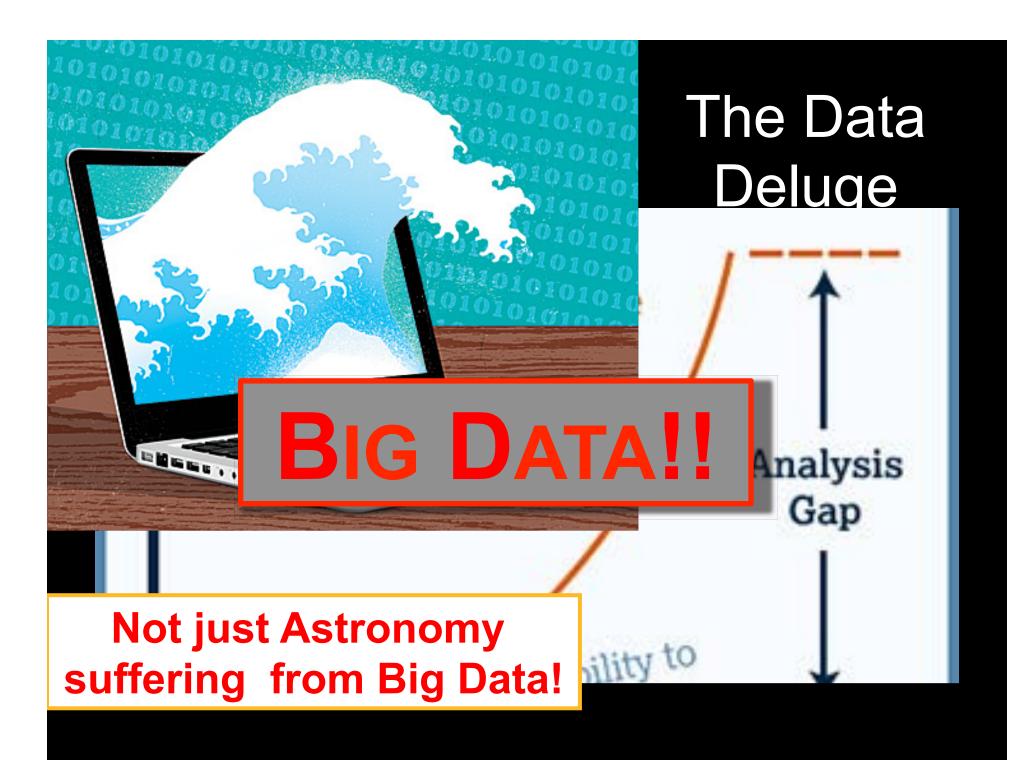
- Largest professional classification is only ~5% of SDSS
- Galaxy Zoo provided 3.3 person years in first 6 months

(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?



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

- 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

ZOØNIVERSE

Blog

0

Projects

About

Alara Ag

Talk

Collect

Daily 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

WH LEFM

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.



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

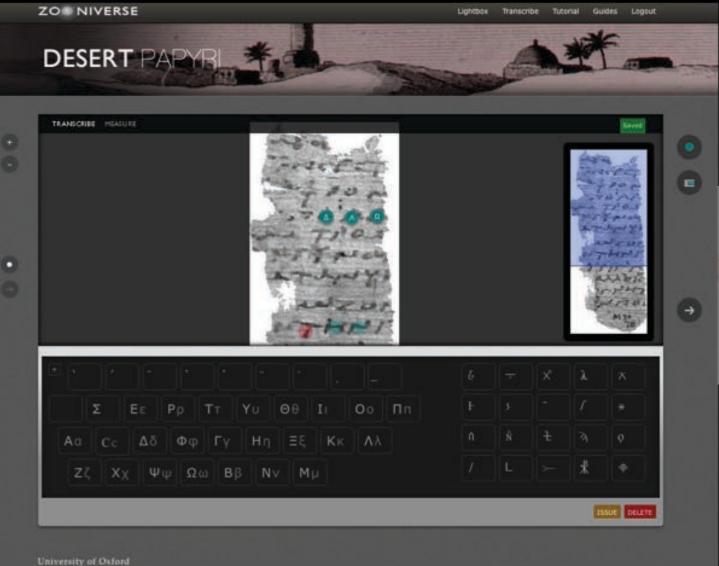
UDUR AUT TO	H	.M.S.	"	Pegarus TR. ", Thursday									2	2-day of					
		From			and	ha	-ge	Ŀ	,	To		T)~	~~	Lee or At				
	- 16	Patent	Dista		True	ons ate	Win	d		the Sea	Height of Barometer and	Ten	T	ature		osition 0800			
10	Time	Log	Milles	Tenths	Course	Revolutions per minute	Direc- tion	Force	Weather	State of the Sea	attached Thermo- meter	Air	Wet Bulb	Sea	2	2000			
-					1						-	1							
0	100	1399	10	8	3516	161									-				
0	200	1410.1	11	1		-									2.3	Dex.	se		
0	300	14-22.5	12	4	<u>_u</u> _	64					-					1			
0.	400	1433.2	10	7		4	3+8	8	oce	6	2.9.81	52	50	4-2	4.5	3	-		
0	500	1445.6	12	4	-11-	14											-		
00	300	1456 - 0	10	4		-									6.0	Hand			
07	7.00	1466.2	10	2							2011								
08	300	1479	12	8.	-4-	"	5×8	8	2e.9.9	b	29.66	57	50	41	8.0		-		
09	000	d Weatl	11	3	33W	iĥe	the l	100				AC		5		c S 2 5 Sher	100		
10	tra	vel the	g	bbe	, colle	ecti	ng v	alu	iab	le	climat	e	dat	a	10.22	Shen	Jur.		

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.



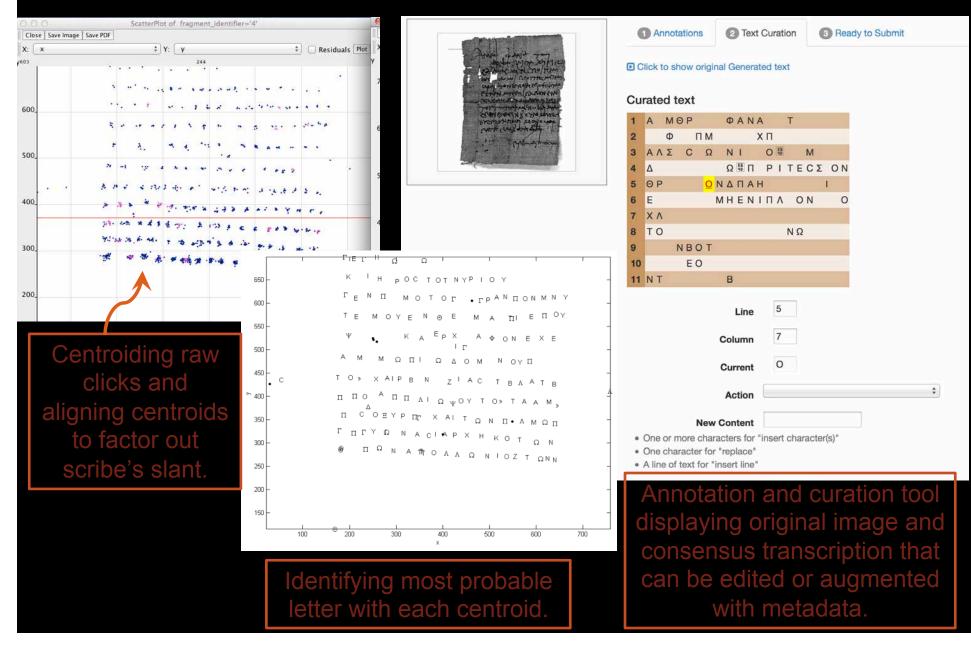
300,000 volunteers

100m





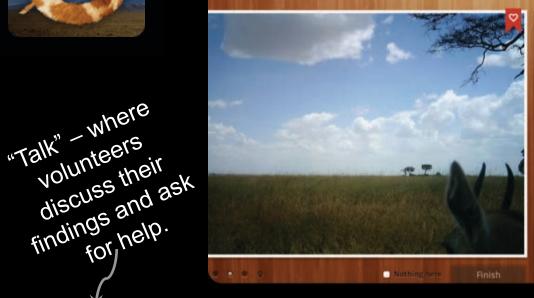
Ancient Lives Project: UMN developed the pipeline to turn clicks into consensus into transcribed text that can be edited & curated.





Anatomy of a Zooniverse Project

SNAPSHOT SERENGETI



Looks like +	٩						
Pattern Colo	r Horns	Tail Bu					
Aardvark	Giraffe	Porcupine					
Aardwolf	Guinea fowl	Reedbuck					
Baboon	Hare	Reptiles					
Bat-eared fox	Hartebeest	Rhinoceros					
Bird (other)	Hippopotamus	Rodents					
Buffalo	Honey-badger	Secretary bird					
Bushbuck	Hyena (spotted)	Serval					
Caracal	Hyena (striped)	Topi					
Cheetah	Impala	Vervet monkey					
Civet	Jackal	Warthog					
Dik dik	Kori bustard	Waterbuck					
Eland	Leopard	Wildcat					
Elephant	Lion (female or cub)	Wildebeest					
Gazelle (Grant's)	Lion (male)	Zebra					
Gazelle (Thomson's)	Mongoose	Zorilla					
Genet	Ostrich	Human					

Snapshot Serengeti











O by parafan 2 hours ago







Great sequence of photos

Baby resting, I hope. Seems to be breathing O by Susan327 2 days ago

by Busar327 2 days ago







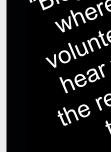


O by parafan 2 days ago

nteracting 11

topi and young?

o by heate3 a day ago







Notes from Nature

Next week, Zooniverse is holding their annual meeting of project science teams. Since Al and Craig are both still in Tanzania, I'm going to be the only Snapshot Sarengeti representative there, but I'm super expited to go. I went to this meeting last year, while we were still developing Snapshot Serengeti, and it was both really fun (the Zooniverse team in Chicago are aweeome) and really useful. Since Zooniverse already had a dozen other projects live, I got a lot of advice from their science team members about what to expect when the Snapshot Serengeti went live, and also tos on analyzing the large data set that results from the project. This year, I'll be one of the people giving tips to the scientists of developing projects.

Speaking of which, if you haven't already, you should go check out Notes from Nature, which launched this week. It's a bit different from Snapshot

Subscribe to RSS



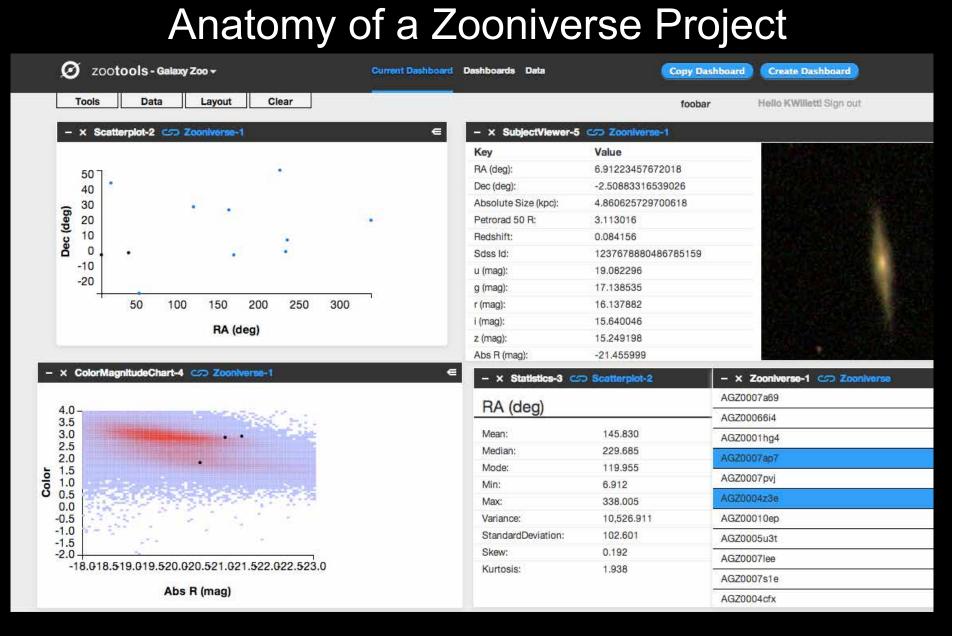
WELCOME

This is the blog of Snapshot Generget from the Zooniverse. Help chers by going on a virtual salari in Tanzania

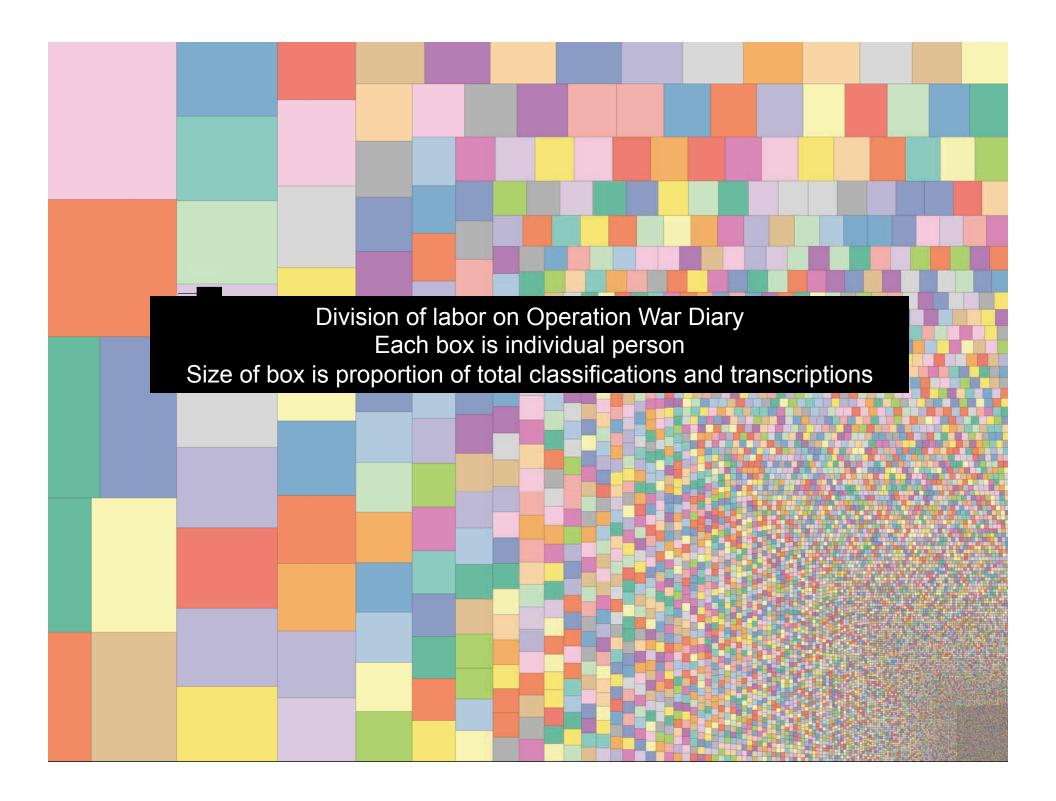
RECENT POSTS

Notes from Nature





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



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

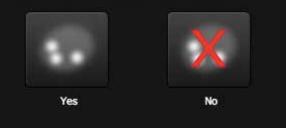


"There are known knowns. These are things we know that we know. There are known unknowns. That is to say, there are things that we know we don't know. But there are also unknown unknowns. There are things we don't know we don't know."

Donald Rumsfeld

Classes of Knowledge Discovery:

Known knowns : Primary task. Data reduction by science team. Does the galaxy have a mostly clumpy appearance?



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:

- Each project has own "talk" URL Develop "object oriented" discussion tool linked dir ightarrowfrom classification task "talk.galaxyzoo.org"
 - More engaging: integrates social media aspects including tagging, creating and sharing collections, discussing observations with peers and team researchers
- Develop a Citizen Science Journal allowing users to publish their work "letters.zooniverse.org"
- Develop more advanced volunteer-based research \bullet enabling tools "tools.zooniverse.org"



чÝ

Talking in the Zooniverse

VISIT MEASURING THE ANZACS CLASSIFY TALK	in the second
Search q	
Notes General comment threads about individual subjects Evan Roberts, Measuring the ANZACs Field Guide <u>an hour ago</u>	10 Participants25 Discussions61 Comments
Troubleshooting Having technical issues with Measuring the ANZACs in particular? Let us know about them here! Bonnie123 Boxes disappearing? <u>16 hours ago</u>	 24 Participants 15 Discussions 83 Comments
The Recruiting Station Just arrived? Say hello here. Evan Roberts, Measuring the ANZACs Welcome to Measuring the ANZACs <u>3 days ago</u>	▲ 10 Participants回 3 Discussions● 21 Comments
Measuring the ANZACs Discussion Discussion about the Measuring the ANZACs project.	 4 Participants 1 Discussion 4 Comments

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

Talking in the Zooniverse

Measuring the ANZACs Talk

Q

Measuring the ANZACs Talk > Notes

Search...

Notes

+New Discussion

FILE.

eld Guide	💄 1 Participant
Evan Roberts, Measuring the ANZACs an hour ago A lot of folks have asked about the different types of documents. We have a s	1 Comment
Subject 1016353	💄 2 Participants
Evan Roberts, Measuring the ANZACs <u>18 hours ago</u> hi @Zanna640. The keyword here is Ballot as you've recognized, so that woul	d 🗩 2 Comments
Subject 994216	▲ 4 Participants● 7 Comments

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



MARK TRANSCRIBE

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.

or

START MARKING

START TRANSCRIBING

Q 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
N	EW ZEALAND EXPEDITIONARY FORCE.
	"H" at
No.	13100 Name: Anth Redent of Regiment or Unit: A Conf. 11 th
	Questions to be put to the recruit before enlistment.
1. 1	What is your same? 1 I
2.	Whara were you born?
B. /	Are you a British subject? 3
	What is the date of your birth ?
	What is your trade or other the second
6. 7	Are you an indextured apprentice? If so, enough 6
7. 1	What was the address at which you last resided? T.
	Have yos passed the Fourth Bducational Standard 8.
9.1	What is the name and address of your present or 9isst suppoyer?
	Are you marrie?? 10
11. 3	Have you ever been sentences a sentences by 11
12.]	Do you now belong to any milliary or naval force? 13. If so, to what corps?
18. 1	Have you ever served in any military or naval force? 13.
14-3	Have you truly statud the whole (if any) of your 14.
10. 3	Have yon been registered for compulsory military 15.
16. 7	Have you ever been rejected as unlit for the military 16.
n, i	strong to be vaccinated or revaccinated and 17.
18, 2	tre you willing to serve in the Expeditionary Force 15
	about a boby be required : For the term of the present European war and for much further period as is notostary to bring the Expeditionary Perces back to New Zoaland and to dibland 17
N.0.0	own Your discharge will not be granted infers your return in New Zeahard unless permission for discharge coewhere be obtained from the
	I, do solomnly deflare that the above answurs made by use to have quastions are true, and that I am willing to fulfil the ergagement made.
	Signature of Reserved. Market and the second of Reserved and
	Signature of Witness
-	Oath to be taken by recruit on attestation.
114	t, Robert John Brath, , do mounty premise and every that I will be faithful and
fie 3	true allegiance to our forwarding Lord the King, his Heirs and Sonosconvened that I will faithfully serve in the New and Milinary Ferress, animolog to my liability moder the Defence Act, and that I will observe and oney all orients of Stajeary, his Heirs and Successors, and of the Generals and Officers set over me, until whall be lawfully discharged, do not, God,
110	Certificate of Magistrate or Attenting Officer.
uen	Construction of an adjustments or Attending Ogner. The above questions were read to the above-shared recruit in my presence. I have taken care that he understands each ion, and that his number to each question has been duly unitered as registed to, and the said recruit has male and signed
1000	eduration and taken the oath buffere me, at
	100 102 + 01 , 191 5 Signature of Alexing Officer Ast. Martchell T
	If nor allowskies to required on this page of the Atsentation, the Attenting Officer should be required to make it and halfed the abgration.

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?

11-3-21 V or delimitand to addiers the - SHEET to Headquarters HISTOR [B.R.-7. Reinforcement : 12 their. Stank Somama Christian Name 100 Beary Smith Robert 13130 John Am 1ºBa dry B B. Religion Last New Zealand address mtractor Occupation e Wharau Cof to Last employer: A Rummer Machinton Name, relationship. Mrs Violet Amith and address of next-of-kin (if not 35 Marine Pide 5 Railway Lerrace Te Wharan resident in New Zealand, insert also Yartinton name and address Stafford Service (Where did he of nearest relative in New Zealand) : Free Rile Eng Service. No. 13130 Renk: Plan go? What did he do? Days. Name. Address .. Service : Service Periods of Service. Theatres of Operation Service. years /22 days. Australasian In New Zealand ; Egyptian years / 5 days. Overseas Wounds Egyptian B.F. yeara 275 days. e-slight Total service Balkan Date commenced duty European 10% Asiatic Wounds Decorations BRITICH WAR MEDAL Sickness VICTURY MEDAL The foregoing particulars are corres 120 Simultur 9.17 or omit to advise this office of any future change of address. tone (ada/8/m-1 mil Sick Egyptian E.F. Total service ... 2 years 75 days. Balkan Death Date commenced duty 14 1-10. Western) 1910-17 European Date finally discharged : 5 . 10 . 15 Asialo Killed in action Died of wound Prisoner sickne orations BRIT H WAR MEDAL Missing MEDAL 12 Prisoner The foregoing particulars are correct Discharge Injuries in or b Signature 93 Ven Good Service N.B. Do not emit to advise this office 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

			The second secon			1		1					
Htten HBA	in the second	STATEMENT of the SERV	ICES of In	ema (Surniame.)	in 1	Will (Christian)	Lam?	Hela	deleto. 8/ 1+20	34			
D Coy. 11th Rei		Regiment or Corps.	Promotions,	Reductions, Cas	malties, &c.	Rank	Date.	Aut	thority for Entry.		Home	Service.	
Occupatio Last empl Name, re and ad	DI	Int les log	hook gran	of 3	0	nle "	21/12/15 11/3/16	Ryo	380		Years.	Days.	
next-of-ki resident Zealand, name an	S	er Blui. O lago Ry	- y	eur Pa	alet NEC	3	14/3/16	P	or us 23. Rouen 1	27/10	Foreign	Service :	
of neares		h. Z. Command deport Codport 1° Bu. Clago Rog T. 4"	her in stringth appro-	from 1. J. Con Hop	Hidde 1 porter V.D. 4	Pte	10.4.17	Paore	C.D. Cod. 10.4.17 a	48.	- 1	146	-
DISCH		1º 130. 000 0 100 1.900	prime. for h	. J. pes 17	· D. Max amo	<u><u>y</u><u>u</u>.</u>	14.4.17.	1). 0.3	So. Money Mana 2	.8. 17	Total S		
Service				-					-		_ /	338	
	C/ No.		in the second						1 744				
Wounds													
	Q.7.	Regiment or Corps.	Place.	Date.	DUCT-SHEE	<u>r.</u>	Punishme		Authority for Pater				
	PET						- Cumbuno		Authority for Entry.				
	124 124 124 124 126	the second second		-									
Sick	C 24 00		total for speciality		teniari mi	02 -10 777	- and	11					
	D.H. 3 D.H. 3												
u ^{ll} Const Balt (0.3. Killed in action	4.												
Died of wounds	8		2					Pana					
Missing Prisoner	" "				1 4 3		in entre	194					
Injuries in or by Service	the Dr				-								
Discharge	Provisio				archene a		Negis S	1					
	Final : a		Lix A 40	P. 1985 54	States and		and and	a strange					
Pension			1	*	4								
		* 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

GitHub, Inc. (US) https://github.com/zooniverse/scribeAPI

Scribe

Scribe is a framework for crowdsourcing the transcription of text-based documents, particularly documents that are not well suited for Optical Character Recognition. It is a collaboration between Zooniverse and The New York Public Library with generous support from The National Endowment for the Humanities (NEH).

CO

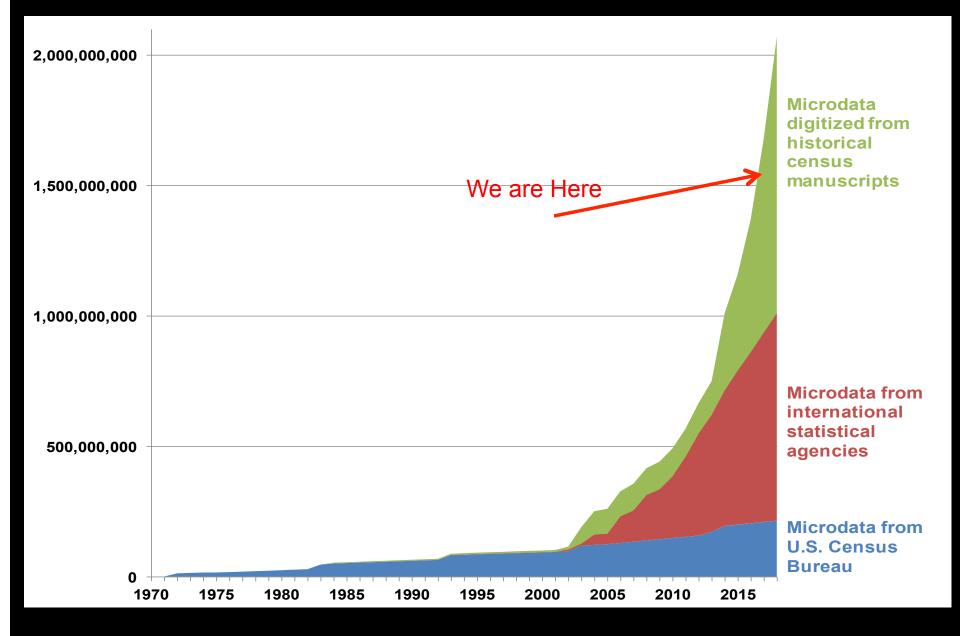
- C

For Project Creators

Are you an organization or individual interested in using Scribe for your next crowdsourced transcription project? Start here!

- What is Scribe and is it for me? Read our Scribe Primer
- Ready to set up your project? Head over to our Project Setup page

Growth of public-use census microdata



State	new york	Incorporated place	Brony	Burn Ward of	ył. city <u>+</u> u	nincorporated place		These segant is required will be used a	by Art of Chaprens. This solution for property desired	and so that the set of the			MMERCE-					B. D. No. 2	E.D.1	No. 3-	12	She
County _	Brong	Township or other division of county	1.2.1_		a A'A'C'I Fu	otitution	itotifation a	nt place having 20 or more b at Date on which satisfies are	Statute:				TION		EDUI	LE	1414	and the second second	ant n			Enu
LOC	LATION HOUSEHOLD DATA	NAME	RELATION	PERSON DESCRIPT	AL EDUCATION	PLACE OF BIRT			RESIDENCE, APRIL	124 67 700					PER	SONS 14	TARS OLD AND OVER		TE.			
1	1 1 1 1 4 4	te of each person whose usual p readence on April 1, 1960, was is household. URE TO INCLUDS: Prome temporally about from house of the date income format.	taces Anisotrative of this years of the provident of the	<u>,</u>	C. Direct (n) C. Direct (n) C. Direct (n) C. Direct (n) Direct (n) Direct (n)	If here is the United States give Bate, Turtbery, or presention. If favoign here, give neces- try is which birthplote was attained as Jackney 1, 1357.	and Arrite Mes	IN WEAT PLA for a person who an April Cull of Same bears form, ease, "Rates place for a person who lived in directed in the Instruction mult address."	CE DED THES PERSON 1 1.1, 1953, was living in the rate for one trying in a def w," insering Coln. 14, 13, or 1 a different phase, enter one. (Enter artual phase of	served brouse bot in the	Ballie villy of	A WORL'S PARA			Hat prints of states patry Gar- cranes outh Cal Hi	Patain Speta Patain Patain Patain California California	OCCUPATION	ON, INDUSTRY, AND CLASS and in public concretences on and correspondences. Solution, as four constraints, Solution, and They is Out HU, (a) If he has not chase of works or (b) if har" in Out HI, and incre Coin. H INDUSTRY	or voters			
	a lana analar a' lan a baabar a' lan a baabar a' lan a lan anal a baabar a' a lan a baabar a' a lan a baabar a'	ne "Ab" eher annen of auch permen. . Childen under ymr of age. Wein "In all has au bene jenn afer anne. bare Suber same permafernisbing infer		Color or Table	Radio	Distinguish Ganda- Breach Bum Ganada- Baghan and Dum Free- Rate (Brey from Both- ers Ireland.	CODE Jam 15	Cip, inwa, or village having \$1.00 or male inhabitants. Enter "2" for all other places.	COURTE STA	19 20		Book a post book a post book and to book out book		111	And	Dention of the Dention of the Dentio	Occupations of particu- ber bind of work, w- frame spinner laborer prost heater music feascher 28	Industry or business ar- corress cull press grocery form shapperd public school		Burberd week	and an interview of the	a new sector and a
		RANFELDT CAR	HEAD 0	MWS		NEW YORK	50.	Same House		M		Yes -		- 1	40	-	TINE KEEPEN	CUT STONE YAR	PW	5	2 3280	_
2 5		- ARV		MWY	8 S No 8 5	NEW YORK.	56	SAME HOUSE				YES -		- 1	40	28	HELPER	CUT STONEYAR	Pur IL	24 4	6 1,165	No
3		S100	-		3 101 No 8 8	SWEDEN	b	SAME HOUSE		N	6	100000000000000000000000000000000000000	to No No	H	-						0 001	No
13		HERE			AND 6 - AI		1		+ No DWEL	Ling:013	Lest.	18					SUPERINTENDAN	BAILLIP Jon			1./	
- 3	825 2 R 40 No 1	ALCO - ANTHON JENNI			8 M No HU	ITALY	75 M	9 SAME HOUSE				Yes -	No No	H	48		LoFT BUILDING	O AMT GAPANY	PW	- 0	2 2400	No
7		- NicHold			2 S NO HU	NEW YORK	-50	SAME House				Yes -		-	40		STOCKCURK	SHIRT MENE COMPANY.	Pw	16115	2 1040	No
8 4			SADDANGATER		9 5 No H4	NEW YORK	jt.	Some HOUSE					o Yes -	-		20	OPERATOR	SHIPT MENG	Pw		5 600	
-9		HERE		BLOCK		1		EGINS B	LOCKE	-> No	WE	14.10	Gen	Bere	72	AND	E"				+	
10	143 3 0 1350 NOF		-		4 M NO 2	ITALY	1	A SAME HOASE					a No Ma							_ 4	201	
n	7/2 / 0 26 /	EUXABETH			8 M NOR	ITALY		A AME HOUSE					ON DIN O	H - 1	44	-	TRUSK	WHOLESAL	0			No
12	743 4 R 28 No 1	MOR. Addastus.	HERD	EWS	M NO 8 M NO 8	NEW YORK	1.00	Some House					o Yes -	H	77	39		OFFICE WOR			6 1302	
13		Augustus	2 7 3 2 2	MWZ		NEW YOR		JARCE (PYNE			-	nu n	U /62	1		1		CITIC MON		-	-12-21	~
15	143 5 R 28 NO 8				O M NO 8	ITALY	8. M	A SAMEHOUSE				Non	10 Yes -	-	3	13	BUTCHER	RESTAURANT	Paul	7/ 15	2 1456	NO
16		MARY	Witz		1 M No 8	ITALY	REN	A SAME HOUSE				We A	to no No	H			614 630				001	
17		WILLIA		MWI	7 S Vesci	NEW YORK	- 12	SAME HOUSE			-	No A	to NO NO	SI	-						00	
18	2/1 1 8 26 1 6	FRED			16 S YESH3	NEW YORM		DAMEHOUSE	-		-		to No No	S	110	20	-	DEPT. N.Y.C			901	
19	743 6 R 28 No R 743 7 R 28 NO R	IGUIRI RALPH	HEAD		ES No 8	NEW YORK		SAME House	<u> </u>	1. 1.	10124	YES -		-			CHAUFFEIR	NYCITY FIL	SW	- 3	6 1000	
20	143 / n 21 mgm	CELEST	WIFE		g M No P	NEW YORK	100	WATERERD SAME Now	DULLIVAN I	EWARK	563	YES No -		-	48		RECEMBER	N.Y.CITY FIL DEATENME BEAUTY SCHOOL	6W	- 0	7 700	
22	74388 28 No 1	BERRY MATHE			6 M NO 8	NEWYORM		SAMERIKE	BRONK	VEWYORK	Xaxo	VES		- 1	40	1		NEW YORK CAN		16	2 1860	N
23		MARGARE			BOM NO P				BRNX A		XoXO	1.	to No No	H	1					1		Ne
24 m		MARSARE			85 No 8	NEW YORK		SAMERANE	BRONS	NEWYORK	Xoxo	Ven		-			ASSEMBLE	MENG-GO	m		3.008	
25	739 9 0 6000 No 7				58 M NO P	YUGodian			BRONK 1		XOXO	11-0		-	44		BAKER	REAFERA	Par	3115	21050	
26		GAZELLA	1 1 2		16 M NO 8	GERMAN			BRING /		3030	1100	10 NO NO		1/2	2.4	Deriver	UNDERWEAL			00	
27		EMORY DAZIA	JA JON SON.		18 S NO HZ	NEWYORK NEW YORK		Some Reca	BRANK A	VEWYORK	1029	and the state of t	- Yes -	2	65	3.4	FILLER	UNDERWEN Y FACTORY SODA WATER FACTORY	8	16 1	6 240	
29		VILMA			6 5 No 8	NEWYORK		SAMEHONE		L M/ MAN			TO NO NO	U							0 0	N
30	4 1 5	EIFERT YOLAND		FW	22 M No 8	NEW YORK	5	SAMEHOUSE							140		ASSEMBU	REALTORY	Pm 496	1/3/1	00	N
31		WILLIA	ME SON IN	MW	31 M NO 8	NEW YORK	1	SAMERACE		NEWYORK	352	No -	- Yes -	-	3 13	13	ELEVATOROF	Bunoint	Pw 30		210/0	N
32	Hard a sa ut		LLA DRUGHTER			NEW YOR		Same How				Ve		-	11	-	0.110	ELECTRIC G	0		00	
33	73710 R 32 No M	FRANC	S& HEAD IS WIFE		30 M No H3	COTLAND	01	ASAMERAKE BLOOMFIELD	ORO NA I	NEW YOR5	5.74	Ver-			40	1.1.1.1	OILER	ELECTRIC	. rw+/+		2 2000	N
33 34 35 36		JANE	0		22 SWOH3	SCOTLAND		A NEW YORK	Berra	Now Veer	10	YES	to No No		48	2	SALESGIR	L DEPT STOR	Ru		0 0	
36		MATHE			19 5 Notte			4 SAME PLAC			10		- Yes -	-	3	64	NEW WARKE	A 1			00	1.00
37		ISABE			57 19 No 8	8 ITALY		In Same House					to No NO		5							1
38	735 11 0 4,000 No 1	ANCIONS CARLO	S HEAD		65 M N. 8	and the second second second		VA SAME HOUSE				- 1 A M A M A M A M A M A M A M A M A M A	to to M		/					-+	00	No
39		MARY	WIFE		SZ M NO D	ITALY		AL SAMEHOUSE					NO NO N		17			EXCAVATIN	e		0 -	M
40	SUPPLEMENTARY	JOSEPH	DI SON		29 M No 10	NEW YOR	<u>r </u>	SAME House	1			YES	. - -		L WOMEN W	10 ANE 1		CONCERN		1CUL	12 520	M
	OUESTIONS	Flace of states of	FOR PE	BOONS OF ALL	AGES TONGES (OR BATTYE LANDFAGE)	VETERARS	-	SOCIAL PROVADT		14 YEARS OLD	-		OBTER	08 84	NE BEEN MA	C_	FOL	FFICE USE ONLY-DO	SOT WEITE 1	THESE CO	ALCHENS	T
Tor P	tersons Enumerated on Lines 14 and 23	Factor of Sailed Sailes		ange L. 1987 Office J Bran	LANDTAGE	Is this person a veter United States militar of the wife, wider of it proved children	torotan	All Peris, brough the filter and all the broads from the present from the	Enter that occupation whi able is work. Who per longest during the past by and usual date of For a person without per black.	ich the permu reports mus is upshirts detern 10 years and at which	as his usual o title this, enter he is physical	computies and that occupation ty able to work.	at which he is physic at which he has we Enter size neual int	111 1 1	1	ables out being	aterie Stille	****				
4		FATURE	NOTELL.	CODE	and an and an and an and an and an and an	an Canada and Can	C. C	And the second s	THUAL OCCUPATI		L INDUSTRY	. Tran	Cons lines	C. P.	-	a new of	5 10		ň			
1	35		87	G	35	H 50 40 41		42 43 44	45		46	47	1	3ª 	40	10 K	LXXO	PQBS				+
14	IMOR. AUGUSTUS	NEW YARK	NEW YORK		NELISH	NO NO N		Yo		_				2		1	41 2	7	11	T	1	
1"	TENKE YILMA		GERMANY	- to the first of	IL BERERIT	114 114 14	×1 1/	I.M.									stantas lands from the stantas		in the second se			

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.