Crowdsourcing: How and why should libraries do it?

Author: Rose Holley, Manager – Trove, National Library of Australia rholley@nla.gov.au
Published: DLib Magazine – March/April edition 2010
Final version: 11 March 2010

Abstract
The definition and purpose of crowdsourcing and its relevance to libraries is discussed with particular reference to the Australian Newspapers service http://newspapers.nla.gov.au, FamilySearch http://familysearchindexing.org, Wikipedia http://wikipedia.org, Distributed Proofreaders http://www.pgdp.net, Galaxy Zoo http://www.galaxyzoo.org and The Guardian MP's Expenses Scandal http://mps-expenses.guardian.co.uk. These services have harnessed thousands of digital volunteers who transcribe, create, enhance and correct text, images and archives. Known facts about crowdsourcing are presented and helpful tips and strategies for libraries beginning to crowdsource are given.

Keywords
Crowdsourcing, social engagement, web 2.0, text correction, digital libraries, digital volunteers, virtual volunteering, wisdom of crowds, citizen science.

1. Background

This article is an abridged version of my private research paper[1]. I gave a keynote presentation[2] on crowdsourcing to the Pacific Rim Digital Library Alliance (PRDLA) annual meeting in Auckland, New Zealand on 18 November 2009. The research was undertaken in my own personal time and contains my own personal views on the potential of crowdsourcing for libraries. These views are not necessarily shared or endorsed by the National Library of Australia (my employer).

2. What is crowdsourcing?

Crowdsourcing is a new term in our vocabulary. It does not yet have a clearly agreed definition, is not recognised by my spellchecker, and does not have widespread usage in the library world. Wikipedia provides a useful starting point to understand the term. An extract of their page[3] from August 2009 reads as follows:

"Crowdsourcing is a neologism for the act of taking tasks traditionally performed by an employee or contractor and outsourcing it to a group (crowd) of people or community in the form of an open call. For example, the public may be invited to develop a new technology, carry out a design task, refine or carry out the steps of an algorithm, or help capture, systematize or analyze large amounts of data (citizen science). The term has become popular with business authors and journalists as shorthand for the trend of leveraging the mass collaboration enabled by Web 2.0 technologies to achieve business goals. The difference between crowdsourcing and ordinary outsourcing is that a task or problem is outsourced to an undefined public rather than a specific other body. In crowdsourcing the activity is initiated by a client and the work may be undertaken on an individual, as well as a group, basis. Crowdsourcing has potential to be a problem-solving mechanism for government and non-profit use."

I want to expand this definition and clarify the difference between social engagement and crowdsourcing. Social engagement is about giving the public the ability to communicate with us and each other; to add value to existing library data by tagging, commenting, rating, reviewing, text correcting; and to create and upload content to add to our collections. This type of engagement is

usually undertaken by individuals for themselves and their own purposes. (In the previous issue of D-Lib I have written about user tagging[4]). Crowdsourcing uses social engagement techniques to help a group of people achieve a shared, usually significant, and large goal by working collaboratively together as a group. Crowdsourcing also usually entails a greater level of effort, time and intellectual input from an individual than just socially engaging. For example correcting the text of a newspaper article, or transcribing a complete shipping record involves more input than quickly adding a tag to a photograph, or rating a book on a scale of 1-5. Crowdsourcing relies on sustained input from a group of people working towards a common goal, whereas social engagement may be transitory, sporadic or done just once. Clay Shirky's book 'Here Comes Everybody'[5] contains some interesting examples of both types of activity that have been enabled by web 2.0 technologies.

Libraries are already proficient in the first step in crowdsourcing: social engagement with individuals, but we need to get proficient in the second step: defining and working towards group goals. Social engagement has happened for years in libraries. In the 'pre- digital library days' a user did not expect to go to a library and have a simple information transaction. They wanted the information but they also wanted to discuss with the librarian (or any other user) what they thought of the latest novel they had just read, the results of their research, what else they know about steam locomotives that was not in the book they just read, or the error they just found in your card catalogue. On the books return desk they tried to sneak back text books that had pencil or worse pen or highlighter underlinings and annotations, without the librarian noticing. If they were noticed they would be fined, or worse banned from borrowing books again! They wrote formal letters of complaint or compliment about library services to the librarian who replied on headed paper. When libraries first started delivering digital resources all these social interactions were taken away from users and they simply got an information transaction by downloading content. It has taken libraries a while to realise that users still want more than a simple information transaction and they want the same and more social interactions than they had in the 'pre-digital' days. In our digital library world they want to: review books, share information, add value to our data by adding their own content, add comments and annotations and 'digital post-its' to e-books, correct our data errors, and converse with other users. And now they are telling us they can do even more, they can organise themselves to work together to achieve big goals for libraries and make our information even more accessible, accurate and interesting. Why are we not snapping up this great offer immediately? How and why should we do it?

3. Why should libraries do it?

Crowdsourcing could bring great benefits to libraries:

- Achieving goals the library would never have the time, financial or staff resource to achieve on its own.
- Achieving goals in a much faster timeframe than the library may be able to achieve if it worked on its own.
- Building new virtual communities and user groups.
- Actively involving and engaging the community with the library and its other users and collections.
- Utilising the knowledge, expertise and interest of the community.
- Improving the quality of data/resource (e.g. by text, or catalogue corrections), resulting in more accurate searching.
- Adding value to data (e.g. by addition of comments, tags, ratings, reviews).
- Making data discoverable in different ways for a more diverse audience (e.g. by tagging).
- Gaining first-hand insight on user desires and the answers to difficult questions by asking and then listening to the crowd.
- Demonstrating the value and relevance of the library in the community by the high level of public involvement.

- Strengthening and building trust and loyalty of the users to the library. Users do not feel taken advantage of because libraries are non-profit making.
- Encouraging a sense of public ownership and responsibility towards cultural heritage collections, through user's contributions and collaborations.

Examples of crowdsourcing goals for libraries could be: getting users to mark the errors in our catalogues; rating the reliability of information/records; adding information to records; verifying name authority files; adding user created content to collections; creating e-books; correcting full text; transcribing handwritten records; and most especially describing items that we have not made accessible because they are not catalogued/described yet. A prime example of this is photographs. The normal procedure in a library is that a photograph is not digitised until it has been catalogued. If instead it is digitised first and users are given the chance to describe the content this would radically open up access to a lot of 'hidden' and difficult to describe photographic collections.

There are very few described examples of successful crowdsourcing projects in libraries because it is not really happening yet. In March 2009 I published my own research into the text correction activity by digital volunteers on the Australian Newspapers Digitisation Program ('Many Hands Make Light Work')[6]. This report generated a huge amount of positive interest in the international library community and resulted in me making contact with several other crowdsourcing projects and undertaking more research in this area. Many libraries are unwilling to take the lead/risk of being the first library to undertake something new on a significant scale, but hopefully the information in this article will lead libraries to move into this area of activity with knowledge of how to do it.

4. Crowdsourcing achievements

Australian Newspapers Digitisation Program, National Library of Australia: Text Correction.
http://www.nla.gov.au/ndp/

The service was released to users in August 2008 and at November 2009 containing 8.4 million articles to search and correct. The public are enhancing the data by correcting text and adding tags and comments. The manager says "This is an innovative library project which is exposing data to the public for improvement. The public are very keen to help record 200 years of Australian history by text correction of historic Australian Newspapers".

Fig. 1. Australian Newspapers – number of volunteers and achievements. Sourced from Rose Holley.

Date	Number of volunteers	Cumulative achievements: Lines of text corrected in newspaper articles
August 2008	Australian Newspapers released	
August 2009	5,000+	4.7 million lines in 216,000 articles
November 2009	*6,000+	7 million lines in 318,000 articles

*Unknown exactly how many volunteers there are since users do not have to register to correct text and users that are registered do not all correct text. This figure is a minimum estimate.

Picture Australia, National Library of Australia: Creation and addition of images.
http://www.pictureaustralia.org/

Picture Australia harvests digital images from over 50 cultural heritage institutions both in Australia and overseas. Since January 2006 the public have been encouraged to add their own digital images to Picture Australia via a relationship set up with Flickr. Any of the contributing institutions can then request permission from the public contributors to upload the contributed pictures into their own collections, while the same pictures also remain visible in the central Picture Australia pool. Fiona

Hooton, Manager of Picture Australia said "As a partner of PictureAustralia, *flickr* is now an open door for individuals to help build the nation's visual record".[7]

Fig. 2. Picture Australia – number of volunteers and achievements. Sourced from Fiona Hooton.

Date	Number of volunteers	Cumulative achievements: Number of public images added to Picture Australia
January 2006	Contribution of images by the public enabled.	
January 2007	unknown	14,000 approx
January 2008	unknown	27,227
January 2009	unknown	42,774
October 2009	2,641	55,664

FamilySearchIndexing, Latter Day Saints: Text transcription of records.
http://www.familysearch.org/eng/indexing/frameset_indexing.asp
FamilySearchIndexing made available millions of handwritten digital images of births, deaths, marriages and census records for transcription by the public in August 2005. Once a record has been transcribed it is included in the free online FamilySearch service, one of the most heavily used genealogy resources in the world. The home page of the site says "The key life events of billions of people are being preserved and shared through the efforts of people like you".

Fig. 3. FamilySearchIndexing – number of volunteers and achievements. Sourced from Mark Kelly.

Date	Number of volunteers	Cumulative achievements: Number of individual names transcribed
August 2005	FamilySearch Indexing on web introduced.	
January 2006	2,004 online volunteers	
January 2007	23,000 online volunteers	102 million
January 2008		217 million
November 2009	160,000 online volunteers	334 million

Distributed Proofreaders (DP), Project Gutenberg: Creation of Ebooks.
http://www.pgdp.net/c
DP, a non-profit organisation, makes public domain books and journal issues into freely available eBooks. DP was formed in 2000. Volunteers find, scan, OCR and mark up the books themselves, page by page. DP is one of the earliest most forward thinking crowdsourcing project. Juliet Sutherland, Manager of DP says "DP has a very strong sense of community and a shared goal to convert public domain books into e-books that are freely available. We grew up from grassroots and operate on a shoe-string budget of $2000 per year. It is a great achievement therefore to have 90,000 volunteers who have created and made publicly available 16,000 texts over the last 9 years. There is still a huge amount of material out there to be transcribed!"

Fig. 4. Distributed Proofreaders – number of volunteers and achievements. Sourced from Juliet Sutherland.

Date	Number of volunteers	Cumulative achievements: E-texts (books and journal issues) created
October 2000	DP founded	
January 2002	232	~ 60
2003	18,469	2,646
2004	29,029	5,912
2005	35,912	8,066
2006	51,098	9,842
2007	64,368	12,021
2008	78,015	14,440
End October 2009 (9 yrs since established)	89,979*	16,000 **

*cumulative registered users. Of these 3000 are active per month.
** approximately 2000 E-texts consistently produced per year.

Wikipedia: Creation of digital encyclopaedia; improving personal names authority file at the German National Library; Wikisource.
http://en.wikipedia.org
There are several projects at the Wikimedia Foundation. The most well know is the free online encyclopaedia created by the public, *Wikipedia*. Andrew Lih says in the conclusion of his book 'The Wikipedia Revolution'[8] , "Wikipedia led the way in demonstrating that the collaborative accumulation of knowledge was not only feasible but desirable. Its neutrality policy, combined with a global team of volunteers helped make Wikipedia not just a clone of existing encyclopedias , but an encyclopedia that made recording human history a revolutionary, collaborative act."

Fig. 5. Wikipedia – number of volunteers and achievements. Sourced from http://en.wikipedia.org/wiki/Wikipedia:Statistics[9]

Date	Number of volunteers	Cumulative achievements: Articles created in English
January 2001	Wikipedia created	
Dec 2001		16,442
Dec 2002		100,000
Dec 2003		189,000
Dec 2004		445,000
Dec 2005		922,000
Dec 2006		1.5 million
Dec 2007		2.1 million
December 2008 (8 yrs since established)	156,000 active monthly volunteers out of 10 million total*	3 million (total of 10 million articles created in 250 languages)

* The number of registered volunteers has grown to millions. Most volunteers are not regular contributors, however. An unknown, but relatively large, number of unregistered contributors also contribute.

In 2005 volunteers at Wikimedia Deutschland worked with the Deutsche Bibliothek (DDB) (German National Library) on the Personal Names Data Authority File project. The Personennamendatei (PND)

has library datasets for more than 600,000 persons and contains more than 2 million names. Deutschland Wikipedia also has a person's name file and 20% of the articles in the encyclopedia are for individual people. A project to enrich articles in the German Wikipedia with PND numbers and links back into the DDB's catalogue was arranged. There were benefits for both organisations since Wikipedia gained persistent and controlled links to literature from and about specific people and the DDB got thousands of corrections and updates to the PND by volunteers who checked and matched each Wikipedia authority file against the PND authority file. After 2 weeks 20,000 Wikipedia articles had links added to the PND. A Wikimedia article[10] and a conference paper[11] outline the project more fully. Another Wikimedia project, Wikisource, is transcribing public domain books into Ebooks.

UK MPs Expenses Scandal, The Guardian: Tagging subject content of records/archives.
http://mps-expenses.guardian.co.uk/
In June 2009 there was controversy over UK MP expenses. This created huge public interest and outrage and the result was that 2 million documents relating to MP expenses were made publicly available as PDFs. The Guardian newspaper built a website in a matter of days that enabled the public to view, read and mark documents they felt should be exposed and investigated. An article[12] was published by the Guardian outlining the lessons learnt about crowdsourcing. This project has significant relevance for archives, manuscripts and records.

Fig. 6. The Guardian MP Expenses Scandal – number of volunteers and achievements. Sourced from Simon Willison and Guardian website.

Date	Number of volunteers	Cumulative achievements: Number of pages reviewed
June 2009	Website released	
First 80 hours	20,000	170,000
November 2009	24,524	215,305

Galaxy Zoo, International University Collaboration: Classification of digital photos
http://www.galaxyzoo.org/
Galaxy Zoo is an online collaborative astronomy project involving several international Universities. Members of the public are invited to assist in classifying millions of galaxies from digital photos. Chris Lintott a member of the Galaxy Zoo team says, "One advantage [of helping] is that you get to see parts of space that have never been seen before. These images were taken by a robotic telescope and processed automatically, so the odds are that when you log on, that first galaxy you see will be one that no human has seen before". [13] The aim is to have "each and every galaxy classified by 30 separate users. The importance of multiple classifications is that it will enable us to build an accurate and reliable database, that will meet the high standards of the scientific community".

Fig. 7. Galaxy Zoo – number of volunteers and achievements. Sourced from Galaxy Zoo website.

Date	Number of volunteers	Cumulative achievements: Classifications of images of galaxies
July 2007	Galaxy Zoo released	
August 2007	80,000	10 million
July 2008	150,000	50 million

World War 2 Peoples War, BBC: Creation of content
http://www.bbc.co.uk/ww2peopleswar
Between June 2003 and Jan 2006 the BBC asked the public to contribute their memories and artefacts of the Second World War to a website. This was answered by 32,000 people who registered and submitted 47,000 stories and 15,000 images. It demonstrated that the public are eager to supplement content on

various topics with both their knowledge and their own digitised content. The topic chosen was one in which many people had first hand experience, knowledge and content. The site was revolutionary at the time because of the high level of social engagement and involvement it gave to the public, and the fact that many of the volunteers were elderly and had not used PC's or the Internet before. Archived screenshots[14] in the project history show how the site worked.

5. Known facts about crowdsourcing

5.1 Volunteer numbers and achievements

Wikipedia, Distributed Proofreaders, Australian Newspapers and FamilySearchIndexing all released their services 'quietly' with little or no advertising, but clear group goals. All had fewer than 4000 volunteers in their first year. These figures rose dramatically in subsequent years as the communities passed the word on and viral marketing (i.e. hearing of the project by blogs, forums, e-mail) took place. Currently Distributed Proofreaders, the Wikimedia Foundation and FamilySearchIndexing have been established for more than 4 years and all have about 100,000 active volunteers. An active volunteer is defined as someone who works at least once a month on a regular basis. They all have far more registered volunteers, with FamilySearchIndexing reporting 160,000 volunteers in January 2009.

5.2 Volunteer Profile

Distributed Proofreaders, FamilySearchIndexing, Wikimedia Foundation and Australian Newspapers have independently undertaken their own analysis of volunteer makeup and have discovered the same things:

- Although there may be a lot of volunteers the majority of the work (up to 80% in some cases) is done by 10% of the users.
- The top 10% or 'super' volunteers consistently achieve significantly larger amounts of work than everyone else.
- The 'super' volunteers have long session durations and usually remain working on the project for years. They are working on your project as if it was a full-time job.
- Age of volunteers varies widely. The 'super' volunteers are likely to be a mix of retired people and young dynamic high achieving professionals with full-time jobs.
- Public moderator roles and roles with extra responsibilities are likely to be taken by volunteers aged 30-40 who are in full-time employment.
- Disabled, sick, terminally ill, and recovering people are among the volunteers since working at home is convenient, gives purpose and structure to the day, and gives feelings of value and reward.
- Many people find the time to do voluntary activities because they do not watch much television and as Clay Shirky describes it use this 'cognitive surplus' time for social endeavours.
- Half of the active volunteers are doing it because they are very personally interested in the subject matter, and half are doing it because they want to do some voluntary activities and see it as a good cause.
- Having a minimum level of computer/keyboard/internet knowledge is not a pre-requisite to volunteering. Many volunteers have low levels of PC proficiency and build up their levels of IT literacy by volunteering for online work. Having never used the Internet or a computer before is not a block for many volunteers.
- Volunteers appreciate that they can learn new things as they go along and many of the projects could be termed 'educational' in some respect.
- Many volunteers (especially genealogists) help on several different online projects.
- Volunteers are much more likely to help non-profit making organisations than commercial companies, because they do not want to feel that their work can be commercially exploited. (This places libraries and archives in a very good position for crowdsourcing.)

- Volunteers continue to work because they find it personally rewarding, and they want to help achieve the main group goal.
- The amount of work volunteers achieve usually exceeds the expectations of the site managers.
- Some volunteers like to be able to choose subjects, and types of work they do, whilst others prefer to be directed to what to do next. Therefore most sites offered a 'pick your work' and a 'do the next thing that needs doing' option.
- Some volunteers like the idea of communicating with other volunteers, but some others just wanted to get on with the job. Generally volunteer moderators were keen communicators and 'super' volunteers were 'head down' types.
- On significant projects e.g. FamilySearchIndexing, Australian Newspapers, Galaxy Zoo, many volunteers describe the work as being 'addictive' or getting 'hooked' or 'sucked in' and time quickly gets away from them, hence they spend far longer than they actually intended to in voluntary work.

5.3 Motivational factors

The factors that motivate digital volunteers are really no different to factors that motivate anyone to do anything [15]:

- I love it (As Clay Shirky would say 'The Internet runs on love').
- It is interesting and fun and I learn new things.
- It is a worthy cause.
- I want to volunteer and give something back to the community.
- I can help in achieving the group goal.
- The goal/problem is so big it is a challenge.
- I am playing an important role in the field of science or history and helping to record, find or discover new things.
- You placed your trust in me and I want to prove I can do it even better than you thought.

"Why do I spend 6 hours a day on Galaxy Zoo? Simply because it gives people who are not lucky enough to be a part of the scientific community a chance to take part in something that furthers the understanding of not only Galaxy's but our future as well. I'm just loving my time here."

"Why do I spend 8 hours a day on Australian Newspapers? The Australian Newspapers digitisation program is the best thing that has ever happened to me in twenty years of family history research. It is a wonderful resource and so valuable for folk who can only do research on-line. Correcting electronically translated text is a worthwhile and enjoyable task and I am happily correcting text to help record Australian history."

By observation and surveys, site managers reported that volunteers were generally highly self motivated but there were a few things that noticeably increased their motivation:

- Adding more content more regularly to the site for them to work on.
- Raising the bar and increasing the challenge/end goal, e.g., identify all the galaxies in the universe; correct all the text in all newspapers; digitise every out of copyright book.
- Creating an online environment of camaraderie for the virtual community by use of forums so that the digital volunteers feel part of a team and can give each other support and help.
- Being very clear about what, how and when things should be done (instructions, FAQ, policies).
- Acknowledgement of the digital volunteers in various ways.
- Rewarding high achieving digital volunteers.
- Being able to see the progress of the big goal and their place in that (by transparent statistics).

5.4 Types of acknowledgement and reward offered

Many of the crowdsourcing sites were unable or had not thought to offer any acknowledgement or reward to volunteers. Being non-profit organisations, most were limited in the type of reward or acknowledgement they could offer anyway. In all cases digital volunteers were proving to do great work without reward systems and had volunteered on the basis that there would be no reward. However some of the sites were now thinking about this more. The following ideas, that cost little or nothing, but worked have been used:

- Acknowledging by naming volunteers (or their handles) on web pages, in newsletters, publicity, on the item they created/amended.
- Volunteers being able to choose public or private profiles so they can be visible if they want.
- Rewarding high achievers in ranking tables.
- Certificates of achievement.
- Promotional gifts e.g. t-shirts, books, vouchers.
- Travel to meet the paid staff on the project.

5.5 Management of volunteers

Most of the crowdsourcing sites had either no paid staff or very limited staffing to manage and co-ordinate hundreds and thousands of volunteers. Because of this they had done two things:

1. Utilised volunteers to moderate/co-ordinate other volunteers and to answer the questions of other volunteers.
2. Implemented IT savvy ways through open source software to manage crowds, communications and processes, e.g., forum and wiki software.

The main task of the paid staff in regard to management of volunteers was to create, establish or endorse guidelines, FAQs, and policies for the digital volunteer processes. The site manager may also keep an eye on the forum activity and spot anything which may become an issue and resolve it through FAQ, policy or guidelines.

All site managers agreed this was the way to handle large online communities. No attempt should be made to seek paid staff to 'manage' digital volunteers. Erik Moeller at Wikimedia Foundation endorses this viewpoint, and they have 10 million registered volunteers.

6. Tips for crowdsourcing

Fig. 8. Rose Holley's checklist for crowdsourcing.

The Thing	The System	The People	The Content
Clear goal	Easy and Fun	Acknowledge	Interesting
Big challenge	Reliable and quick	Reward	New
Progress	Intuitive	Team support	Lots
Results	Options	Trust	History/Science

After talking to other crowdsourcing site managers and asking them "what lessons have you learnt?" and taking my own experiences into account I am able to provide a combined, comprehensive list for librarians of tips for crowdsourcing. No site I looked at had utilised all the tips. If a new site follows

these tips I have no doubt they will be very successful. The tips are illustrated with example screenshots where appropriate.

Tip 1: Have a transparent and clear goal on your home page (which goal MUST be a BIG challenge). You might know what you want to do, but you must tell your volunteers this as well and keep the message clear and prominent. It helps if your goal is massive and appears to be unachievable, or if you keep upping the size of your goal or the task.

Fig. 9. Goal - Family Search Index.

Fig. 10. Goal - Australian Newspapers.

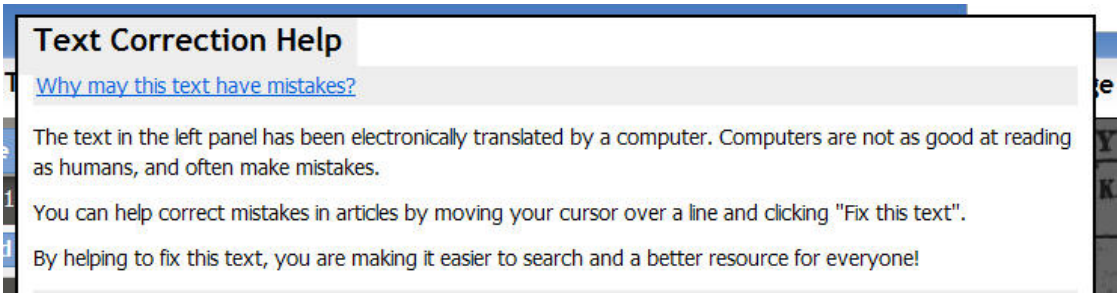
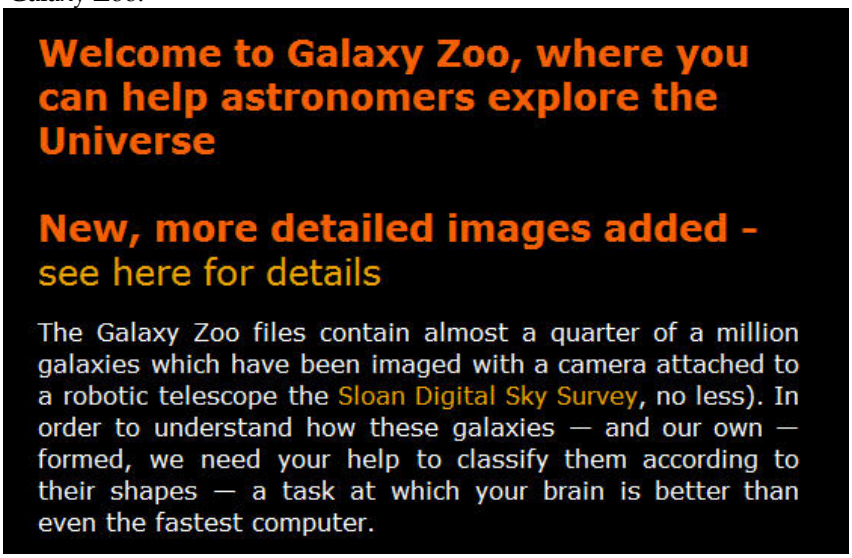


Fig. 11. Goal - Galaxy Zoo.



Tip 2: Have a transparent and visible chart of progress towards your goal.

Simon Willison of the Guardian says "Any time that you're trying to get people to give you stuff, to do stuff for you, the most important thing is that people know that what they're doing is having an effect. It's kind of a fundamental tenet of social software. … If you're not giving people the 'I rock' vibe, you're not getting people to stick around." You must let volunteers know how well they are doing to keep their motivation up.

Fig. 12. Progress - The Guardian MP's expenses – home page.

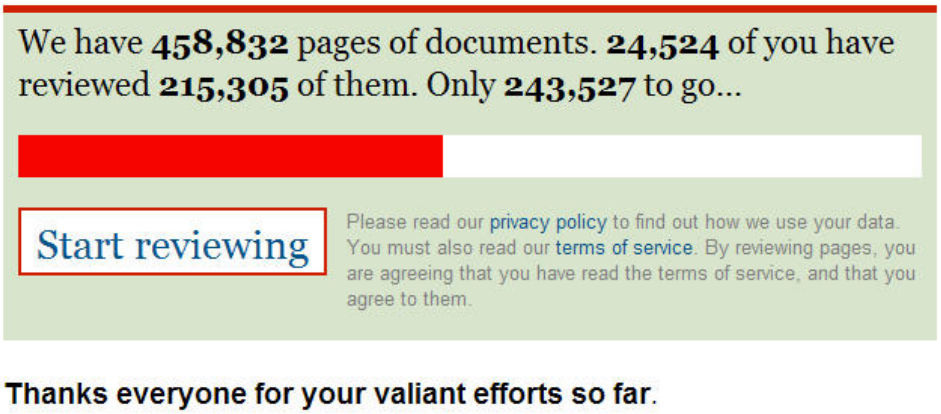


Fig. 13. Progress - The Guardian MP Expenses – MP page.

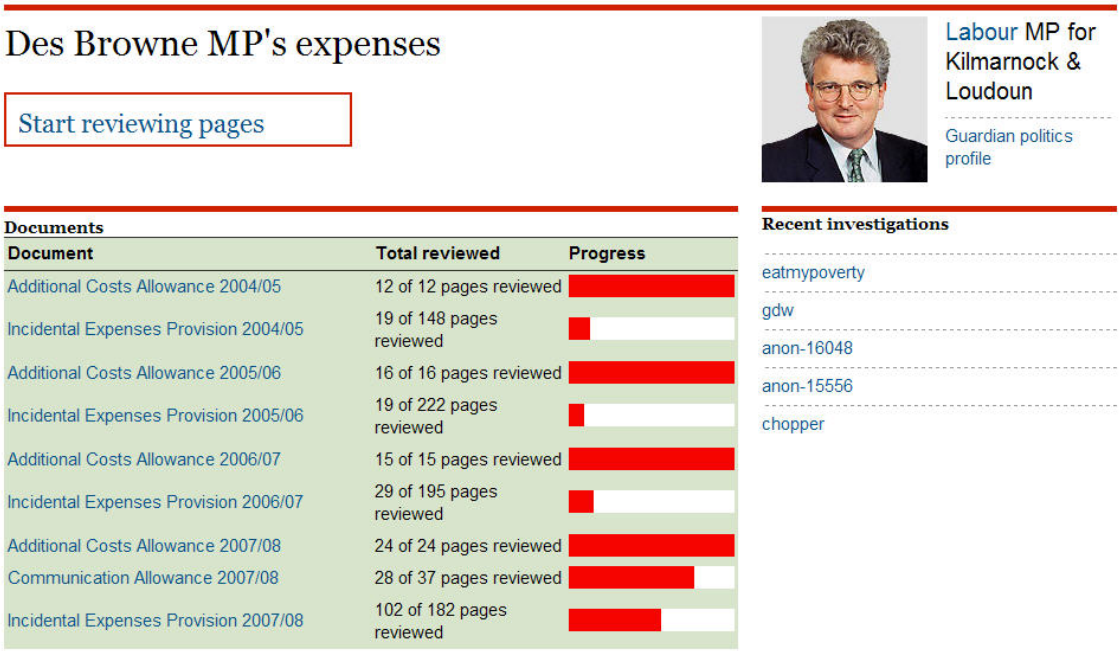


Fig. 14. Progress – the Distributed Proofreaders.

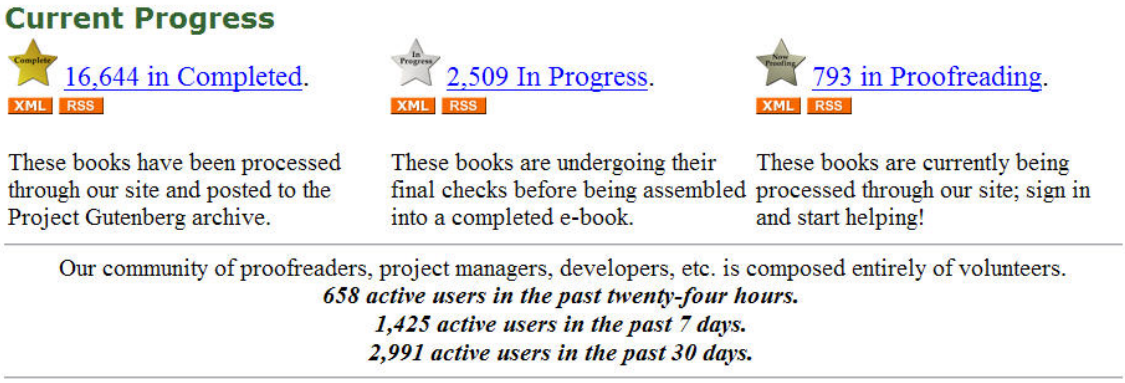


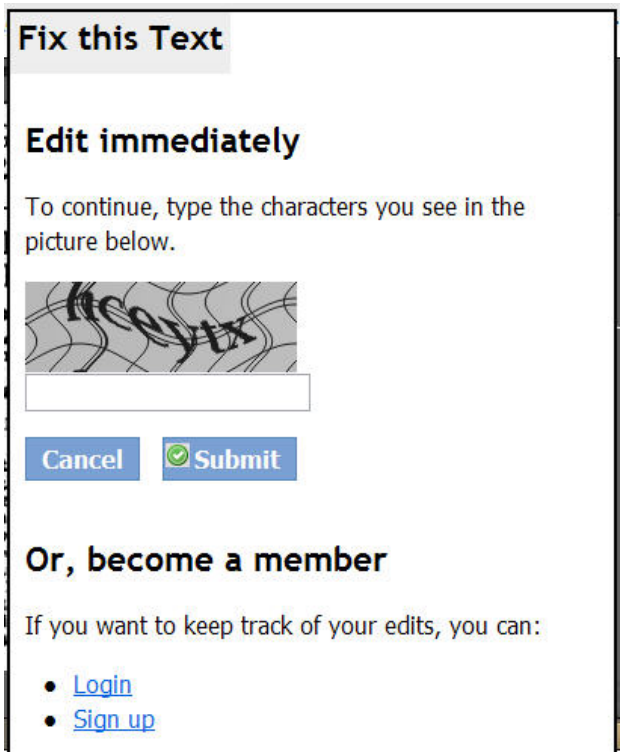
Fig. 15. Progress – Wikipedia.

Fig. 16. Progress – Galaxy Zoo.

Tip 3: Make the overall environment easy to use, intuitive, quick and reliable.

The usability of your system is really important. If you haven't done usability testing beforehand then solicit and act on user feedback as you develop the system. Load testing is also necessary since speed and reliability of the system are as important as usability. Uses will drop off as quickly as they came if the system fails them in this respect. Simon Willison of the Guardian said "We kind of load-tested it with our real audience, which guarantees that it's going to work eventually." Wikipedia noted in their statistics "The big slowdown in the rate of article creation in June–July 2002 was caused by major server performance problems, remedied by extensive work on the software". Make it as easy as possible for users to do work and options for this include no requirement to login (e.g., Australian Newspapers), a very simple sign up process and an easy login method.

Fig. 17. Easy and Quick to use – Australian Newspapers no login required.



Tip 4: Make the activity easy and fun.
Make it like a game and keep it light hearted. After all the volunteers are doing this for nothing. The easier and more fun it is the more likely they are to join in and keep doing it.

Fig. 18. Easy and Fun – The Guardian MP's expenses.

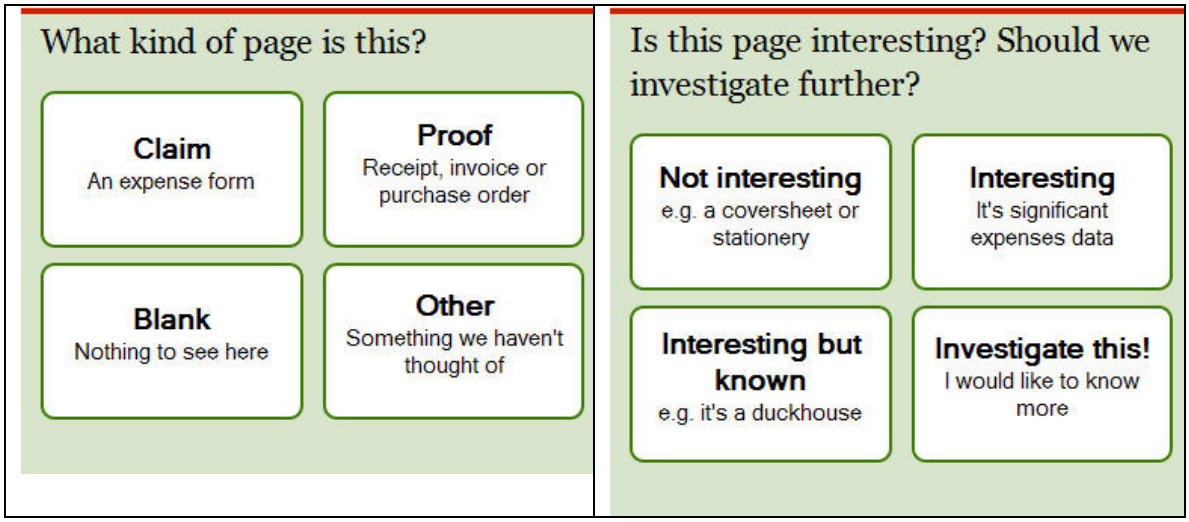


Fig. 19 Easy and Fun – Galaxy Zoo.

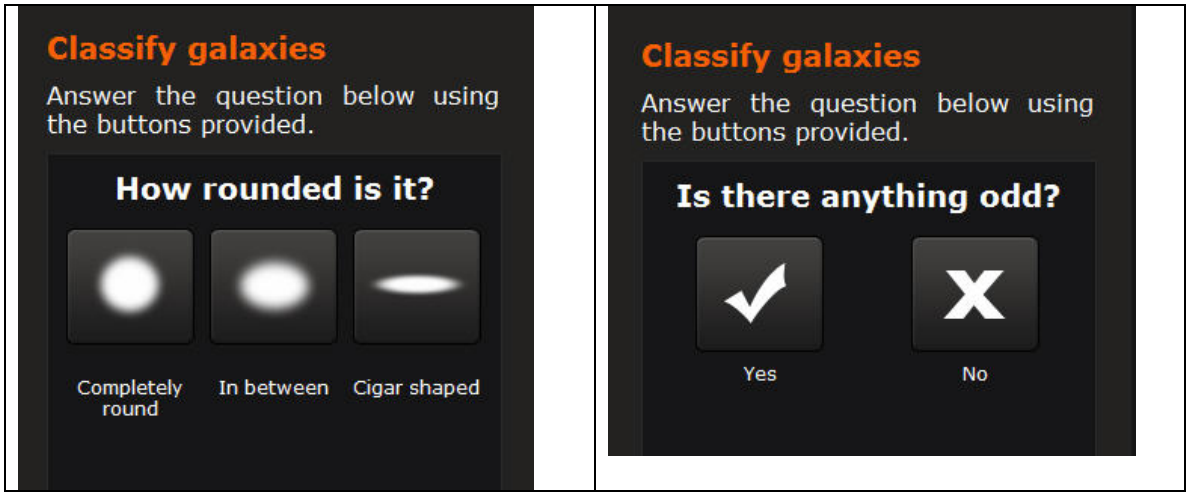
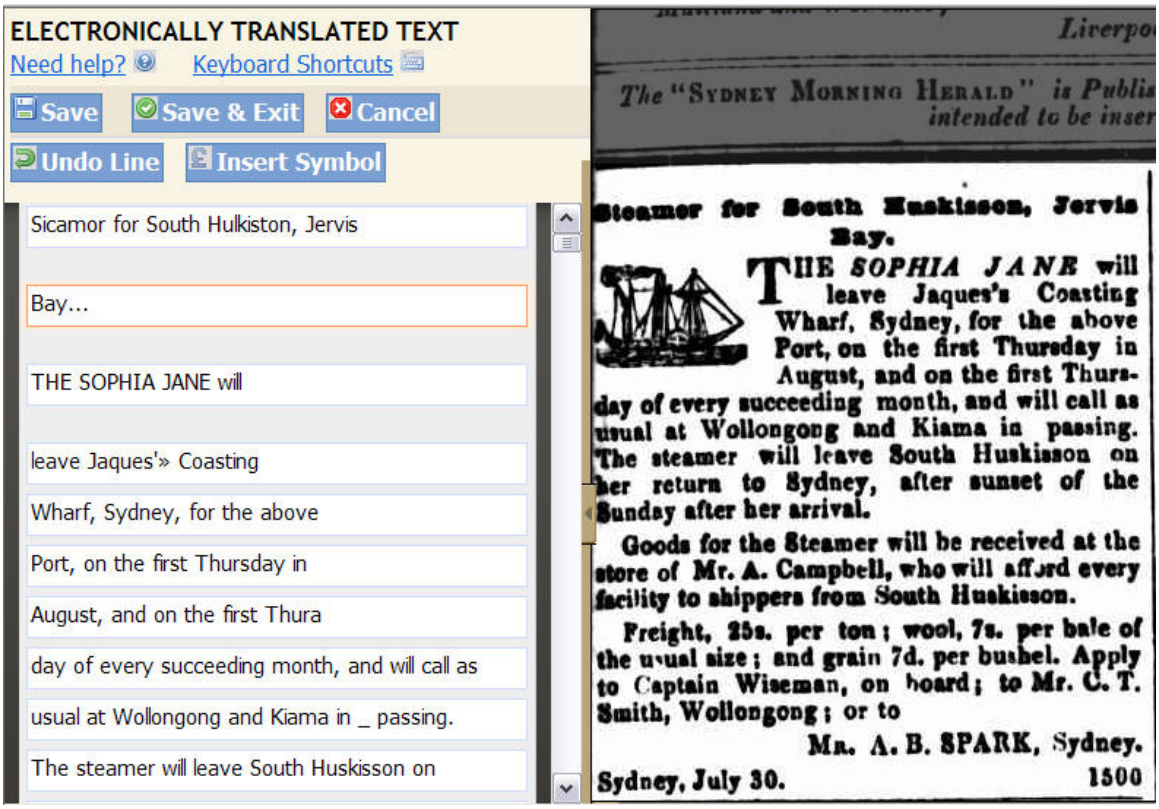


Fig. 20. Easy and Fun- Australian Newspapers.



Tip 5: It must be interesting.
Something about the content must be really interesting. Regular drawcards are history, science, personal lives, scandals, genealogy, and animals.

Tip 6: Take advantage of transitory and topical events if they help you.
Topics in the news can have a big impact on user activity, the primary example being the UK MP scandal which led to 20,0000 volunteers signing up and completing work in the first two weeks. Take advantage of news events, special occasions, historical anniversaries etc. to get your volunteers to do more work or special targeted bits.

Tip 7: Keep the site active by addition of new content/work.
Keeping the site current, expanding the amount of data and continually developing the site are all important in keeping your volunteers motivated. Crowdsourcing is still quite new and there is no one site that has got it all totally right. New development as well as new content keeps volunteers busy.

Tip 8: Give volunteers options and choices
Some volunteers like to be able to choose subjects, and types of work they do, whilst others prefer to be directed to what to do next. Good ideas are to have a 'pick your work' and a 'do the next thing that needs doing' option. Distributed Proofreaders, the Guardian MP Expenses and Wikipedia all give both choices. In addition, Wikipedia and FamilySearchIndexing give options to select projects from different countries, and Wikipedia and Galaxy Zoo also have different language interfaces. Distributed proofreaders, Wikipedia and others also have different tasks for novices and experienced users. All these things extend your volunteer base.

Tip 9: Make the results/outcome of your work transparent and visible.
Some interesting results (aside from achieving your main goal) may come out of the work and it is important to share these with the volunteers as soon as you find them, because if it wasn't for them the those results would never have happened. Volunteers are especially interested in new discoveries and research arising out of what they have done.

Fig. 21. Results – Galaxy Zoo.

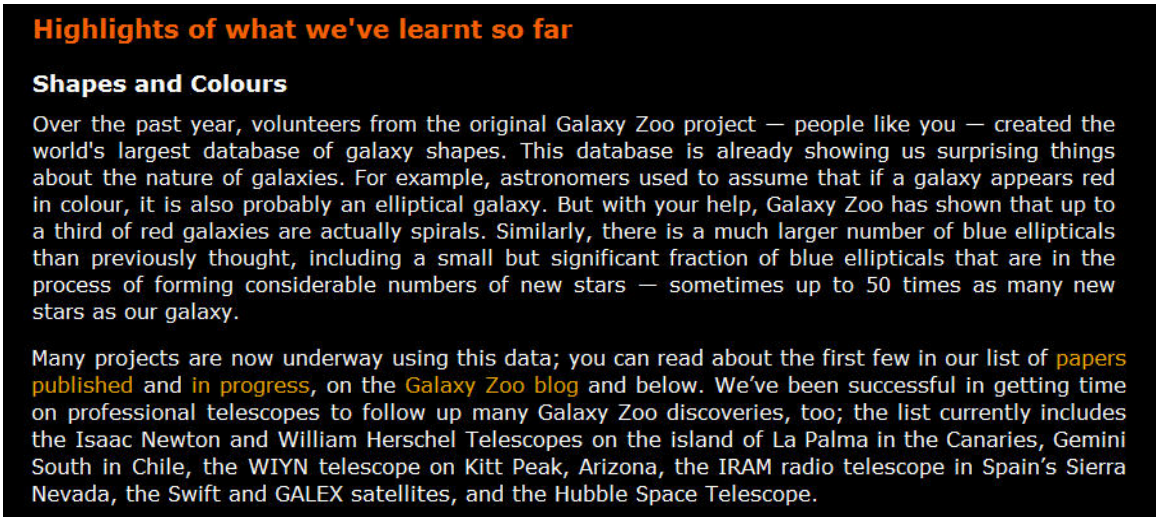


Fig. 22. Results – Distributed Proofreaders.

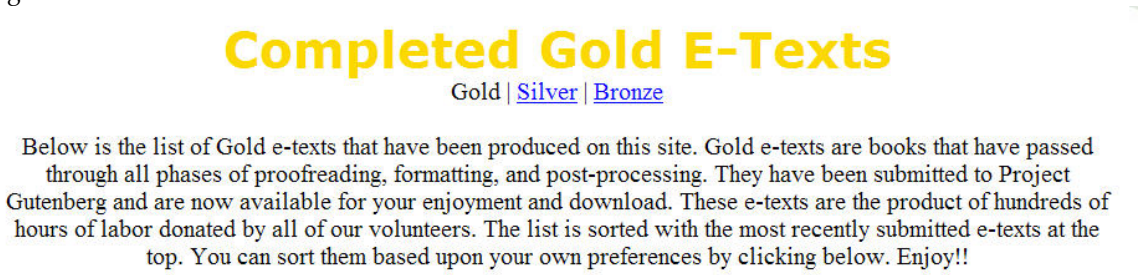


Fig. 23. Results – The Guardian's MP's expenses.

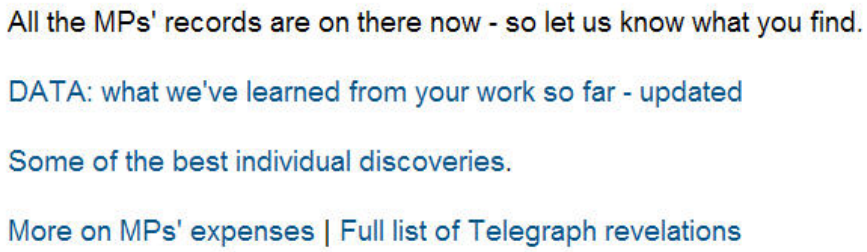


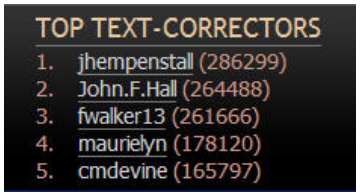
Fig. 24. Results – BBC People's WorldWar2.

Tip10: Let volunteers identify and make themselves visible if they want acknowledgement.
Let users have the option to use/display their real name or a pseudonym, to add a photograph, and to make themselves visible to other volunteers and/or the users of the site. Names may be credited on the site and/or against the item they have worked on.

Fig. 25. Acknowledgement Picture Australia.

Tip 11: Reward high achievers by having ranking tables and encourage competition.
Putting in a ranking table is easy and costs nothing, yet it makes a big difference to your volunteers and helps to motivate them.

Fig. 26. Reward – Australian Newspapers.



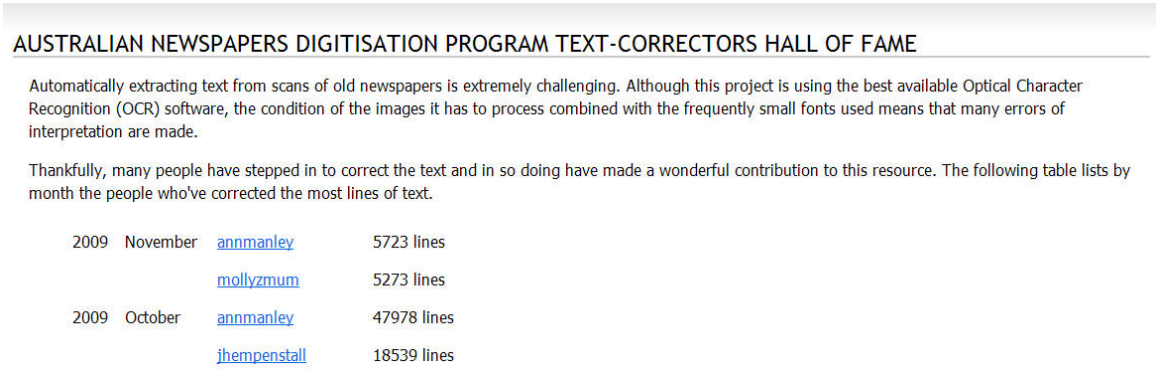
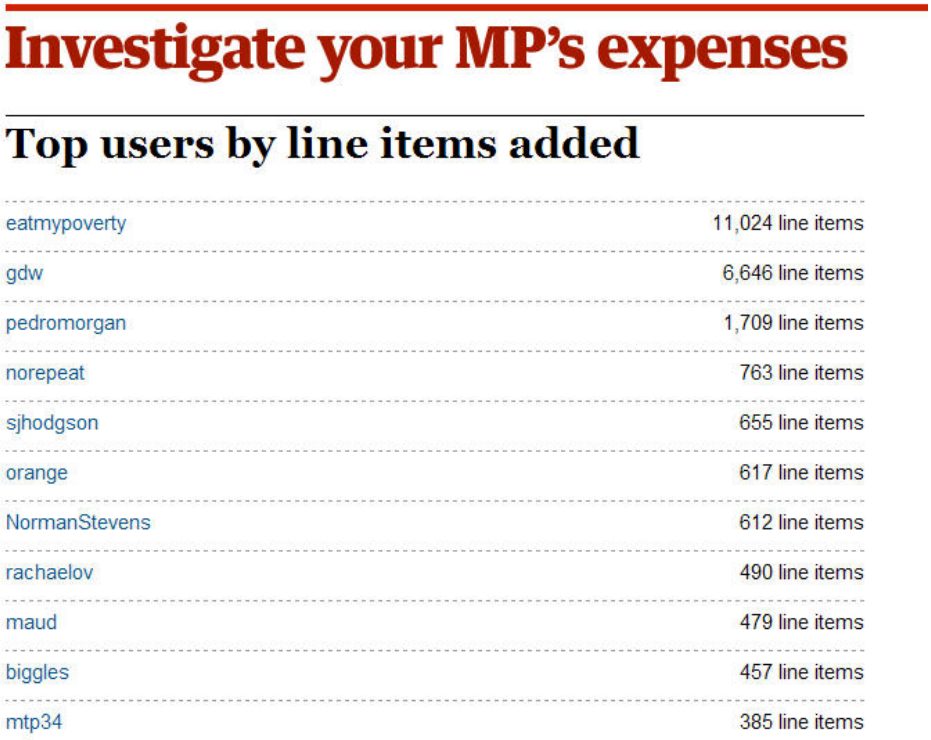


Fig. 27. Reward – Guardian's Expenses.



Top users by pages reviewed

eatmypoverty	26,384 votes
gdw	10,122 votes
norepeat	3,440 votes
egg	2,415 votes
Billfred	2,287 votes
pedromorgan	1,769 votes
maud	1,728 votes
biggles	1,322 votes
NormanStevens	918 votes
anon-17732	900 votes
anon-4353	888 votes
shelleyb	877 votes
christo_method	873 votes

Tip 12: Give the volunteers an online team/communication environment to build a dynamic, supportive team environment.

There are plenty of online forum and wiki packages to choose from. Galaxy Zoo has an extensive communication environment with forum, blog, and FAQ.

Tip 13: Treat your 'super' volunteers with respect and listen to them carefully.

Remember that your 'super' volunteers are usually doing more than half of the entire work so treat them well. Anything they have to say to you should be of interest, especially their feedback on your system and process because they are your heaviest site users. If they recommend something, give it go.

Tip 14: Assume volunteers will do it right rather than wrong.

Experience shows that the greater the level of freedom and trust you give to volunteers the more they reward you with hard work, loyalty and accuracy. Rather than assuming everything will go wrong and spending valuable time putting systems in place to stop vandalism, assume volunteers will do their best and monitor and help each other. Give them as much freedom as you are able.

7. Next steps for libraries

The potential of crowdsourcing for libraries is huge. Libraries have a massive user base and both broad and specific subject areas that have wide appeal. Libraries could get hundreds of thousands of volunteers if they really publicized and appealed for help. Anyone with an Internet connection is a potential volunteer. A shift in thinking is required by libraries to fully embrace the potential of crowdsourcing. Up until now libraries have been in control of creating, collecting and describing data and have held the power to do this. Individuals as well as organisations such as Google can now easily create, organise and describe content which is seen by some as threatening the role of libraries. Giving users the freedom to interact with and add value to data as well as create their own content and upload

it into our collections is what users want, and helps libraries maintain their relevance in society. Libraries need to think globally rather than individually about crowdsourcing. There could be a centralized global pool of volunteers and projects rather than each library trying to establish its own volunteer base. Digital users do not care about institutional walls. Libraries know this and have worked hard to break down the walls or make them invisible to users when they are searching digital collections. There should be no walls in crowdsourcing projects either. For example if someone wants to improve resources on shipping lists – they should be able to come to a central portal to find all the projects and countries involving shipping lists instead of having to discover and find for themselves that there are shipping lists which can be corrected in 3 different projects, for example, Australian Newspapers, Mariners and Ships in Australian Waters, and FamilySearchIndexing.

Libraries and archives will never have the resources to fully do what they or the users want so crowdsourcing is an opportunity that should be seriously considered. Governments in the UK and Australia are now taking digital social engagement seriously and looking at developing policies for this and the utilisation of web 2.0 technologies into their government departments, including libraries. The Australian Government 2.0 Taskforce[16] is in the final stages of preparing its recommendations. But libraries do not need to wait for taskforce instructions to engage with users, this has been their 'modus operandus' for years, and is seen by many librarians simply as excellent customer service. So what should libraries be thinking about in relation to crowdsourcing?

- How can we build back digital social engagement into all our activities as normal business?
- What do we want help with? (Indexing, transcription, creation of content)
- Why do we want this help? (To improve data quality, to socially engage, to get new content, to stay relevant?)
- Can we build partnerships with existing crowdsourcing non-profit making organisations to share use of existing volunteers, especially Distributed Proofreaders and Wikipedia?
- Can we build our own sites that incorporate the crowdsourcing tips using the open source software developed by others? (Easy, quick, rewards, acknowledge, communicate)
- Is it possible to work towards establishing a global pool of volunteers and a crowdsourcing portal that has no country or organisation boundaries?
- Can libraries market crowdsourcing effectively in the online environment?
- How should we structure the user generated content behind the scenes (add into our own data, or keep in separate locations and layers)?
- When can we build the changes and activity into our strategic plans?
- How we can change the 'power' thinking of strategic staff to 'freedom and potential' thinking to make this happen on a mass scale?
- When and how are we going to start crowdsourcing?

8. Conclusion

Crowdsourcing has not been attempted on any significant scale by libraries to date, but could prove to be the most useful tool a library can have in the future. If the facts known about crowdsourcing and the tips outlined in this article are applied any crowdsourcing project that is 'for the common good' and initiated by a non-profit making organisation such as a library is likely to be successful. If the public are given a high level of trust and responsibility they will respond with loyalty and commitment as has been demonstrated in the crowdsourcing sites discussed. There is huge potential for libraries to harness digital volunteers. Libraries need to give up 'power and control' thinking and look to freedom instead. Harriet Rubin, business publisher and author talking about success says "Freedom is actually a bigger game than power. Power is about what you can control. Freedom is about what you can unleash"[17]. And librarians need to be courageous about this. Dr John C. Maxwell , leadership expert and author talking on how to generate momentum in the workplace says "Passion energizes your talent and rubs off on those around you. If you have courage then you will influence people based on your passionate

convictions"[18] . Do we have the courage, and dare we give users something greater than power – freedom?

9. References and Footnotes

I would like to acknowledge and thank the people below who gave their time to me to provide statistics and quotes, and discuss crowdsourcing techniques and social engagement ideas and strategies. They also provided candid information on the lessons they had each learnt about their respective crowdsourcing projects.
Kent Fitch (Lead System Architect, National Library of Australia)
Juliet Sutherland (Manager, Distributed Proofreaders)
Liam Wyatt (VP Wikipedia Australia)
Erik Moeller (Deputy Director, Wikimedia Foundation)
Mark Kelly (FamilySearch Support Manager Australia/PNG, FamilySearchIndexing).
Fiona Hooton (Manager, Picture Australia)

Urls of crowdsourcing sites:
Australian Newspapers: http://ndpbeta.nla.gov.au
Picture Australia: http://www.pictureaustralia.org/
FamilySearchIndexing http://www.familysearchindexing.org/home.jsf
Distributed Proofreaders: http://www.pgdp.net/c
Wikipedia: http://www.wikipedia.org/
UK MP's Expenses: http://mps-expenses.guardian.co.uk/
Galaxy Zoo: http://www.galaxyzoo.org/
BBC WorldWar2 Peoples War: http://www.bbc.co.uk/ww2peopleswar

Footnotes:

[1] Holley, Rose (2009) Crowdsourcing and social engagement: Potential, Power and Freedom for Libraries and Users. Research Paper. http://eprints.rclis.org/17410/

[2] Holley, Rose (2009) Crowdsourcing and social engagement: Potential, Power and Freedom for Libraries and Users. PowerPoint presentation given at PRDLA 2009 meeting. http://www.prdla.org/2009/10/crowdsourcing-and-social-engagement

[3] Wikipedia (2009) Article on crowdsourcing: http://en.wikipedia.org/wiki/Crowdsourcing viewed August 2009.

[4] Holley, Rose. (2010) Tagging full text searchable articles: An overview of social tagging activity in historic Australian Newspapers August 2008 – August 2009. D-Lib Magazine, vol 16, no 1/2, 2010 http://www.dlib.org/dlib/january10/holley/01holley.html

[5] Sharky, Clay (2009) Here Comes Everybody: How Change Happens When People Come Together, Penguin Books, ISBN 9780141030623.

[6] Holley, Rose. (2009) Many Hands Make Light Work: Public Collaborative OCR Text Correction in Australian Historic Newspapers, National Library of Australia, ISBN 9780642276940 http://www.nla.gov.au/ndp/project_details/documents/ANDP_ManyHands.pdf

[7] Hooton, Fiona. (2006) Picture Australia and the flickr effect. Gateways, number 80, April 2006. http://www.nla.gov.au/pub/gateways/issues/80/story01.html

[8] Lih, Andrew (2009) The Wikipedia Revolution: How a bunch of nobodies created the world's greatest encyclopedia, Aurum Press Ltd, ISBN 9781845134730

[9] Wikipedia (2009) Wikipedia Statistics: http://en.wikipedia.org/wiki/Wikipedia:Statistics. Viewed October 2009.

[10] Voss, Jakob (2005). Metadata with Personendata and beyond. Wikimania 2005 paper. http://meta.wikimedia.org/wiki/Transwiki:Wikimania05/Paper-JV2

[11] Danowski, Patrick (2007). Library 2.0 and User Generated Content: What can the users do for us? Staatsbibliothek zu Berlin. 73rd IFLA Conference Paper, 19-23 August 2007, Durban, South Africa. http://ifla.queenslibrary.org/IV/ifla73/papers/113-Danowski-en.pdf

[12] Anderson, Michael (2009) Four crowdsourcing lessons from the Guardian's (spectacular) expenses-scandal experiment. June 23 2009. http://www.niemanlab.org/2009/06/four-crowdsourcing-lessons-from-the-guardians-spectacular-expenses-scandal-experiment/

[13] Wikipedia (2009) Article on Galaxy Zoo: http://en.wikipedia.org/wiki/Galaxy_Zoo viewed October 2009.

[14] BBC (2009) WW2 People's War: Project History: How the site worked: User Journey. http://www.bbc.co.uk/ww2peopleswar/about/project_07.shtml#personalpages. Viewed August 2009.

[15] Jorgensen, John (2007). 21 Proven Motivation Tactics. Published online in Pick the Brain, August 23 2007. http://www.pickthebrain.com/blog/21-proven-motivation-tactics/

[16] (2009) Australian Government 2.0 Taskforce website, blog, draft report: http://gov2.net.au/

[17] Fast Company (1998) The Fast Pack. Fast Company Magazine, Issue 13, January 31, 1998, page 5. http://www.fastcompany.com/magazine/13/fastpack.html

[18] Maxwell, John C (2010) Momentum breakers vs momentum makers. http://www.giantimpact.com/articles/read/article_momentum_breakers_vs_momentum_makers/