Project:	Synthesis of systematic resources
Project acronym:	SYNTHESYS3
Grant Agreement number:	312253
Workpackage:	3: Innovation, Impact and Sustainability
Deliverable number:	3.4
Deliverable title:	Crowdsourcing report: Phase 2
Deliverable author(s):	Laurence Livermore[1], John Tweddle[1], Lisa French[1], Sarah Phillips[2] ,Lucy Robinson[1], and Vincent S. Smith[1] [1] The Natural History Museum, London [2] RBG Kew, Surrey
Date:	February 2015

Making molehills out of mountains: crowdsourcing digital access to natural history collections

SYNTHESYS3 NA3 - Work Package 3 Innovation, Impact and Sustainability

Task 1.4 Pilot study into optimal crowdsourcing processes for NH institutions

Report Authors: Laurence Livermore[1], John Tweddle[1], Lisa French[1], Sarah Phillips[2] ,Lucy Robinson[1], and Vincent S. Smith[1]

[1] The Natural History Museum, London
[2] RBG Kew, Surrey

Report Background

SYNTHESYS3 is a European Union-funded Integrated Activities grant which aims to create an accessible, integrated European resource for researchers in the Natural Sciences. The Joint Research Activity (JRA) is one of its three main activities and aims to improve the quality of and increase access to digital collections and data within natural history institutions' virtual collections. The Networking Activity 3 (NA3) ensures that SYNTHESYS3 has a clear pathway to impact by making certain that outputs are well publicised to, and utilised by, the broader research and development community – and that the work undertaken and benefits created continue beyond the life of the project

One of the JRA objectives is to implement a crowdsourcing platform for SYNTHESYS3 collaborators and one of the NA3 objectives collaborates with the JRA to review optimal crowdsourcing processes for natural history institutions. As part of the NHM's and RBG Kew's contribution to these objectives we have produced a report on the current approaches and underlying technology for crowdsourcing metadata enrichment of digital images. This report provides recommendations for developing a crowdsourcing strategy for natural history collections. The initial research for the report was presented at The Society for the Preservation of Natural History Collections (SPNHC) Conference in June 2014. The findings of the report were presented to the NHM's Digital Collections Programme Project Board in January 2015.

Table of Contents

Abstract

The development of cheap mass imaging techniques has made transcription the bottleneck in digitising natural history collections. Crowdsourcing, or in this context “crowdscience”, has been successfully used across different scientific communities and has the potential to become an effective method for transcribing and georeferencing natural history data. We review current crowdsourcing platforms and communities and discuss their potential role in improving digital access to natural history collections worldwide.

Introduction

Crowdsourcing is a relatively new term but the idea of scientific research being conducted by volunteers or amateurs is well established in both the natural history community and in online communities. There is a lack of a consistent definition for crowdsourcing which for science-based projects, fall under the broad umbrella of citizen science. Estelles-Arolas *et al* (2012) attempted to reach a consensus definition and came up with the following properties:

1. Involves a crowd
2. Has a clear task and goal
3. Rewards the crowd
4. Has a distinct crowdsourcer
5. Benefits the crowdsourcer
6. Is an online and open participatory process

There have been numerous examples of successful scientific and cultural crowdsourced (also referred to as crowdscience) projects over the past ten years (Westphal et al, 2005; Lintott et al, 2008; Owens, 2011; Ridge, 2014). The largest of these is Zooniverse, a web portal hosting over 20 distinct projects, with more than 1.1 million users (Zooniverse, 2014). During the first year of Zooniverse’s Galaxy Zoo project, 150,000 volunteers classified more than 50 million galaxies (Galazy Zoo, 2014), and during the 14 months that Galaxy Zoo 2 was active it is estimated that total volunteer effort was equivalent to that of one person working for >110 years (Lintott, no date).

The range of crowdsourcing tasks is diverse and ranges from transcription of natural history specimen labels and diaries to decision tree-based classification of galaxy morphology (Galaxy Zoo, 2014). A common feature of these projects is that they are based on tasks that are well-suited for human intelligence rather than computers.

Many scientific crowdsourcing projects are research-based with a goal of using the results in formally published papers, sometimes with significant contributing members of the “crowd” community as co-authors for example Hanny van Arkel in the discovery of a new astronomical

object (Józsa et al, 2009). Often the resulting data is open and freely usable increasing its potential future impact on scientific research.
Inspiring volunteers to participate and then rewarding their efforts is a key element of project design. Understanding the interests, motivations and expectations of potential volunteers is critical if a project is to succeed in yielding both high quality scientific data and a rewarding volunteer experience. Commonly cited reasons for initial participation in a crowdsourcing project include interest in the subject area, the opportunity to learn new knowledge, altruism, social interaction and involvement in a collective endeavour that the participant is confident will generate useful scientific findings. Methods for supporting and rewarding on-going involvement include rapid feedback, the opportunity to progress and develop skills and knowledge, user-focused forums, gamification, leaderboards and digital badges.

Natural history collections are vast and exciting treasure troves of biodiversity data but the process of digitising them still remains a mammoth task. It has been estimated that the total number of specimens in natural history collections is around 2 billion (Ariño 2010). In the past three years multiple natural history collections and organisations have launched their own crowdsourcing projects. These focus predominantly on transcription of either specimen labels, collection registers or diaries.

Applied human intelligence is still required for label transcription, even with the application of optical character recognition and computer vision software. Some of the issues that are very challenging to solve computationally are:

- Diversity and irregularity of labels e.g. shape, size, contents (Figure 1)
- Recognising and mapping of label data to atomised fields is complex
- Label data can be duplicated
- Label data can be irrelevant or contradictory
- Mixture of handwritten and printed text (Figure 1)

Figure 1 - Examples of insect labels in multiple languages with historical taxonomic names, locality names, handwritten and printed text.

From all the projects we observed, all seem to get completed. On the Zooniverse platform Notes from Nature was the 10th most popular project although it's not stated how popularity is determined (02/10/14). On the Smithsonian Transcription Centre the Bumblebee project Set 12 was the 8th most popular project (02/10/14). Note that book (diary/register/field notebook) transcription projects are ranked more highly and some are natural history related. If looking at complete projects: 3 natural history Passiflora Plants projects are in the top 10 (3, 5 & 7). As with Notes from Nature it is unclear how popularity is being ranked.

The limiting factor appears to be the time and preparation to create projects and their images on the platform rather than a lack of interest or participation from volunteers. A lack of support for the volunteers in the participating natural history institute, e.g. having a dedicated community manager providing timely responses and interactions, can also be a limiting factor on creating new projects (Sarah Phillips per. comms., 2014).

In summary, it is already evident that the "crowd" of volunteers can do great things with data. Not only with transcriptions, which is the focus of most crowdsourced natural history projects, but with a range of other tasks such as classification and assisting 3D object reconstruction (MicroPasts, 2014) that may be possible to utilise in the future.

This report:

- Reviews platforms used for crowdsourcing projects in natural history collections;
- Provides case studies of natural history crowdsourcing projects;
- Summarises the motivations of volunteers;
- Recommends strategies for successful crowdsourcing in natural history collections;

Crowdsourcing Platforms

In the following section crowdsourcing platforms are described and compared.

The prerequisites for a platform to be included in this report were that it:

- had at least two distinct transcription-based natural history projects
- was currently active and openly accessible to new volunteers

Platforms in scope (abbreviations used in brackets):

- Atlas of Living Australia: DigiVol (DV) - http://volunteer.ala.org.au/
- herbaria@home (h@h) - http://herbariaunited.org/
- Les Herbonautes (LH) - http://lesherbonautes.mnhn.fr/
- Notes from Nature (NfN) - http://www.notesfromnature.org/
- Smithsonian Digital Volunteers: Transcription Center (TC) - https://transcription.si.edu/

Overview of major platform differences

Among the five platforms the method of data entry for transcription takes one of two forms: multikeying, where multiple unique volunteers transcribe data multiple times without verification or reviewing (NfN, LH); single keying usually accompanied by a data review process either by volunteers or a project administrator (DV, TC, h@h).

The support for project management and administration as a service differs greatly between platforms. DV has data templates, project creation wizards and user management available to project administrators in contrast to h@h which is solely managed by the platform administrator.

The georeferencing tools available to projects also varies between platforms: in NfN there are no inbuilt georeferencing tools; DV provide a tool that allows volunteers to manually or automatically generate coordinates with an uncertainty radius.

There is generally little support among the different platforms for transcription using a mobile device. Some platforms have more responsive/fluid layouts than others but none have a dedicated app or mobile version of the page. To our knowledge none of the platforms have shared data about the type of devices volunteers use to contribute to projects.

Atlas of Living Australia: DigiVol

Atlas of Living Australia's DigiVol (DV) is a general natural history collections platform that was publicly launched in 2011. The scope of projects in the DV is not limited to Australian collections, nor Australian specimens.

DV organises crowdsourcing projects into separate 'expeditions', and promotes different projects as the 'Virtual expedition of the day'. The homepage for DV displays a list of some expeditions with a percentage for overall completion and an honour board, showing the top daily, weekly, monthly and overall volunteer.

Each expedition has its own homepage which gives a brief overview of the project and displays a map of where transcribed specimens have been collected. Within each expedition is a list of volunteers who have contributed to the project, with each volunteer given a special title depending on the number of transcriptions they have completed.

DV requires the volunteer to register and login before transcribing. Registration is for the whole Atlas of Living Australia website which provides services and citizen science activities in addition to DV. During registration there are a number of required fields, including two fields asking the volunteer what their primary and secondary use of the website will be.

The volunteer transcription interface varies slightly between expeditions depending on which fields are displayed, for example, not all projects use DV's mapping tool. All the fields are displayed on screen underneath a large zoomable image of the object to be transcribed **(Figure 2)**.

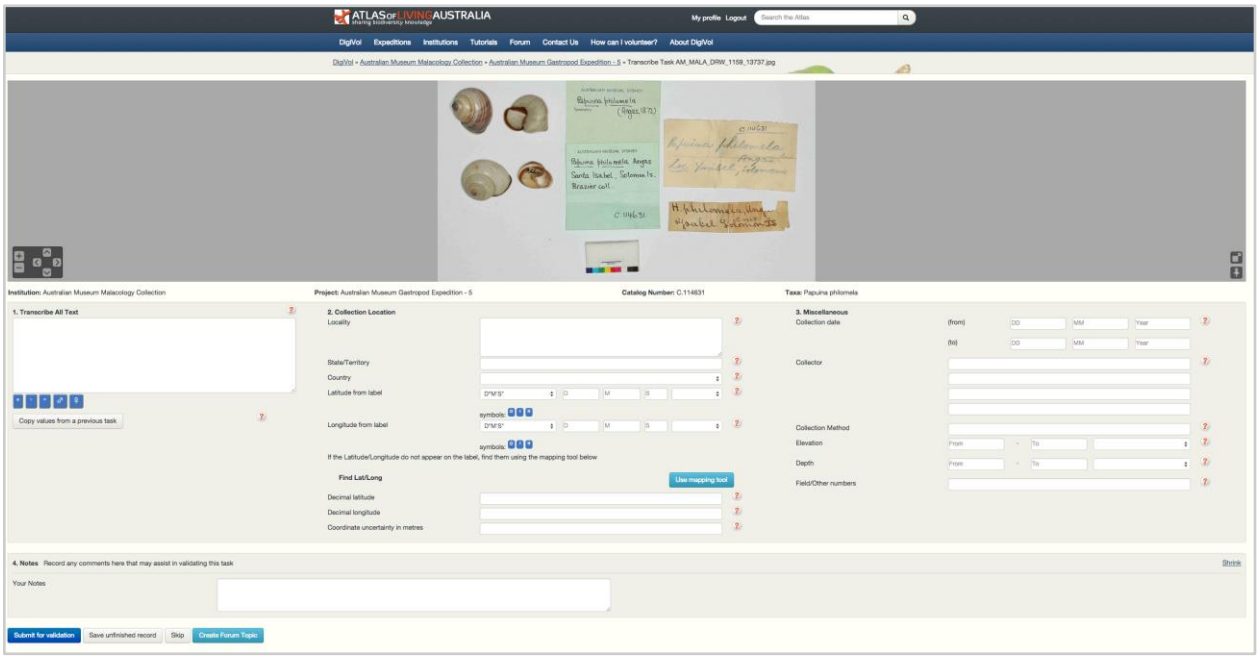

Figure 2 - Atlas of Living Australia's DigiVol transcription interface

Each field is transcribed once and subsequently verified by project administrators or volunteers with specific verification permission. The mapping tool attempts to automatically determine coordinates based on volunteer entered values but will allow the volunteer to manually select the location of the specimen on a map. Volunteers can also specify an uncertainty radius for the location.

Help is provided to volunteers through a generic specimen label tutorial, expedition-specific tutorials and field-specific help text in the transcription interface. All of the help materials can be managed by the institution running the expedition. Volunteers can also record comments to assist with record validation or start a forum topic to discuss a particular task with the DV community and expedition administrator. Automatic notifications can be setup so that administrators are aware of new comments or posts.

DV allows volunteers to view their transcription history with links to previously submitted or saved transcription tasks. Depending on the status of the task a volunteer can return to a task and make further edits. Volunteers can also see whether their tasks have been verified.

DV is used by multiple Australian and non-Australian organisations for transcription and crowdsourcing, some of which have run multiple expeditions. Institutions can create their own institutional page with a picture and short description under which their expeditions are listed, which enables volunteers who have a specific affinity with a particular institution to easily identify the expeditions belonging to that institution.

Atlas of Living Australia and DigiVol are open source and available from https://github.com/AtlasOfLivingAustralia/volunteer-portal

The distinguishing features of DV are:

- Mapping tool - none of the other platforms have a mapping tool
- Complete transcription - the only platform where projects undertake comprehensive label transcription
- Rich administration and expedition management tools (uses TDWG standards)
- Support for creating institutional subsites

herbaria@home

herbaria@home was one of the first dedicated tools for transcribing natural history data, with a public launch in 2007. It focuses exclusively on the transcription of herbarium sheets from UK and Irish collections, and is supported by the Botanical Society of the British Isles (BSBI).

Current active projects are listed on the h@h homepage along with a progress summary. Projects do not have their own specific project pages, but volunteers can click to be allocated a set of herbaria specimens to transcribe from a particular project or herbarium. The website also gives volunteers the option to search for herbarium sheets to transcribe from a particular genus or species, a feature which is unique to h@h. Volunteers can sign up for an e-mail via a 'watchlist' when sheets of particular taxa, from specific localities or collectors have been added to the site.h@h requires the volunteer to register and login before transcribing. Registration requires the volunteer's first name, surname and email address.

The volunteer transcription interface does not vary between projects. All fields are presented on screen with both a large image of the herbarium sheet along with a thumbnail for quickly moving between areas of interest when zoomed in (**Figure 3**). Each field is transcribed once and subsequently verified by the community of volunteers. The current taxonomic name is usually provided but it is possible for volunteers to record previous combinations and determinations. The georeferencing fields make extensive use of autocomplete and dropdown lists to reduce error. h@h is the only platform which supports Ordnance Survey grid references. The main fields all have 'illegible' flags.

Help is provided to volunteers through a herbarium sheet tutorial and through extensive hover-over help text in the transcription interface. The website allows volunteers to request feedback about the sheets transcribed, which will post a message on the forum. This allows the community at h@h to help the volunteer complete the transcription, or answer any questions. The h@h user community have gathered and written reference information to aid with transcription such as on collectors, maps and gazetteers and taxonomic resources.

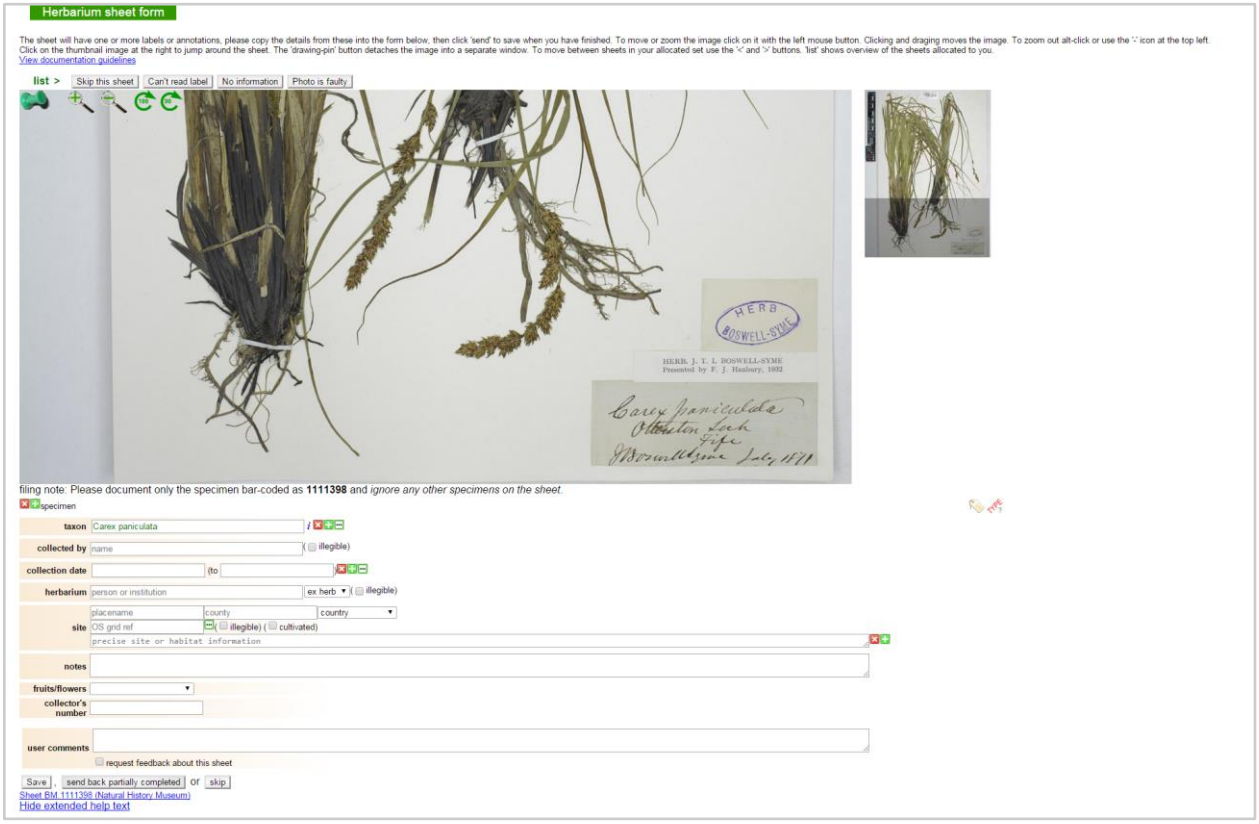

Figure 3 - herbaria@home transcription interface

h@h allows volunteers to view their transcription history including viewing previously submitted or saved transcription tasks. Depending on the status of the task a volunteer can return to a task and make further edits. Volunteers can also see whether their tasks have been reviewed by another volunteer.

h@h has been used to transcribe herbarium sheets from more than 15 herbaria, with collection sizes ranging from small (<500 sheets) to large (>100,000). All herbaria are either in the UK or in Ireland as are the majority of transcribed specimens.

Over 140,000 British herbarium sheets have been transcribed since June 2006. The most specimens transcribed in a day was 274 in 2010, with the most specimens per person a day being 210 in 2012. There have been periods where no specimens have been available on the h@h website to transcribe. The top ten users on the website have contributed 61% of the data, and the top 20 users 81%. Most of these top users have been active members for several years.

h@h is currently closed source but the code is potentially available upon request from the project's lead software architect, Tom Humphrey.

The distinguishing features of herbaria@home are:

- Transcription projects of very different sizes (<100 to >100,000)
- Watchlist - notifies volunteers of new sheets that meet their criteria of interest
- Volunteers can choose to be assigned sheets from particular taxa/localities
- Community-created resources and wiki
- Structured recording of previous combinations and determinations

Les Herbonautes

Les Herbonautes is a French language crowdsourcing website that publically launched in 2012 and focuses exclusively on the transcription of herbarium sheets from the herbarium of the Muséum national d'Histoire naturelle, Paris.

The crowdsourcing projects are split into missions with two promoted missions featured on the homepage. Volunteers can also choose to view all the missions available, with each mission showing a percentage completion rate. The homepage also contains an interactive activity feed, listing the most recent transcriptions by volunteers.

The homepage for each mission gives a brief overview of the project, as well as listing the current mission progress, the number of volunteers and the top contributors. The mission has a sidebar for volunteer discussion, as well as an activity feed showing the transcription activity specific to that mission. There are additional tabs on a project page that list subscribing members, a journal, specimen map and project comments.

LH does not require user registration but anonymous users are always limited to transcription of country and no other fields. Volunteer registration allows for tracking of contributions and requires minimal registration (username, email and password). Volunteers can unlock badges and levels based on the number of transcriptions that they complete. Volunteers with a higher level are able to transcribe more than those with a lower level, for example new "level 1" users are only able to transcribe country whilst experienced "level 6" volunteers can undertake full

georeferencing on a map or record coordinates. Volunteers can have their own avatar, which will show if they make a comment or are a top contributor to a mission.

The volunteer interface is the same for all the missions that were viewed by the authors but most transcription fields (anything beyond country) were unavailable for new users. Transcription fields are displayed in a column on the left hand side of the page with a large zoomable image of the herbarium sheet on the right hand side. Each field can be transcribed multiple times by different volunteers but there is no volunteer review process.

Projects (1st October 2014) are herbarium sheet label transcriptions and are all exclusively from the Paris herbarium (L'Herbier de Paris) although some have research collaborations with other French institutes.

Les Herbonautes is currently a closed-source project but there are plans to release the code under an open source licence in late 2014 (Simon Chagnoux, personal communication, 6th October 2014).

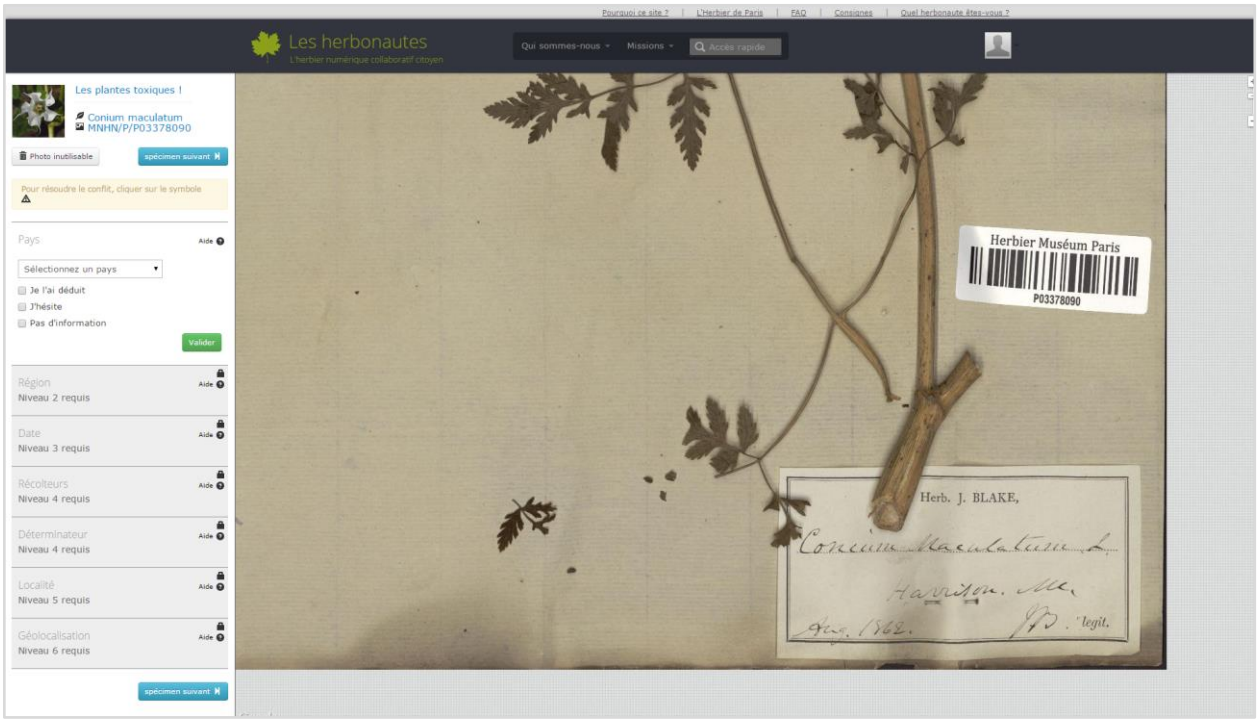

Figure 4 - Les Herbonautes transcription interface

The distinguishing features of Les Herbonautes are:

- Skill/experience based transcription - experienced volunteers can undertake more challenging tasks

- Registration not required - anonymous users limited to transcription of country information
- Interface and collections are French language-based with potential for multilingual support

Notes from Nature

Notes from Nature (NfN) launched in 2013 and is part of the larger Zooniverse crowdsourcing platform, which is a popular citizen science website containing projects on a diverse range of topics. NfN is a general natural history collections platform and has projects to transcribe registers, herbarium sheet data and pinned insect specimen labels.

Projects on the website are split into different collections. As of October 2014 there are four collections: SERNEC, Essig Museum, Natural History Museum and New York Botanical Garden. Specimens are continually added to each collection, rather than creating a new project when one set of data is complete.

Each collection has a homepage detailing further information about the project including the number of transcriptions completed for the project, as well as the number of volunteers taking part but it does not list the top contributors. Volunteers are incentivised through a badging system, with new badges awarded for increasing numbers of transcriptions completed. Each collection awards its own set of badges, for example the New York Botanical Garden awards a ‘Spore’ badge for one transcription, and a ‘Mushroom’ badge for one hundred transcriptions.

NfN does not require registration in order to transcribe and allows anonymous unregistered volunteers to participate in transcribing. Registration for NfN is for the whole of the Zooniverse platform which provides access and volunteer statistics to all the Zooniverse projects/platforms. Required fields are minimal (username, password, email address, real name).

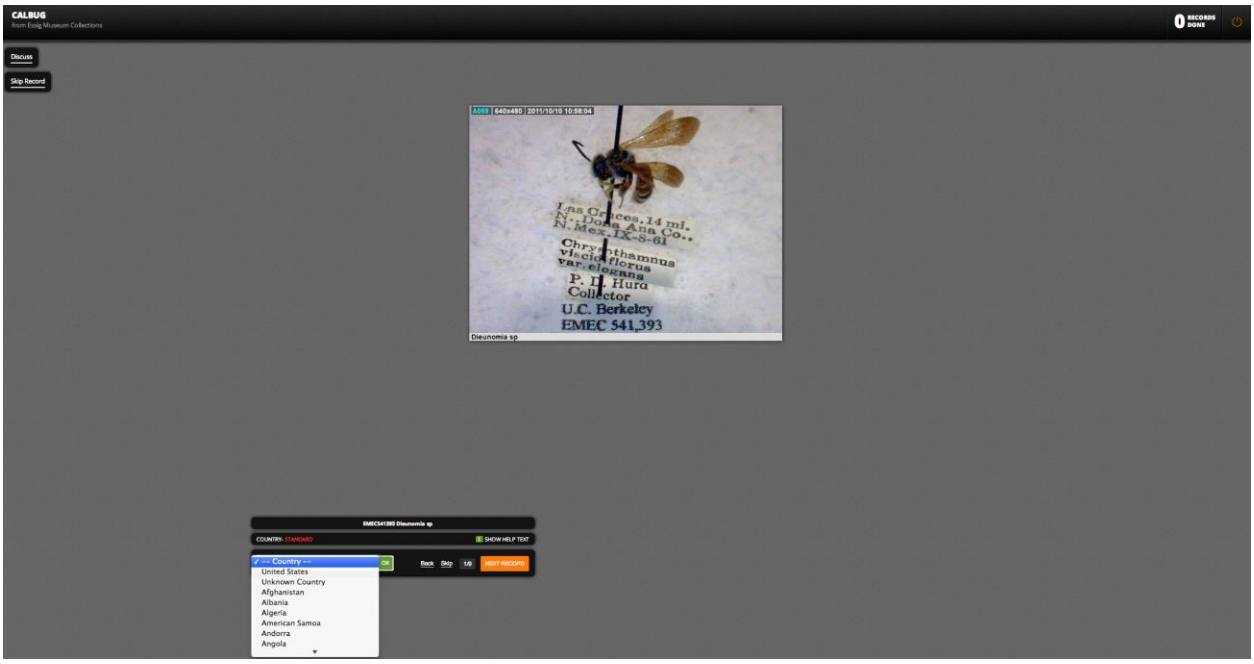

Figure 5 - Notes from Nature transcription interface

The transcription interface is slightly different for each collection. The Natural History Museum interface asks users to transcribe ornithological registers by allowing volunteers to create rows on the image of the book where the text is written. They are then asked to transcribe the information on that row. The New York Botanical Gardens asks volunteers to select the specimen label on the herbarium sheet, and then brings up a dialogue box for the volunteer to complete. There is no option for the user to manually zoom into the specimen for any of the projects. Each field is transcribed multiple times by different volunteers. Volunteers can read how data validation takes place by accessing the NfN blog, which is linked to on each project's homepage under a heading 'Notes on our Numbers'. There are no in-built tools to assist with verifying and managing entries by project managers.

Help text is provided for each field, with the volunteer being able to select 'Show Help Text' to view this. They can also select 'discuss', which will allow them to discuss the specimen on a forum. Selecting discuss also gives the volunteer the option to share the specimen on social media (Facebook, Twitter or Pinterest). Unlike other crowdsourcing websites there is no option to go back and edit a specimen once completed.

Notes from Nature is open source but parts of the Zooniverse code are not (pers. comms. Deb Paul).

The distinguishing features of Notes from Nature are:

- Single/limited transcription fields displayed at one time

- Moveable transcription fields
- Specimens are continually added to projects
- Does not allow users to revisit specimens
- Awards badges for transcribing certain numbers of specimens

Smithsonian Digital Volunteers: Transcription Centre

The Smithsonian Digital Volunteers Transcription Center is a crowdsourcing platform for the transcription of objects from multiple Smithsonian institutions, not just natural history collections. It publically launched in July 2013, with transcription projects including botany specimens, diaries and field notebooks.

The Transcription Centre splits the transcription into projects, which can be searched by looking through project categories e.g. ‘Biodiverse Planet’, ‘Mysteries of the Universe’. Projects can also be browsed by institution.

Project home pages in the transcription centre show a pictured list of all objects within a project. It shows the number of volunteers contributing, and has a project progress bar which shows progress for both transcription and reviewing. The Transcription Centre asks volunteers to both transcribe specimens themselves, but also to review the transcription of other volunteers. The pictured list of objects has a colour indicator showing whether the specimen needs transcribing, or if it needs review by other volunteers (and if so, how many reviews are needed). Volunteers can transcribe either by selecting ‘start transcribing’, in which case a random specimen is allocated, or they can select one of the items in the pictured list.

Volunteers are able to transcribe without logging in, but registration is required to review the transcriptions of others. Registration is simple, requiring a username and email address, and works across all Smithsonian transcription projects.

The transcription window has a zoomable image of the specimen, with a click through list of fields for the volunteer to transcribe. This window is similar for volunteers who are reviewing rather than transcribing. The transcription window has a link to social media, allowing volunteers to link to the specimen through Twitter or Facebook. The website automatically has the specimen on the right hand side, with the transcription fields on the left, but it does allow the volunteer to switch to a horizontal layout if they prefer.

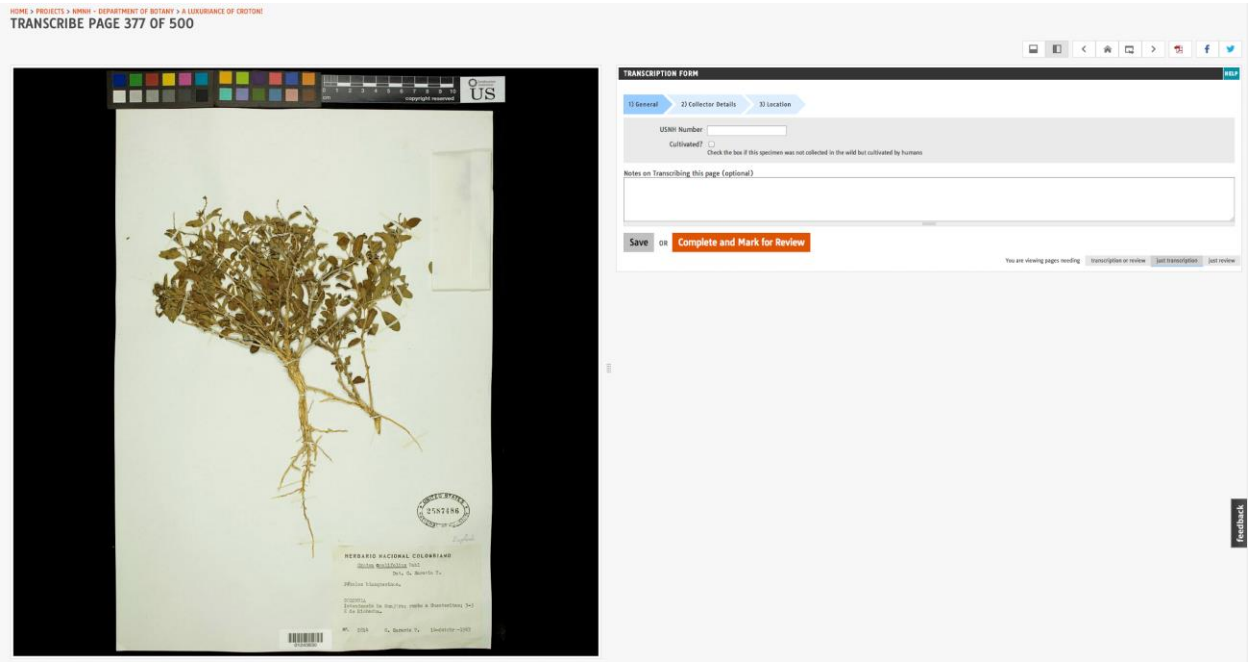

Figure 6 - Smithsonian Digital Volunteers: Transcription Centre

Each field has a help text which hovers on the side when the volunteer selects the field. They can also access a tutorial by selecting 'Help'. They are able to leave notes in a section called 'Notes on Transcribing this page (optional)' which would allow the volunteer to comment if there were difficulties in transcription, and the reviewer would then be able to see this.

The review and data validation process is more transparent on the Smithsonian website, with each specimen requiring review by other volunteers. The review process asks a volunteer to check the transcription of another, and then either mark the specimen as complete or reopen it for editing if changes are needed. The original transcriber can also edit the specimen.

The distinguishing features of the Smithsonian Digital Volunteers Transcription Center are:

- Review by other users is required, rather than an optional or hidden function
- Review and validation more public than other platforms
- Designed to allow easy import of data into collections management system (*http://goo.gl/JjMAQN*)

Project Case Studies

In the following section specific project case studies are described for some of the platforms.

Atlas of Living Australia: DigiVol

Host institute/collection: RBG Kew
Project lead: Sarah Phillips

Kew have run two DigiVol expeditions on the Atlas of Living Australia website: one for Angolan Lamiales, and the other on the *Drimiopsis* of Africa. Both tasks have been completed by volunteers on the platform, with the data subsequently validated by the Kew team.

The time to set up the expeditions on the website took longer than for herbaria@home but the team at Atlas of Living Australia were proactive in helping this set up. The platform has a wizard function to help with expedition creation with step-by-step guides. Images are uploaded directly to the site, allowing the administrator to create expeditions, add the background text and a photo for the expedition. A Kew specific PDF tutorial was been created for volunteers, which can be reused for other expeditions. This means that Kew have more control than at herbaria@home, but the overhead for creating expeditions is higher.

DigiVol allowed Kew to change the existing templates to suit their own needs - they were able to add their own fields and choose how the fields are mapped to DarwinCore. Records can be downloaded in DarwinCore-Archive format at any time, and so there is no reliance on the technical team to receive transcription data. There is however a period of learning needed to maximise the potential of the templates and administration tools. There are probably still potential improvements to be made to the templates and validation rules for any subsequent expedition as Kew learns about the flexibility of the system.

Atlas of Living Australia recommend that expedition sizes are relatively small, in the range of 200-2000 tasks to maintain volunteer motivation. This required Kew to split the tasks into smaller subtasks, which is slightly more work in comparison to herbaria@home which allows much bigger tasks (10,000+) which can constantly be added to. DigiVol strongly discourage adding specimens to an expedition once it is set up.

Validation was not always completed in the platform itself but undertaken after the data was downloaded from the system and prior to data import to the Herbarium collections management system, thus enabling bulk cleaning of data. There was a lack of staff resource to promptly validate all records within the system, although Kew validated a proportion of records from each volunteer to give prompt feedback on their transcription and to thank them for taking part in the expedition to encourage them to continue transcription. Once an institution has built up a volunteer base within DigiVol enthusiastic good quality transcribers can be given the rights to validate other transcribers' records reducing the management overhead of the institution. Another option could be to train an on-site volunteer to validate the records transcribed if institutional staff resource was an issue.

Use of the platform was free but ALA/DigiVol are considering a shared institutional development or subscription-based model in the future. Project administration is currently a potential issue as the user rights of administrators are based on trust - all administrators can modify global templates and download data from other projects.

There were various issues with the resulting data that made it harder to subsequently import it into Kew's collection management system. This included: data exported in amalgamated fields in inconsistent ways; unit stored in the same field as the value e.g. "900: 900 metres"; character encoding (problems with UTF-8). It may be possible to improve on these issues through improvements to validation rules for data entry via the template which needs to be further explored by Kew.

Angolan Lamiales was set up using 1,277 specimens images captured from a previous project where the label data had yet to be captured. Volunteers started to transcribe specimens on the 4th June 2014 and most specimens were completed by the 19th August 2014. In total eight volunteers contributed transcriptions however one volunteer contributed 87% of the data. Kew did very little publicity to attract volunteers, only sending out a couple of tweets while they tested the system. The main contributor was an existing volunteer just a few emails between Kew and the volunteer maintained her interest in the expedition. At the start of the project she admitted that she found the tasks a bit daunting but then started to enjoy it more. At the end of the project she stated that she enjoyed the work and felt like she had acquired knowledge of the geography of Angola and a few words of Portuguese. There were around 40 records which she found too hard to translate, these were mainly Baum collections with handwritten labels in German. These labels were challenging however she probably would have been able to complete them if she was pointed to the full text of the Kunene-Sambesi Expedition which can be found on the internet.

Angolan Lamiales project Link: http://volunteer.ala.org.au/project/index/3996284

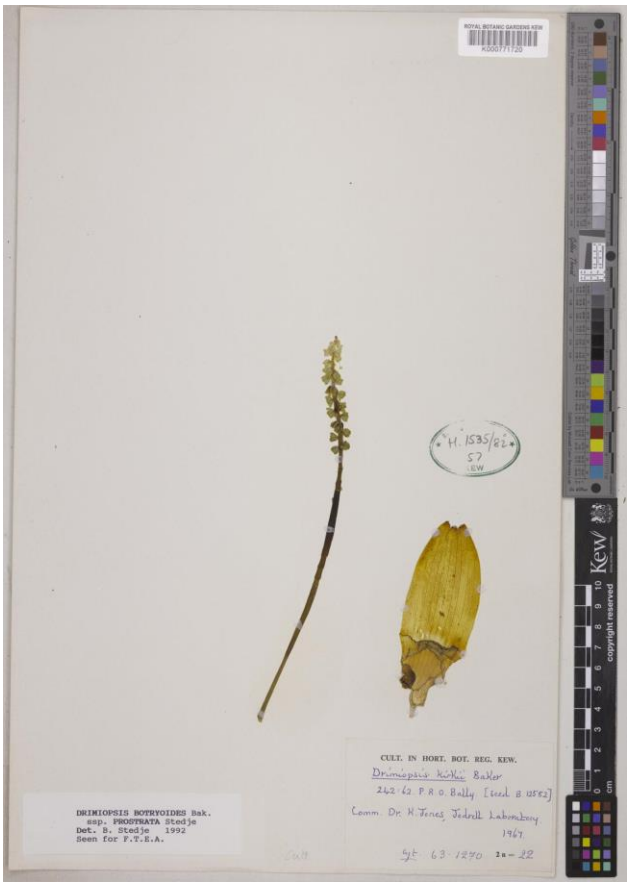

Figure 7 - Transcribed *Drimiopsis* herbarium sheet [available at http://apps.kew.org/herbcat/getImage.do?imageBarcode=K000771720]

Drimiopsis of Africa was a small test project conducted by RBG Kew based on a research data request from an external researcher. 153 specimens of *Drimiopsis* (Asparagaceae), an African plant genus, were imaged and put online for transcription.

Volunteers started to transcribe specimens on August 1st 2014 and all the specimens had been transcribed by mid-September (~6 weeks). In total seven volunteers contributed transcriptions with two making the majority (68%) of the contributions.

At the end of the project all the volunteers were emailed to thank them and to provide links to the online specimens with the transcribed data on Kew's online herbarium catalogue. The researcher who originally requested the data was also pleased with the outcome.

Drimiopsis of Africa project link: http://volunteer.ala.org.au/project/index/4725119

Drimiopsis of Africa dataset: http://collections.ala.org.au/public/show/dr1695

Herbaria@Home - RBG Kew

Host institute/collection: RBG Kew
Project lead: Sarah Phillips
Project duration: May 2012 - September 2014 - with gaps where there were no images.
Project link: http://herbariaunited.org/specimensearch/?search=search&Inst=17.5#searchlist

Kew have used h@h to transcribe over 13,000 British herbarium sheets of different plant families since 2012. All submitted sheets have been transcribed by volunteers on the platform with some of the sheets verified by volunteers.

Submitting herbarium sheets for transcription was quicker using h@h than using DV because the administrative overhead was minimal and unlike DV it is possible to keep adding to the number of tasks (herbarium sheets that need transcribing) once the host institute has digitised specimens. Most of the administrative work is undertaken by the project's lead developer (Tom Humphrey's) as there is currently no front end interface available to administrative users. This also means that the configuration of the project and customisation of fields by the Kew team was minimal making it harder to import data from h@h into Kew's collection management system.
When data is requested from Herbaria at Home for import it is delivered as an XML file. The structure of the XML means that each specimen record is broken up into several components, one for each element (i.e. comments, determinations, Provenance, Locality). For us to import it into our catalogue it is necessary to transform into a flat file into Excel and import into Access. In total data can be captured in c.90 fields allowing comprehensive label data capture. For a very small sample of records there were multiple entries for a single specimen and in these cases it was necessary to manually look through these records to decide which to import or if the data should be consolidated.

Use of the platform was free and is currently being maintained by the Botanical Society of British and Ireland. There is further potential for this site but it is currently limited by lack of resources to administer the portal, load images to the site and to maintain and update the IT Infrastructure. If these issues could be addressed there is great potential to increase the volunteer user base through the BSBI community, (Alex Lockton per comms. 2014). Currently transcription is aimed only at British specimens.

herbaria@home has a very active botanical community and Kew found that the georeferencing work, combined with grid referencing and validation led to good quality data being returned. There was strong volunteer knowledge on both UK collectors and localities. The interpretation of

handwriting was comparable to that of the digitisation team at Kew. Volunteers often gave added value to the records by adding very useful user notes to the record. It was not necessary for Kew staff to validate or give feedback to transcribers as experienced transcribers on the site undertook that role.

Notes from Nature - NHM

Project leads: Tim Conyers & Robert Prys Jones
Project link: http://www.notesfromnature.org/#/archives/ornithological

The Ornithological project was one of the three initial project proposals that was supported by Notes from Nature and went live in August 2013. The ornithological project is managed by the Natural History Museum, London and aims to transcribe ornithological registers from scans and import this data into the museum's collections management system.

Three Bird Group registers were used as a trial transcription project, with the three registers chosen being considered as easy to tackle. The registers were considered easy due to the legibility of the handwriting and the consistent layout. There were a total of 1037 register pages to be transcribed.

The interface development for the website took over 18 months. Initial plans were to record all details in the register, but during the course of the project it was decided to only record specific pieces of data:

Page Number
Registration Number
Scientific Name
Location
Collection Date

Daily activity from August 2013 to March 2014 is shown in the Figure 8.

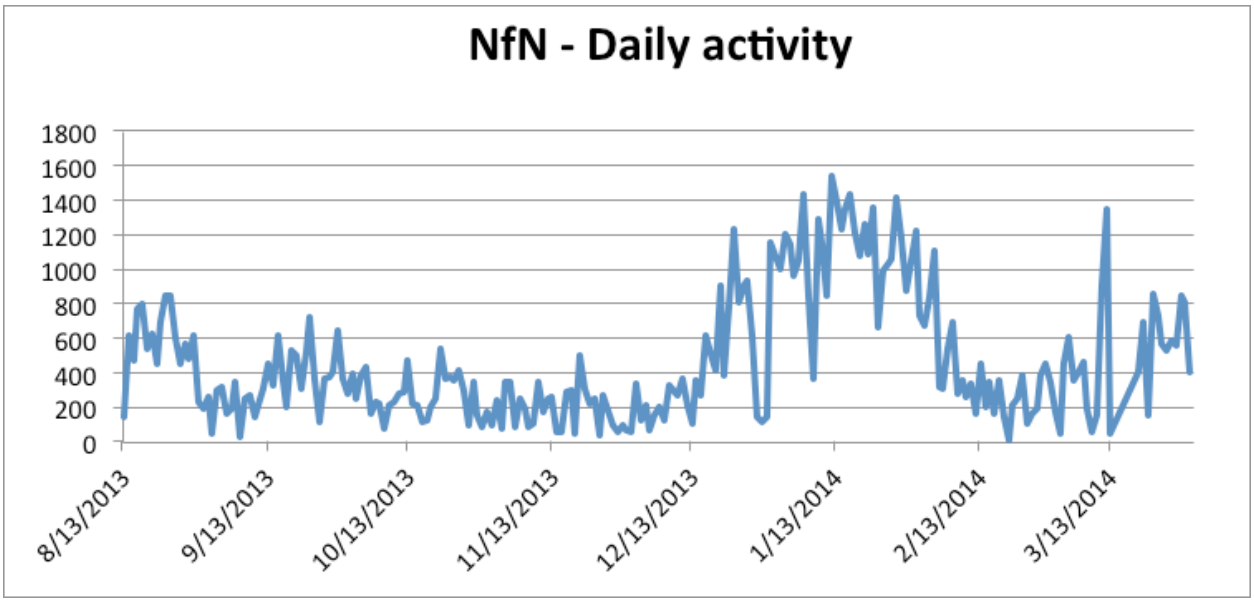

Fig 8: Daily Activity from Notes from Nature

By March 2014 there were over 100,000 transcriptions completed by over 1000 users. Often in crowdsourcing projects there are a few users who contribute toward the majority of transcriptions, and in the ornithological project one user completed over 30,000 of these transcriptions. Each transcription was completed by more than one user, with the aim being to have more confidence that the data was correctly transcribed.

There have been two issues with the data returned from this project. The first is the data validation process: deciding which of the multiple transcriptions should be used. The second is how to reconcile this data with the existing records in the collections management system. Some records were missed during the transcription process, with the team estimating 90% of records were transcribed. It is difficult to know how many records are missed in transcription projects using written material, as it requires a guess of how many records there are per page. In this project, it was only possible to measure the number of transcriptions completed due to a parallel digitising project.

The initial development of the Notes from Nature website was slow, partially caused by the decision making process. There are currently difficulties with developing the site further, despite a number of issues being documented.

The level of publicity for the project was not as high as the NHM team had hoped. The site was eventually linked on the NHM website, and an article appeared in the museum's magazine but neither were published until much later in the project.

The team have commented that there could have been more interaction between the scientists on the project and the users. NHM staff did contribute a few blogs on the website, but more could have been done. This would cost more in staff time and resource, but this type of interaction may be critical if hard core users were to lose interest.

Motivating the crowd: social aspects to crowdsourcing

The relevance of understanding volunteer motivations and expectations

Participation in a citizen science project is not guaranteed and there are many examples of projects that failed because the motivations and expectations of participating volunteers were not adequately considered (Roy et al., 2012). If the project does not appeal to potential users, then uptake can be poor. Similarly, participation can drop-off rapidly if the user interface is overcomplicated, unresponsive or does not provide the expected level of support and feedback (Tweddle et al., 2012; Pocock et al., 2014). Initial studies also show that there is a clear link between the level of engagement and motivation of participants, and both the longevity of participation and the quality and volume of data that are produced (see Roy et al., 2012).

It follows that understanding the interests, motivations and expectations of potential volunteers is critical if a project is to succeed in yielding both high quality scientific data and a rewarding volunteer experience.

Why do people take part in crowdsourcing projects?

Understanding why people choose to contribute their time and skills to crowdsourcing endeavours is an active field of research and there is still much to learn (Ellwood et al., in press). The following discussion summarises key findings to date.
Current thinking accepts that:

a. Motivations will vary between participants, both within and between sectors of society, and often within participants through time (Ellis & Waterton, 2005, Ellis, Waterton & Wynne, 2010, Grove-White *et al.*, 2007, Miller-Rushing *et al.*, 2012; Roy et al., 2012).
b. For mass participation projects it will never be possible to understand the motivations of all potential participants (e.g. Tweddle et al., 2012).

Hence, studies pertaining to motivation tend to focus on finding the areas of greatest agreement amongst volunteers.

Participation can be driven by both intrinsic and extrinsic motivational factors. Intrinsic factors relate to internal desires to perform a particular task. e.g. because it is enjoyable, intellectually

challenging, relates to an existing interest, feels morally worthwhile, or supports personal learning and skill development. Extrinsic factors are external to the individual and not directly related to the task being performed. The task is a means to an end, for example as a route to gaining financial or other reward (see discussion in Alam & Campbell, 2012).

The study of Rotman et al. (2012) identifies two critical decision points: the initial decision to participate and the decision to continue participating once the initial task has been completed. Motivations behind these two decisions can be quite different, as supported by studies of online (Raddick et al., 2013) and field-based citizen science (e.g. Roy et al., 2012).

(i) The initial decision to participate

An emerging consensus is that the initial decision to take part predominantly relates to intrinsic factors (Rotman et al., 2012; Alam & Campbell, 2012). Ridge (2014, pg. 294) argues that people "identify and support causes and projects that provide them with a sense of purpose, get meaning from doing things that matter to them and find a sense of belonging by being part of something bigger than themselves". The desire to 'record, find or discover new things' and contribute to delivery of an ambitious group objective can certainly be important motivations among crowdsourcing volunteers (Holley, 2010) and this has parallels within the amateur naturalist community (Grove-White et al., 2007). Common themes include the opportunities to learn and develop new skills, and to contribute to a socially-relevant project that the participant is confident will produce useful scientific research outcomes. Provision of a rich and compelling scientific narrative is central to success.

The example of the *Australian Newspapers Digitisation Programme* (Alam & Campbell, 2012) is pertinent in a transcription context. Here, contributors involved in text correction activities initially took part for intrinsic reasons, including a range of personal and community-based motivations (e.g. altruism, learning new knowledge, involvement in a collective endeavour). Similarly, participants in the *Transcribe Bentham* project (Causer & Terras, 2014; Ridge, 2014) reported that their initial decision to participate resulted from their interest in the subject area and crowdsourcing as a technique, and a sense of altruism (being part of something that would benefit the wider community). The study also captured some of the factors that limited participation, including difficulties deciphering Bentham's handwriting and aspects of the platform design.

Surveys of participants within Galaxy Zoo emphasize that most users expressed more than one motivating factor (Raddick et al. 2010), but that the opportunity to contribute to scientific research formed the primary reason for participation (Raddick et al., 2013). This is perhaps unsurprising given that involvement in science research is the primary marketing tool for Galaxy Zoo. The finding is though common to a range of other citizen science projects (e.g. OPAL, 2013). Other motivations for involvement within astronomical crowdsourcing projects included personal enjoyment, interest and enthusiasm for the topic of the project, learning new

knowledge, progression, and contribution to a cooperative endeavour (Nov et al., 2011; Gugliucci, 2014; Raddick et al., 2013).

An aesthetic appreciation of the material to be studied (Grove-White et al., 2007; Raddick et al., 2013) and social interaction (Tweddle et al., 2012) have also been cited as drivers for initial involvement in a range of citizen science projects.

Most studies to date have surveyed participating individuals, with fewer data available for school groups. Within field-based citizen science there is though strong evidence that the following factors help to inform the decision to participate: personal (i.e. teacher's) interest in the project; fit with curriculum; ease of involvement (ready-made, accessible materials and lesson plans); rapidity of feedback; location and cost (e.g. OPAL, 2013; Robinson et al., 2013). Anecdotal evidence indicates that teacher's greatly value participation in citizen science as a way to demonstrate the scientific process to their students (OPAL partnership, unpub.). A growing library of crowdsourcing project-based teaching resources is available within the Zooniverse platform (https://www.zooniverse.org/education), though to date these do not cover Notes for Nature.

(ii) Maintaining volunteer participation

Across the Zooniverse platform, a large proportion of classifications have been observed to originate during an initial spike of interest for the project concerned (Lintott, C. no date). Similarly, Nov et al. (2011) report that many digital citizen science projects can suffer from high rates of user drop-off through time and that individual contributors often reduce their input after an initial period during which they explore the project. These observations are supported by the OPAL field-based citizen science portfolio (e.g. Bugs Count project, unpublished data), for which clear spikes in data submission accompanied project launch activity, with lesser spikes following subsequent media events.

An implication of these findings is that repeated publicity activity may be necessary to encourage a flow of new users to the project website, particularly within platforms that do not have an established community of participants. Similarly, it is sensible to invest resources in the development of mechanisms that reward and encourage longer-term involvement of participating volunteers. Although data are currently sparse, this is another area of active research.

Potential forms of reward include rapid feedback, the opportunity to progress and develop skills and knowledge, social interaction (including via user-focused forums), gamification, leaderboards and other incentives such as digital badges.

There is growing evidence that thanking participants for their efforts and demonstrating how their contributions are helping to progress the project's scientific goals can be a powerful motivator. Prompt and appropriate feedback can form one of the clearest rewards for participation and play a key role in preventing participant drop-off (Roy et al., 2012; Tweddle et al., 2012). Within web-based crowdsourcing projects, feedback typically occurs through task completion counts, project blogs and forums.

Task completion counts can be an effective method to show the participants' individual contribution, as well as overall project progress (Ellwood et al., in press), an approach that has been employed very successfully across the Zooniverse platform (Zooniverse, 2014). The use of blogs and social media to communicate project progress and activities is both instant and has the added benefit of advertising the project to potential new participants. Online project-focused forums in which knowledge can be shared and questions asked, and results interpreted and discussed by project scientists and participants have proven to be a highly effective way to deepen feelings of 'ownership', provide opportunities for social interaction and maintain motivation amongst participants (Hars and Ou, 2002; Simpson, 2014). Forums can also help to develop a sense of community within the project (Iocovides et al., 2011) and build skills (see Ellwood et al., in press). In the case of Galaxy Zoo, forum-based discussions have also led to the generation of unexpected scientific findings (Chris Lintott, pers. comm.).

Within field-based citizen science, the most successful projects in terms of participant retention are often those that provide volunteers with the opportunity to develop their skills and knowledge and progress from straightforward tasks to more complex tasks and responsibilities (Roy et al., 2012). There are various ways in which this can be enabled within a crowdsourcing project including a tiered engagement model, whereby a range of tasks that vary in duration and complexity/required experience are provided.

Leaderboards and gamification (i.e. development of game-like environments) can help to maintain the involvement of some participants, although currently available evidence is mixed (Iocovides et al., 2013; Lintott, 2012; Bowser et al., 2013; Prestopnik and Crowston, 2011). Studies to date do not cover specimen transcription-style projects and further research is required to assess whether there are trade-offs with the adoption of a gaming approach, such as the quality of resultant data. Relevant studies that consider the motivational impacts of digital badges as a form of reward are also currently lacking.

Volunteer behaviour

Studies show that volunteers do not behave as a single population. Some complete an activity once and then leave the project, whilst others participate repeatedly over longer time-frames (e.g. Raddick et al., 2013). A smaller subset of volunteers engage more deeply, for example by

contributing to discussion forums, or helping to interpret and publish results (Cardamone et al., 2009; Józsa et al, 2009).

A key finding across projects is that a small number of contributors are often responsible for generating a disproportionately high volume of results (Simpson, 2014; SubseaObservers, 2014). Others donate a considerable amount of their time to support discussions on project forums or message boards and mentor new volunteers (e.g. Herbaria@home, Tom Humphrey, pers. Comm.). This suggests that supporting the activity of such 'super-contributors' can be highly beneficial to both the scientific output of the project and to the citizen scientists themselves. Conversely, the development of mechanisms to increase the involvement of the general - less frequently contributing - crowd may also be worth investigating (Franzoni and Sauermann, 2014).

What does this mean for project developers?

The above findings highlight that projects should aim to be personally and socially relevant to the intended audience(s) and incorporate appropriate types of reward if they are to attract and, where desirable, maintain participation. As the motivations and interests of participants can be quite different to those of the project developer, effective communication and dialogue can often be central to success.

This is particularly pertinent to crowdsourcing projects, which have to date largely followed a top-down (contributory cf. Bonney et al., 2012) model of development, in which the project goals, dataset and analytical approaches (i.e. tasks) have been set in advance by the project lead, rather than in collaboration with potential participants.

Examples from the cultural heritage (Ridge, 2014; Blaser, 2014) and field-based citizen science sectors (e.g. Roy et al., 2012) show that a more collaborative approach, which factors in suggestions from participants can both aid the development of project protocols and interfaces and refine the research questions themselves. At its simplest this could take the form of user-testing, whether as pilot testing early within the project design process, or through a (cyclical) process of user-testing and evaluation pre- and post-launch (see Tweddle et al., 2012 and Pocock et al., 2014 for detailed discussion). At a more involved level, the project could be developed as a true collaboration within which representatives of the participant community form part of the core project team and help to shape project design, implementation and analysis. It is evident that depth of involvement and feeling of control over the scientific process are powerful motivators and can lead to deep, long-lived engagement and high quality of output alike (Roy et al., 2012). If embraced, such two-way communication can subsequently have a positive influence on the realised scientific and engagement outcomes of a crowdsourcing project.

This observation links to an evolving line of thought which expands the value that crowdsourcing of object digitisation and related research activities can bring to museums. Rather than focusing purely on the scientific and collections access benefits, it frames crowdsourcing as an opportunity for deepened visitor engagement. Through the act of participating, people have the opportunity to engage with specimen collections far more deeply than is possible via other routes; the act of participation means that rather than simply receiving information, people are creating it (Owens, 2012). Owens writes that “Far better than being an instrument for generating data that we can [then] use to get our collections more used, it is actually the single greatest advancement in getting people using and interacting with our collections”. Although currently driven by the cultural heritage museums and galleries sector, this concept is equally applicable to natural history museums and may be a fruitful area for future research.

Discussion and Recommendations

Establishing a crowdsourcing programme has the potential to bring a natural history collection a number of benefits which include:

- An effective way of engaging members of the public
- Increased rates of collections digitisation
- Greater scientific outputs
- A stronger online presence

However, a crowdsourcing programme requires organisational investment in a number of areas:

- Technical infrastructure (e.g. software, hardware and developers)
- Communication, outreach and support (e.g. dedicated staff time to develop and provide feedback to an external community, internal project manager and scientists)
- Effective project selection (e.g. strong narrative, potential scientific outputs, public appeal, well-structured tasks of known complexity)
- Preparing underlying data (e.g. data for autocomplete fields such as collector names or localities)
- Post-processing and subsequent import into institutional collections management system

Based on the feedback gathered from crowdsourcing case studies, experience from running citizen science projects and guidance issued by some of the platforms we have summarised some general recommendations for creating effective crowdsourcing projects:

- Each project should have a clear design rationale with cultural or scientific benefits;
- Projects should be actively promoted (PR strategy/comms plan);
- Institutions should make it clear how and when the data will be publically available through an open access policy;

- Projects should be broken up into discrete parts - volunteers enjoy completing projects rather than having endless tasks (pers. comms. Deb Paul; most platforms have 1000-5000 specimens per project with the exception of NfN);
- Platforms should utilise existing data standards where possible (TDWG/ISO);
- Ensure there is adequate help and documentation on both platform and project level;
- Always display specimens along with labels where possible (volunteers enjoy seeing specimens);
- Record and analyse project activity (e.g. using Google Analytics);
- Allow volunteers to work on projects and tasks of interest (e.g. specific location or taxa);
- Provide post-project updates (e.g. resulting papers, data sets, public outreach etc);
- Allow volunteer communities to establish and support themselves (e.g. the wiki/forums model of h@h and object-orientated forums within Zooniverse);
- If appropriate and feasible, consider differentiating between users of different skill/experience levels (e.g. reviewing contributions or harder tasks);

Future Research

This report does not comprehensively address and answer questions about crowdsourcing but aims to act as a starting point for institutes who are considering creating a crowdsourcing programme.

We recommend that additional research is undertaken to address the following questions and data deficiencies:

- Detailed technical information about existing crowdsourcing platforms and their underlying technology
- A/B testing of platforms and their unique approaches
- Potential of OCR to make label transcription more efficient e.g. do some sorting and organising prior to a crowdsourced project
- Optimal size of a project, e.g. the number of tasks - There is anecdotal evidence that people enjoy regular smaller projects, even if they are parts of the same larger project
- Further understanding of crowd motivation in natural history crowdsourcing
- Importance of multilingual interface support and multilingual transcribers (our collections have data in multiple languages and alphabets, some locality information is very obscure)
- Best way to evaluate volunteer experience
- What opportunities are there to crowdsource physically on site as a visitor activity?
- What are the best method for motivating the crowd?
- Sustainability of crowdsourcing platforms

- Can we come closer to matching the popularity of crowdsourcing projects like Galaxy Zoo?
- How can we support/move the crowd from simply digitising specimen information for future research purposes, to performing research activities themselves?

Recommended Reading

Evaluation - Phillips, T. B., M. Ferguson, M. Minarchek, N. Porticella, and R. Bonney. 2014. User's Guide for Evaluating Learning Outcomes in Citizen Science. Available online: http://www.birds.cornell.edu/citscitoolkit/evaluation

Evaluation - RCUK. 2011. Evaluation: Practical Guidelines. Research Councils UK. 49 pp. Available online: http://www.rcuk.ac.uk/Publications/policy/Evaluation/

Crowdsourcing resources for teachers and educators - https://www.zooniverse.org/education

Hill A, Guralnick R, Smith A, Sallans A, Gillespie R, Denslow M, Gross J, Murrell Z, Conyers T, Oboyski P, Ball J, Thomer A, Prys-Jones R, de la Torre J, Kociolek P, Fortson L (2012) The notes from nature tool for unlocking biodiversity records from museum records through citizen science. ZooKeys 209: 219-233. doi: 10.3897/zookeys.209.3472

Acknowledgements

We would like to thank the following people for providing data and feedback on this report: Simon Chagnoux, Tim Conyers, Libby Ellwood, Paul Flemons, Tom Humphrey and Deborah Paul.

References

Alam, S.L. and Alam, J. (2012). Crowdsourcing Motivations in a not-for-profit GLAM context: The Australian Newspapers Digitisation Program. 23rd Australasian Conference on Information Systems 3-5 Dec 2012, Geelong. Available: https://dro.deakin.edu.au/eserv/DU:30049107/alam-crowdsourcingmotivations-2012.pdf

Ariño, A. (2010) Approaches to Estimating the Universe of Natural History Collections Data. Biodiversity Informatics 7: 81-92. https://journals.ku.edu/index.php/jbi/article/view/3991/3805

Blaser, L. (2014). Old Weather: Approaching Collections from a Different Angle. In: *Crowdsourcing Our Cultural Heritage*, ed. M. Ridge (Ashgate).

Bonney, R., Ballard, H., Jordan, R., Mccallie, E., Phillips, T., Shirk, J. and C. Wilderman, C. (2009) Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report. Center for Advancement of Informal Science Education (CAISE), Washington, D.C.

Bowser, D. Hansen, J. and J. Preece (2013). Gamifying Citizen Science: Lessons and Future Directions. *Gamification Research Network, CHI 2013*. Available: http://gamification-research.org/wp-content/uploads/2013/03/Bowser_Hansen_Preece.pdf

Cardamone, C., Schawinski, K., Sarzi, M., Bamford, S. P., Bennert, N., Urry, C. M., Lintott, C., Keel, W. C., Parejko, J., Nichol, B., Thomas, D., Andreescu, D., Murray, Phil, Raddick, M. J., Slosar, A., Szalay, A. and J. Vandenberg. (2009) Galaxy Zoo Green Peas: discovery of a class of compact extremely star-forming galaxies. *Monthly Notices of the Royal Astronomical Society*, 399 (3). pp. 1191-1205. ISSN 0035-8711 10.1111/j.1365-2966.2009.15383.x

Causer, T., and Terras, M. (2014) Many hands make light work. Many hands together make merry work: *Transcribe Bentham* and crowdsourcing manuscript collections. In: *Crowdsourcing Our Cultural Heritage*, ed. M. Ridge (Ashgate).

Ellis, R. & C. Waterton. (2005) Caught between the cartographic and the ethnographic imagination: the whereabouts of amateurs, professionals, and nature in knowing biodiversity. *Environment and Planning D-Society & Space* 23, 673-693.

Ellis, R., Waterton, C. & B. Wynne. (2010) Taxonomy, biodiversity and their publics in twenty-first-century DNA barcoding. *Public Understanding of Science* 19, 497-512.

Ellwood, E.R., B. Dunckel, P. Flemons, R. Guralnick, G. Nelson, G. Newman, S. Newman, D. Paul, G. Riccardi, N. Rios, K. C. Seltmann and A. R. Mast. (in press). Accelerating digitization of biodiversity research specimens through online public participation. *BioScience*.

Estelles-Arolas, E., and F. Gonzalez-Ladron-De-Guevara. (2012) Towards an Integrated Crowdsourcing Definition. *Journal of Information Science* 38.2 (2012): 189-200

Franzoni, C. and H. Sauermann. (2014) Crowd science: The organization of scientific research in open collaborative projects. *Research Policy,* 43, 1– 20.

Galaxy Zoo (2014) *Galaxy Zoo*. Available from: http://www.galaxyzoo.org/ [Access 27th August 2014]

Grove-White, R., Waterton, C., Ellis, R., Vogel, J., Stevens, G. & B. Peacock, B. (2007) *Amateurs as experts: harnessing new networks for biodiversity.* Lancaster University, Lancaster.

Gugliucci, N., P. Gay. and G. Bracey. 2014. Citizen Science Motivations as Discovered with CosmoQuest. Ensuring STEM Literacy: a National Conference on STEM Education and Public Outreach ASP Conference Series, Vol. 483. Astronomical Society of the Pacific.

Holley, R. **(**2010) Crowdsourcing: How and Why Should Libraries Do It? *D-Lib Magazine* March/April 2010, Vol 16, Number 3/4.

Iacovides, I., Jennett, C., Cornish-Trestrail, C., and A.L. Cox. (2013) Do games attract or sustain engagement in Citizen Science? A study of volunteer motivations. *Proceedings of the 2013 Conference on Human Factors in Computing Systems (CHI-2013), Extended Abstracts,* 1101–1106. New York: ACM.

Józsa, G. I. G.; Garrett, M. A.; Oosterloo, T. A.; Rampadarath, H.; Paragi, Z.; van Arkel, H.; Lintott, C.; Keel, W. C.; Schawinski, K.; Edmondson, E. (2009). Revealing Hanny's Voorwerp: radio observations of IC 2497. Astronomy and Astrophysics 500 (2): 33-36. DOI: 10.1051/0004-6361/200912402

Lintott, C. (no date). The two-and-a-bit page guide to running a Zooniverse project. Available online: http://static.zooniverse.org/www.citizensciencealliance.org/downloads/zooniverse_guide.pdf

Lintott, Chris J.; Schawinski, Kevin; Slosar, Anže; Land, Kate; Bamford, Steven; Thomas, Daniel; Raddick, M. Jordan; Nichol, Robert C.; Szalay, Alex; Andreescu, Dan; Murray, Phil; Vandenberg, Jan. (2008) Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society* 389 (3): 1179-1189. DOI: 10.1111/j.1365-2966.2008.13689.x

MicroPasts (2014) *MicroPasts*. Available from: http://micropasts.org/ [Accessed 27th August 2014]

Miller-Rushing, A., Primack, R. & R. Bonney. (2012) The history of public participation in ecological research. *Frontiers in Ecology and the Environment* 10, 285-290.

Nov, O., Arazy, O. & Anderson, A. (2011) Dusting for science: motivation and participation of digital citizen science volunteers. *iConference*. Seattle, USA.

OPAL Community Environment Report: exploring nature together (2013). ISBN: 978-0-9574377-0-8. http://www.opalexplorenature.org/CEreport

Owens, T (2011) Crowdsourcing the Civil War: Insights interview with Nicole Saylor. The Signal: The Library of Congress Digital Preservation Blog, Dec. 6. Available at http://blogs.loc.gov/digitalpreservation/2011/12/crowdsourcing-the-civil-war-insights-interview-with-nicole-saylor/. (Accessed: 13 October 2014)

Owens, T. (2012). Crowdsourcing cultural heritage: the objectives are upside down. Blog, March 10. Available at: http://www.trevorowens.org/2012/03/crowdsourcing-cultural-heritage-the-objectives-are-upside-down/ [Accessed 04/01/15]

Pocock, M. J. O., D. S. Chapman, L. J. Sheppard and H. E. Roy. 2014. *Choosing and using citizen science: a guide to when and how to use citizen science to monitor biodiversity and the environment.* Centre for Ecology and Hydrology, UK. 28 pp.

Prestopnik, N.R. and K. Crowston. (2011). Gaming for (citizen) science: exploring motivation and data quality in the context of crowdsourced science through the design and evaluation of a social-computational system. *IEEE Seventh International Conference on e-Science Workshops, 2011*, 28-33.

Raddick, M. J., G. Bracey, P. L. Gay, C. J. Lintott, C. Cardamone, P. Murray & J. Vandenberg. (2013) Galaxy Zoo: Motivations of citizen scientists. *Astronomy Education Review, 12, 1.*

Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Murray, P., Schawinski, K., Szalay, A. S., & J. Vandenberg. (2010) Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers. *Astronomy Education Review*, 010103.

Ridge, M. ed. (2014). Crowdsourcing our cultural heritage (Ashgate)

Robinson, L.D., Tweddle, J.C., Postles, M.C., West, S.E., & J. Sewell (2013) *Guide to running a BioBlitz.* Natural History Museum, Bristol Natural History Consortium, University of York and Marine Biological Association. http://www.bnhc.org.uk/wp-content/uploads/2014/04/BioBlitz-Guide-2013.pdf

Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C., Lewis, D. and D. Jacobs. Dynamic changes in motivation in collaborative citizen-science projects. *Proc. CSCW 2012*, ACM Press, (2012), 217-226. http://dl.acm.org/citation.cfm?id=2145238

The Royal Society (2012) Science as an open enterprise: The Royal Society Science Policy Centre report. Available at: https://royalsociety.org/~/media/policy/projects/sape/2012-06-20-saoe.pdf (Accessed: 8 October 2014).

Simpson, R. (2014) Citizen science with the Zooniverse. Citizen Cyberscience Summit 2014, London.

SubSeaObservers. (2014) Available from: http://subseaobservers.com/ [Accessed 5th January 2015]

Tweddle, J.C., Robinson, L.D. Pocock, M.J. & H.E. Roy (2012). *Guide to citizen science: developing, implementing and evaluating citizen science to study biodiversity and the environment in the UK.* Natural History Museum and NERC Centre for Ecology and Hydrology for UK-EOF. ISBN: 978-1-906698-37-9. http://www.ukeof.org.uk/resources/citizen-science-resources/

Westphal, A. J.; Butterworth, A. L.; Snead, C. J.; Craig, N.; Anderson, D.; Jones, S. M.; Brownlee, D. E.; Farnsworth, R.; Zolensky, M. E. (2005) Stardust@home: A Massively Distributed Public Search for Interstellar Dust in the Stardust Interstellar Dust Collector

Zooniverse. (2014) *Zooniverse - Science Projects*. Available from: https://www.zooniverse.org/projects [Accessed 21st August 2014]