Crowd Lab-ing Brain Research

Wendi He

Master of Science

School of Informatics

University of Edinburgh

2018

Abstract

This dissertation describes the methodology taken for the proof of a concept citizen science platform aimed at investigating a solution for crowdsourcing data collection. A web-based citizen science platform is developed by using Django web framework has successfully addressed this problem. The main tasks for development including: Protocol design, Web framework construction, database design, UI design. The result from data collection demonstrate that this project it is an accessible solution for crowd sourcing gathering with a guaranteed data quality.

Acknowledgements

Many thanks to my supervisor, Prof. J Douglas Armstrong and Dr. Valentina Ferlito who provide constant support to me. I am really grateful that I am the one of the School of Informatics.

And I love my grandpa forever.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Wendi He*)

Table of Contents

1 Introduction **1**

1.1 General Motivation 1

1.2 Citizen Science 1

1.3 Objective 3

1.4 Contribution 3

1.5 Structure of the thesis 4

2 Background and Related work **5**

2.1 Parkinsonism and Targeted Population 5

2.2 Citizen Science Project 6

2.3 Crowdsourcing 7

2.4 Related work on Crowdsourcing Citizen Science Platform 8

3 Methodology and Design **10**

3.1 Original User Requirement Analysis and Prototype design 10

3.1.1 Volunteer Engagement 10

3.1.2 Requirement strategies 11

3.1.3 Principle of citizen science 13

3.2 D.A.Lab Technology Architecture 14

3.2.1 Web framework 14

3.2.2 Database Design 16

3.3 UI Design-Related Approaches 19

3.3.1 Color design in UI 20

3.3.2 Animation added in interface 20

3.3.3 Page Display and functional description 21

3.4 Data collection 25

3.4.1 Manual-input validation 26

3.4.2 Survey Design 27

4 Analysis and Evaluation **28**

4.1 Self-Evaluation 28

4.2 Usability Evaluation 29

4.2.1 Basic Profile of Participants 29

4.2.2 Exposure to Parkinsons Disease 30

4.2.3 Task Completion Rates 30

4.2.4 User experience analysis 31

4.2.5 User Engagement Analysis 34

4.2.6 Overall score of D.A.Lab 35

4.3 Data Quality Evaluation 35

4.3.1 Data Integrity 35

4.3.2 Data Consistency 36

5 Limitations and Further Work **39**

5.1 User Stickiness and Motivation 40

5.2 Cognitive Friction 40

5.3 Reward Mechanism 41

5.4 Web Traffic Data 42

5.5 Community-based Character 42

6 Conclusion **44**

A Main code for Data Validaion **45**

A.1 HTML form submission attribute 45

A.2 Javascript API 45

B Survey Design **47**

Bibliography **49**

Chapter 1

Introduction

In this chapter, Im going to introduce the general motivation, objective, contribution of the project. The introduction is partly taken from my IPP.

1.1 General Motivation

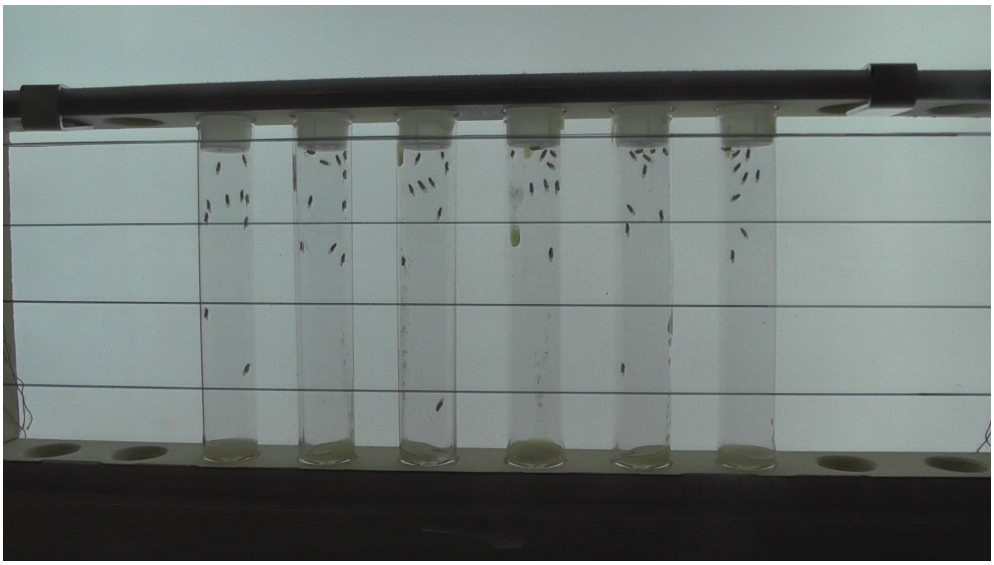
Parkinson's disease is a long-term degenerative disorder of the central nervous system which mainly affects the motor system [35]. This disease usually caused rigity, tremor and bradykinesia bradykinesia in the early stage. Besides, more problems would occur with their thought and behaviors[Sveinbjornsdottir]. Other symptoms include sensory, sleep, and emotional problems. Thus, Parkinson's patients are slower in action and thinking than unaffected people in terms of cognitive and physical abilities.

Nowadays, most of researches was established based on Parkinson's disease. The scientists form the University of Edinburgh are searching for the new treatment of Parkinson's disease from the regular drug process on Drosophila. The assessment of neurodegenerative phenotypes in Drosophila by climbing assays requires the significant annotation of a large amount of data sources. Thus, the purpose of a project is to design a platform that allows distance users to score simple behaviors of Drosophila in some sample images was created. For this motivation, the definition of 'citizen science' is introduced.

1.2 Citizen Science

Citizen science, as a new civil collaborative scientific fashion has been developed in recent years for encouraging professional or non-professional people getting involved

Figure 1.1: One of the sample images for volunteers to do annotation tasks



into scientific research for universities academic proposal, public governmental organization or commercial company, typically through the processes of gathering, transforming, analyzing and curating metadata [46]. It has recruited large amounts of personnels for basic works and provided more rational debates and solutions for scientific research which undoubtedly provides material fundamental by human interacting. Since 1982, it was the first time for Stebbins gathered volunteers, amateurs, defined this public involvement as a method for collecting life interests from different hobbyists by some entertained relaxation[45].

In recent years, public involvement has been applied in scientific discovery and its development with is confirmed based on data analysis of related article on web of Science and Scopus[19]. It means, the increasing reliance on citizen science makes it possible to address the scientist problems and also provide a platform for vigorous volunteers who do not need any specialized expertise or background to engage in projects.

With the development of technology and social network, participation in online activities have made marvelous progress. The existing projects indicate the value of collecting data from volunteers and amateurs, the collaborative potential become a heated topic as which could inspire developer to build a better infrastructure or platform to understand the motivation of participators and then promote a cooperative partnership. In last few years, as the research sometimes has the requirement of excursion geographically and asynchronously, the budget and practical problem need to be considered. Scientist nowadays are searching for more efficient and appropriate solutions on col-

lection data. The traditional presentation of citizen science,like emails,questionnaire survey,instant messaging, text messaging [4]has no longer to satisfied users' need.

Thus, the web-based platform play a main-streamed role in citizen science project. As computational systems, they are human-agent collectives that leverage human computation and crowd-sourcing to help professional organization process huge amounts of research data to assist related work. For fostering a appropriate and sustainable community, human-computer interaction and pervasive computing concept are necessary to take into consideration. As volunteers participate in projects spontaneously, they have right to know motivations and profits. Thus the platform is obliged to encourage users to participate, enhance their social responsibility and enthusiasm for science which need to improve the usability,interactivity,incentive and intrinsic business logic.

1.3 Objective

Drosophila.Annotation.Lab is a project concept initiated to provide patients afflicted with Parkinsons disease a direct voice in scientific research, with the ultimate goal of finding a cure. In this context, persons affected by Parkinsons disease as well as concerned individuals motivated to discover treatments for the disease, become active participants at all stages of the research and therapeutic development process.

The project provides some simple annotation tasks for volunteers without any geographical hindrances. Users who join this community receive adequate and usable tutorial resources for the preparation of annotation tasks. At the same time, they would gain the sense of achievement when they make certain contributions to the scientific research.

1.4 Contribution

D.A.Lab collaboration assist scientists and researchers in the collection and analysis of data from cluster of images gathered from Drosophila climbing experiments for drug discovery and screening tests for Parkinsons Disease. Volunteers are able to participate and perform annotation tasks on image files. This platform aspires to provide an effective communication channel available to volunteers for information contribution and to project researchers for broad data source collection. Some specific characters such as interaction and information flow, determine the performance and value of the platform function.

Since the implementation of the D.A.Lab, we have successfully received accessible data from volunteers for assisting neuronscience scientists to do further research on the treatment of Parkinson's disease. Furthermore, the related assessments and discussions based on existing results will be used for continuous improvement of the platform.

1.5 Structure of the thesis

Chapter 2 will give a detailed description of the development of citizen science project combined with its accessibility on Parkinson's disease. Chapter3 will elaborate on the methodology applied to the web development. Chapter 4 covers the evaluation of our results and chapter 5 describe the limitation at the current stage and give direction of further work. Chapter 6 existed for a short summary of the whole project.

Chapter 2

Background and Related work

In this chapter, I will introduce some cutting-edge citizen science projects with considerable scales and growing user bases as well as the application of crowd data resources. The method and prepared work of my project would also briefly included in this chapter with a more reasoned introduce about the definition of Parkinson's Disease which aimed at increasing designing usability and user-friendliness. This is partly taken from my IPP.

2.1 Parkinsonism and Targeted Population

In Douglas' paper, Parkinson's Disease is defined as:

> **The clinical diagnosis of Parkinson disease (PD) is based on the identification of some combination of the cardinal motor signs of bradykinesia, rigidity, tremor, and postural instability [16]**.

As several clinic signals have been mentioned in Donley's paper, the cardinal features of Parkinson patients was embodied in degenerative motor disorder. Making patients alleviate suffering from Parkinsonism, complementary therapies(CTs) was introduced to improve and recover some basic movement and thinking abilities. As for distanced patients, scientists have experienced on mailing surveys to them which contained questions on CTs used, effectiveness, and for what symptoms were the CT helpful. The result shows that complementary therapies including some restorative actions are both helpful for monitor and non-monitor symptoms [17, 33]. It means that we could find a way for Parkinson's patients to push the drug discovery by doing some simple operation, such as providing artificial annotation for Drosophila climbing test, as well as

relieving symptoms. This helps them achieve self-value with increasing social responsibility and make more connections with scientists and information about Parkinson's disease. By this way, considering the limitation of patients' movement, the platform could gaurantee that they have rights to choose an environment in such a convenient way to get involved in these projects in a web-based foundation. Moreover, the pages should be user-friendly to operate.

2.2 Citizen Science Project

In recent years, a large-scale study pattern requires a considerable amount of data to be collected across time and space. The web-based data-gathering technique was born to obtain crowdsourcing data which motivates citizens to integrate scientific data source [60]. Citizen science, scientific research could be conducted by non-specialists, has found ways for the amateur enthusiast to contribute their strength to further studies. It has recruited large amounts of personnel for essential works and provided more rational debates and solutions for scientific research which undoubtedly provides material fundamental by digitized human interacting. Participants who get involved in a citizen science project could also be called 'citizen scientists' [43]. Their contributions to scientific research are mostly based on simple operations such as data classification, annotation, and analysis. For ensuring the data quality, participants are always given enough tutorial information on citizen science platform. The evaluation and monitoring mechanism, as well as knowledge dissemination, are motivating social learning and responsibility. Therefore, the available results of data-gathering always perform as an indicator of some trend, change or regulation which could be quoted in scientific publication to inform population management decisions [12, 48].

Citizens have been involved in public scientific research for a long time. Many excellent citizen science projects have emerged from previous works. In history record, the first event could be tracked in China related to locust outbreaks [51]; the history of the first Christmas Bird Count was held by the National Audubon Society in the turn of the 20th century [47]; the trend of birds distributions and behaviors has been revealed by data from spontaneous birds-watchers in North America [18]; in the Philippines, hunters and anglers in protected areas monitor and react to changes in usage of resource linked with populations of wildlife, etc. [15]; to find impact brought about by gas extraction, Pennsylvanian residents participate the activity of monitoring water turbidity and conductivity as well as macroinvertebrate populations etc. [58].

As the research sometimes has the requirement of excursion geographically and asynchronously, the budget and practical problem need to be considered. Scientist nowadays is searching for more efficient and appropriate solutions for collecting data. The increasing reliance on citizen science makes it possible to address the scientist problems and also provide a platform for vigorous volunteers who do not need any specialized expertise or background to engage in projects.

Therefore, designing and building a credible, feasible and interesting platform for supporting citizen science is directly determine the quality of a part of research processing and data collection from the volunteer.

2.3 Crowdsourcing

'Crowdsourcing' was first coined by Howe in 2006. It is a sourcing model in which individuals or organization could obtain required services and ideas with advantages of its speed, cost, flexibility and quality[38]. Silvetown[43] reported that the emergence of the Internet has provided users with increased labor resources from crowdsourcing as well as free intelligence and skills. The process of crowdsourcing work was usually divided into discrete, repeatable micro-work [59] which always be accomplished for a larger overall project. Crowd-sourcing techniques have proven to be an effective method to provide valuable contribution by thousands of participants in the scientific research process. The human combination of intuition and perception with ability in pattern recognition and analysis could, for a wide variety of scientific tasks, far outstrip the performance of even sophisticated automatic systems [44]. For providing a solid scientific foundation and supported development for crowdsourcing process, the diversity and adequacy of the models need to be defined appropriately. The features of participants must be taken into consideration as the data they provide depends on their knowledge background and skills. Moreover, the interaction with crowd members; the inherent uncertainty and disagreements in crowd answers; and evaluation metrics are used to capture the cost-effective of the crowd [5, 6].

In recent years, the online crowdsourcing method has gained attention from educators, researchers, scientists, economist, and policymakers because of the ongoing advances in Internet and computing technologies [57] which is popularized in different citizen science projects.

2.4 Related work on Crowdsourcing Citizen Science Platform

Related to my project, I suppose that there are two factors are crucial to citizen science initiatives:

- **Web-based community:** Social elements and communication activity are two perspectives for online interaction among users, scientists, and developers which push for project process.
- **Crowdsourcing:** Large-scale projects attract volunteers and amateurs to provide sourced information

There is a hobby or 'avocational bias' [53] of citizen science result in different degree of investigation of citizen science projects. Web-based platforms enlisted crowd help in these way:

1. Data collection [25, 36, 56]/Data classification [37]; /Data analysis [27, 52]
2. Collective intelligence [52]

In this way, Zooniverse, as the world's largest citizen science platform, is hosting thousands of projects which consisted of more than 1.6 million registered participants contribute to research projects led by hundreds of researchers. Raw resources related to the research are presented on the web page in the form of pictures, video, audio, etc. for the huge amount of registered users acquiring online or downloading. Furthermore, simple step-by-step instruction will show by each project to volunteers how to annotate, classify, identify as real researchers do.[44]

Compared with my project itself, there is some similar platform released on Zooniverse which are aimed at encouraging volunteers to do data annotation from given resources. Galaxy Zoo, the first Zooniverse project, was launched July 2007 and motivate over 1 million citizens to participate in the classification of galaxies pictures[30]. Annotating astronomical photographs in this project is one of the main operations for individuals to work as a galaxy investigator. In its new edition(Galaxy Zoo Quench), users could participate in a complete set of project processes individually to classify, analyze, discuss and collectively write an article for submission [20].

One of another outstanding web-based project named "Worm watch lab" on Zooniverse is the world's largest and the most popular platform for scientists to discover how

genes affect behaviors on nematode worms by receiving data from volunteers based on watching a video and spotting eggs worms laying [2].

Besides, Planet Hunter is also an astronomy citizen science project for participants to filter data collected from the telescope. The task requires participants annotates on light curve images to record the apparent presence of transits.[28]. Zooniverse provides a usable framework for different projects, and the interfaces of every page are fully functional with each module has a rich source of information. Not only the annotation task itself could support the research process, but also the implementation of ”Forum” and ”Blog” as important characteristics of the community are equally necessary. Especially for newbies, when they feel ambiguous about the annotation tasks, they could acquire more experience from the discussion board, and users feedback to understand. When they navigate pages, they might have more additaional meterials to learn from others to gain additional motivation on continuing the tasks.

Thus, creating the citizen science platform has the potential to facilitate research by using available large-scale data, however, validating the results is challenging. The data collection and further validation will be based on data quality, platform feasibility and corroborated output in my project.

Chapter 3

Methodology and Design

This chapter covers the implementation of the web platform including the prototype design based on user requirement and citizen science principle; the technology architecture about its framework and database design; the approaches of UI design and the method for data collcetion and validation.

3.1 Original User Requirement Analysis and Prototype design

3.1.1 Volunteer Engagement

User engagement is a prerequisite and criterion for platform construction, as voluntary participation denotes no obligatory enforcement and participation in current citizen science projects. Haklays research paper indicated that people who devoted themselves to citizen science projects always blurred the concept of ”who the real scientist is”. It is important to identify the boundaries for this definition. Ghose [21] described that a project can only be considered as a citizen science project if its participants decide to frame it as such.

A projects building framework is vital in determining some scientists expectations. In the past, as for scientists, the activity always emphasize on recording and visulizing data that follow a certain protocol rather than some other implementation of community-style blocks such as chatting or blog views for interaction. [26]. In this way, the user engagement would be reduced significantly.

Therefore, in recent years, the large-scaled citizen science platform are tend to do more investigation on how to motivate more users, even potential users, to take part

into the projects.

There are increasing demands for passionate volunteers desiring to make contributions to meaningful and interesting projects as well as for innovative citizen science projects that motivate and involve people from around the globe. Web-based citizen science is classified into three subcategories: VC(volunteered computing) [7], a more cognitive VT(volunteered thinking) [10] and the most current conception of PS(participatory sensing) [55]. Continued involvement signifies an increasing crowdsourcing data requirement that will be fulfilled through easier methods. Some outstanding projects show evidence that the community elements to encourage user interaction, and the gamification are two of the most effective ways to provide entertainment and participation for users in a project design. This is accomplished by implementing serious volunteer thinking games that represent scientific human-computation tasks [13, 14].

Gamification is an operative reference method for gathering volunteers and leading more users to science-related knowledge background, where the macroscopical model is simplified into operation with curiosity and enjoyment. In this setting, the rigorous conceptualization of project design as a research tool is necessary to accomplish broader acceptance. However, based on the main task in D.A.Lab project, gamification is difficult to appropriately applied in just annotation operation in the current stage. But it will still a field to be worth investgating.

3.1.2 Requirement strategies

To analyse basic user requirements and engagement strategies, it is important to identify the target population those who may want to take part or benefit from citizen science projects .

The main purpose of the D.A.Lab project is to motivate citizen science amateurs interested in discovering cures for patients afflicted with Parkinsons disease. This project strategy will be especially beneficial to Parkinsons disease patients who are unable to join practical laboratory tasks (because they either reside far from our laboratories or the disease affects their manual dexterity). Patients can still become Remote Crowd-Labbers, affording them the ability to make personal contributions to scientific research aimed at finding effective related treatments by marking and scoring bunches of images showing Drosophila in vials with experiment under therapeutic drugs. Listed below are some features, functions, and user requirements provided by project initiators that

make D.A.Lab a simple and very user-friendly interface:

- Users:

  Distance volunteers form the core of D.A.Lab. These volunteers gain remote access to the homepage by a simple and quick email registration and login process which is equipped with a password retrieval system. Each username login represents a personal account with an email address and password. The user can update his/her personal profile by accessing the account user center.

- Tutorial Guide:

  For users new to the platform, an understanding of the projects mission and vision is paramount. The tutorial guide is displayed on the homepage of the website which has adequate visualized introduction as well as step-by-step instructions. The guide will visually illustrate and guide participants how to annotate data just as real scientists and researchers do. There is an option to simply read the instructions should the user believe the tasks are easy to perform. Aside from the text and graphic guides, a link to an even more detailed tutorial video is located at the bottom of tutorial page, which is designed for people who need more clarification and instructions to accomplish annotation tasks. The compulsory tutorial guide setting may be reviewed if we decide to add more elements and update the annotation task processes to be more complicated or challenging. Currently however, it remains as an optional tutorial mode.

- Annotation Task:

  It is a simple task open to volunteers of all age levels. The task involves annotating the number of Drosophila in six different vials where one vial has four tube slots. Volunteers can accomplish this task with just a simple swipe of the finger, after reading the required detailed textual tutorial guide or watching the detailed tutorial video guide. Users can freely choose the images they want to annotate by swiping left and right on the scroll bar located at the top of the annotation page. To enable users to make more accurate annotations, especially for those who are near-sighted and elderly [29],a magnifier function that covers the entire area of the annotated images is provided for even greater user-friendliness and intuitiveness.

- Community-based character

Project provide its participants opportunities for self-fulfillment as well as a venue to subsequently enjoy social relationships that brew from their personal interactions with other users and volunteers [20]. For sustainable long-term user engagements and interactions, adding more community-based character is necessary. A "Feedback" page was taken into consideration wherein individuals can find a direct link to questionnaires and comments associated with the project. It is a venue where users have the freedom to express anything they want in text that will be displayed on the sites feedback page. The "User Center" is available to all users where they can modify personal information, to include personal avatars, nicknames, birthdays, mobile phone numbers, and addresses. These supplementary features are vital in building long-term and sustainable scientist-volunteer relations. This social interaction can be accomplished in user forums, comment broads, questionnaires, online videos for help and tutorials, real-time message exchange systems, and the like.

3.1.3 Principle of citizen science

A new principle was purposed by Galaxy Zoo [37]

- Do not make participants feel that they are wasting time.

  Ensuring that the data contributed by the participants is valid and available for the further research or publication. Participants are generally reluctant to spend time doing meaningless things. The suffcient information related to the project itself should be a support to inform participants of what they are doing and what sort of outcome they will get.

- Don not do anything computationally that you dont need humans to do.

  For instance, it is possible to recognize Drosophila automatically based on computer vision algorithms, but in the current state of the art, artificial annotation is more reliable and effective than computing. In the future work, the large amount of data generated from human are useful for machine learning to validate and evaluate whether the user count a right number as most users do.

These two rules are encapsulated into decision-making at D.A.Lab to guarantee that users contribution to Parkinson's disease would be of value to patients, scientists and themselves.

3.2 D.A.Lab Technology Architecture

3.2.1 Web framework

D.A.Lab is based on a widely-accepted data collection approach: A mass collaboration involving a large number of volunteers distributed across time and space could submit individual data recordings from their respective observations and analyses utilizing web-based technology. With limited time for project development and no prior experience in web development, I launched the project with a web framework. A web framework generally shortens the web-building process by providing an existing structure for application development. This allowed me to pay more attention to the applications business logic instead of the routine elements [39]. Web development is a fast-paced industry, with new frameworks and libraries popping up every other day. Web stacks comprise of multiple languages used concurrently a software bundle that runs everything from the front-end to the back- end, from databases to server administration, to rendering the website [3].

Comparing six available mainstream web frameworks (Flask, Pyramid, web2py, Zope2, Grok, Pecan and Falcon), Django is the most popular one with diverse features to meet the full-stack development. It is based on the Python programming language and refers to some excellent websites and applications (Google, Youtube, Instagram, Yahoo map, Pinterest etc.) [42].

3.2.1.1 MTV structure

The D.A.Lab platform is designed utilizing the Django web framework, which is a full-stack, batteries-included Python web framework that encourages rapid development and pragmatic design. It takes care of the laborious hassles related to web development. A Python web framework includes necessary features such as URL routing based on regular expression, ORM (object-relational mapper), authentication, template engine, and database schema migrations. Furthermore, Django does not require any front-end framework for creating dynamic web-services as it utilizes its own template language mixed with HMTL. A Model-Template-View paradigm was implemented in the Django framework [54].

- Models(M)
  - A model is the definitive source of information about data. It contains the essential fields and behaviors of the data being stored. Generally, each

Key Features of Django	
Django Comes with "Batteries-Included"	Authentication, URL routing, a templating system, an ORM and database scheme migration.
Free API	No additional coding required as a simple command is good enough to begin generating APIs.
Unique Admin Interface	The framework empowers you to quickly and easily generate an administration site straight from an application's models.
Code Layout	Django makes it easier to plug new capabilities in your product by using things called *applications.*
Django' s ORM Takes Care of Databases	No hassles of dealing with the SQL, which is mostly used to query the database for the data needed. Django works out-of-the-box with relational database management systems like MySQL, SQLite, and Oracle.
Robust Built-in Template System	Django' s templates allow developers to build entire, dynamic websites out of a very small amount of front-end code.
Simple, Readable URLs	Easier than other frameworks for URL construction, thus it could create simple and readable URLs based on regular expression.
Automatically Creates Tables in Database	Creating missing table by executing the migrate command in Django.
Easy Database Migrations	Easy to track database schema and its associated changes. Migration names help in managing version control, and a plethora of options are available to merge versions and make modifications.
Security	Default protection against XSS attacks, CSRF attacks, SQL injections, clickjacking, user management, cookies, email header injection, cryptography, directory traversal etc.

Figure 3.1: The key future of django web framework which was concluded in Manmohan's article[31]

model maps to a single database table. As the data access layer, it provides the API to access database and allows the developer to pull data from your database without knowledge requirement on the intricacies of the underlying database. The model usually also provides an abstraction layer with the database, so that the same model could be applied to multiple databases.[1]

- Templates(T)
  - A Django template is the presentation layer for models, which display the page contents to users directly and obtain user inputs from front-end layers. This layer contains presentation-related decisions for displaying content on a web page. In this project, Hypertext Markup Language (HTML) is used for layout, which is rendered with a context combined with Cascading Style Sheets (CSS) for interface styling. Javascript and Ajax (Asynchronous JavaScript and XML) for client side code to optimize interface styling and to enhance user interaction with the web pages.[1]
- Views(V)
  - A Django view represent business logic layer which contains the logic that accesses the model and defers to the appropriate template(s). This, in turn, acts as a connection between models and templates. A view function is simply a Python function that takes a web request and returns a web response. This response can be the HTML contents of a web page.[1]

In conclusion, although multiple language makes the process of the web development difficult, the MTV architecture in the Django framework provides a way to separate the web interface (template layer) from the domain logic (model and view layers) making it easier to implement and modify each of these components independently. The View layer receives the request from the user's browser, invokes a desired functional operation and returns the template (i.e. the graphical data representation) as a response to the browser.

3.2.2 Database Design

The conceptual schema is a concise description of the users data requirements and includes a detailed description of the entity types, relationships, and constraints. The D.A.Lab database is aimed at satisfying users requirements with simple structures and

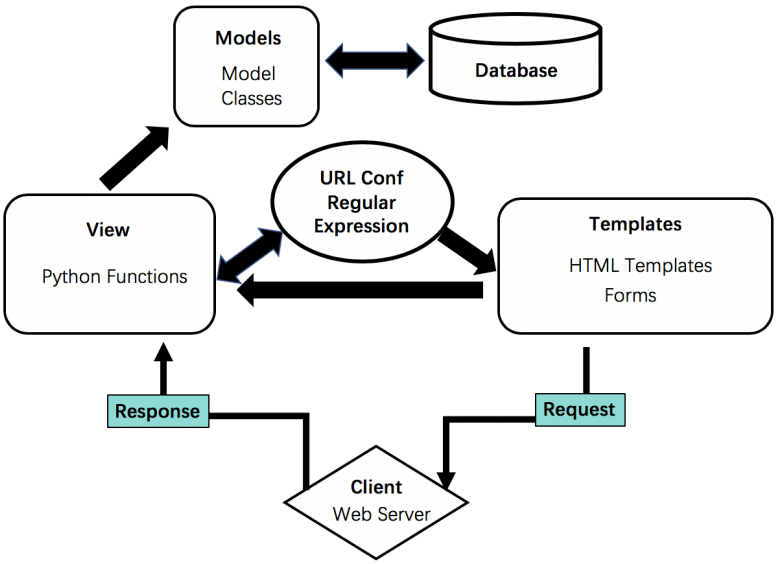


Figure 3.2: The Response Pattern of MTV Structure

easy administration, which mainly stores data about user profiles, user comments, sample image files, correspondent annotations to each image, and the interaction among these three. Users gain direct access to sample images by entering the annotation interface and accomplishing annotation tasks by inputting integers in a 6 * 4 form field before submission. A fundamental setting contained in this database is that a single annotation task is essentially identified as a user interaction with an image.

[Need to do more introduction of the relationship between entities in a textual way? [e.g. The structure of database is clear and simple to build a mainly logic relationship between user, annotation data and drosophila climbing images. User_id as a foreign key.....?]

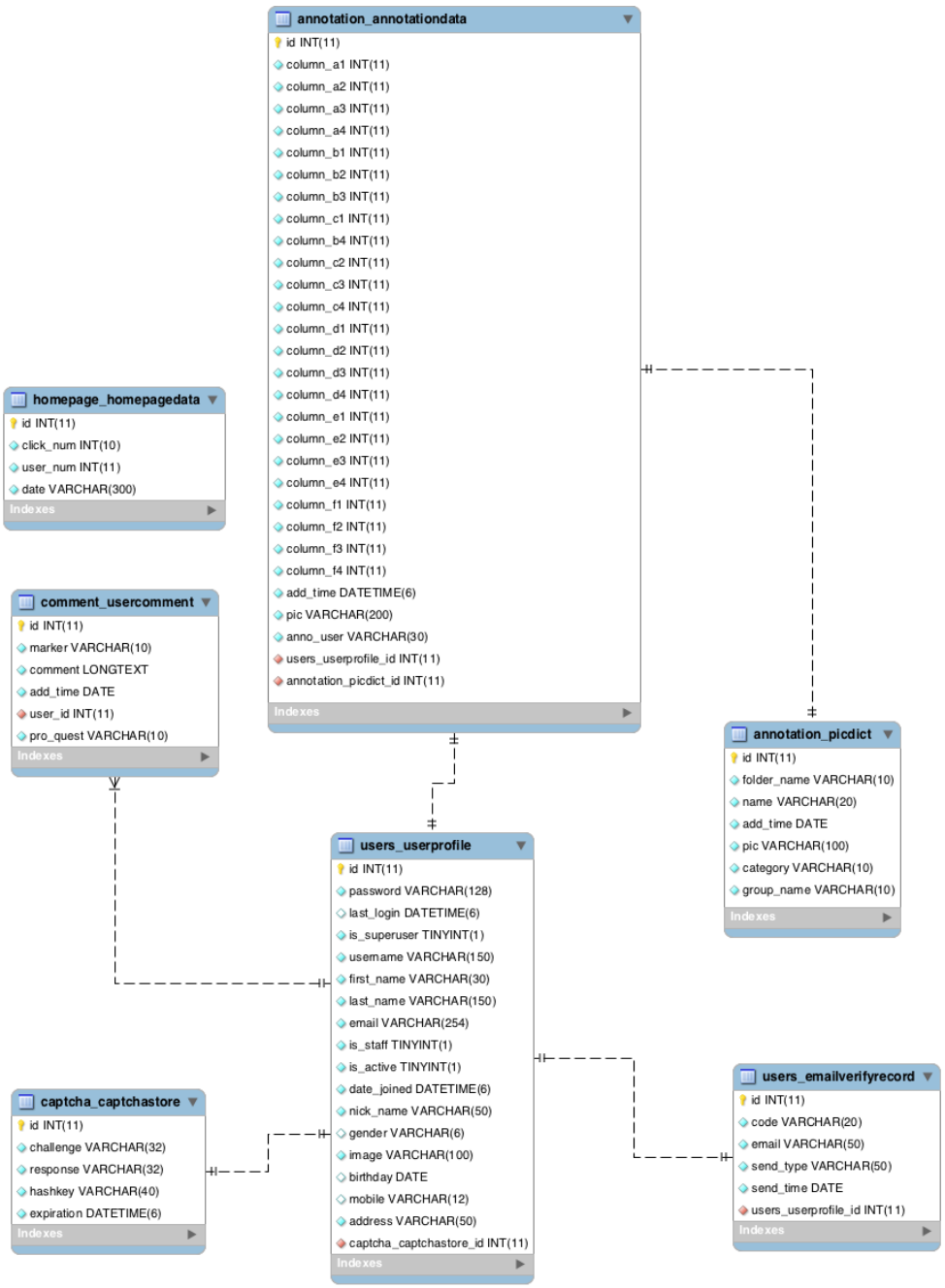


Figure 3.3: The key future of django web framework which was concluded in Manmohan's article[31]

The fully-automated Management Backend in Django framework: I only needs to use ORM and perform simple object definitions. It can automatically generate database structure, which comes as a function in this full-featured management background. ORM(Object-relational mapping) in Pythons web framework interacts with the back-end database. This back-end framework assists the user to best generalize and define

data within the site for future projects developments that will utilize the same or similar web architecture.

My project employs two tool applications for visualized data management:

- **MySQL Workbench** MySQL has excellent GUI tools in the MySQL Workbench, which is a unified visualized tool for database architects, developers, and DBAs. It provides data modeling, SQL development, and comprehensive administration tools for server configuration, user administration, backup, and much more.

- **Xadmin** Xadmin is a drop-in replacement for the Django admin system that comes with lots of goodies such as fully-extensible plugin supports and pretty UI framework based on Twitter Bootstrap. Xadmin is only accessible by superusers to manage back-end database directly. Its built-in enhanced filters are very convenient for data query, and each management interface can optionally display the required fields in each model, avoiding data fields in the interface from being too redundant while managing the database quickly and efficiently.

3.3 UI Design-Related Approaches

A good platform design ensures the best usability and motivation. Identification with the online citizen science platform is the primary motivational factor that is highly relevant to understanding user response in collaborative efforts [34, 41].

The interface should have its own internal reconciliation, which is the most fundamental requirement of users. An unaesthetic interface made up of rough colors, imprecise lines, and uncoordinated fonts may be unlikely to inspire user confidence emotional interaction of the user's engagement. UI design directly or indirectly affects usability of the platform, user friendliness, and user satisfaction.This user-centric design approach is based on a historical experience perspective of classic cases or general user surveys [24]. While considering user operation practicality, emotional and psychological satisfaction should also be fully considered to improve user engagement [32].

3.3.1 Color design in UI

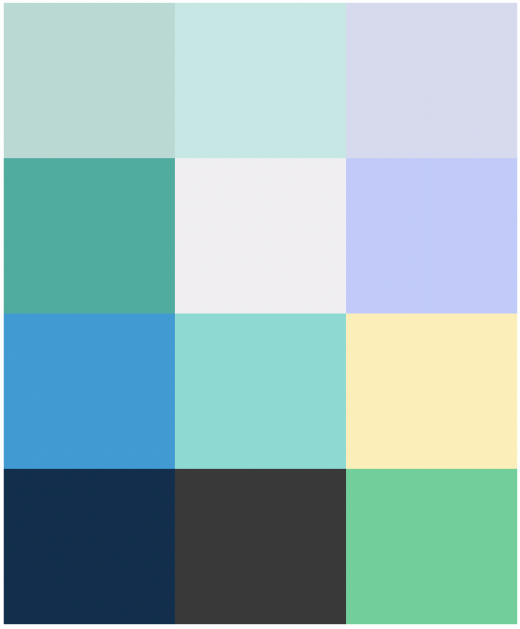


Figure 3.4: Main color scheme is D.A.Lab interface

Color preference is sort of the main indicator in users' visual experience. It is fairly prevalent in color psychology [22, 49] to use neutral shades and cool color in the overall interface to create an evocative tone for users that enhances concentration and helps alleviate visual fatigue(Fig 3.4).

3.3.2 Animation added in interface

Based on previous analysis on human processor model it is difficult for a human eye to perceive the anything change if the animation duration is under than 230ms [11]. The optimal duration suggested for animations is that it should not be shorten than 1000ms on mouse input. From this perspective, some interesting cartoon icons in some expressive form(e.g. big Drosophila icon, "Start!" button and "Tutorial" button on the homepage etc.) could produce a shake-response results from users' operation, when their mouse cursor hovers in an adjacent area.

Designed to simulate the tremor action of Drosophila, such an animation would make the icon more vivid, arousing an engagement and interest whilst attracting user attention, which facilitates an increased likelihood of mouse clicks to explore the interface.

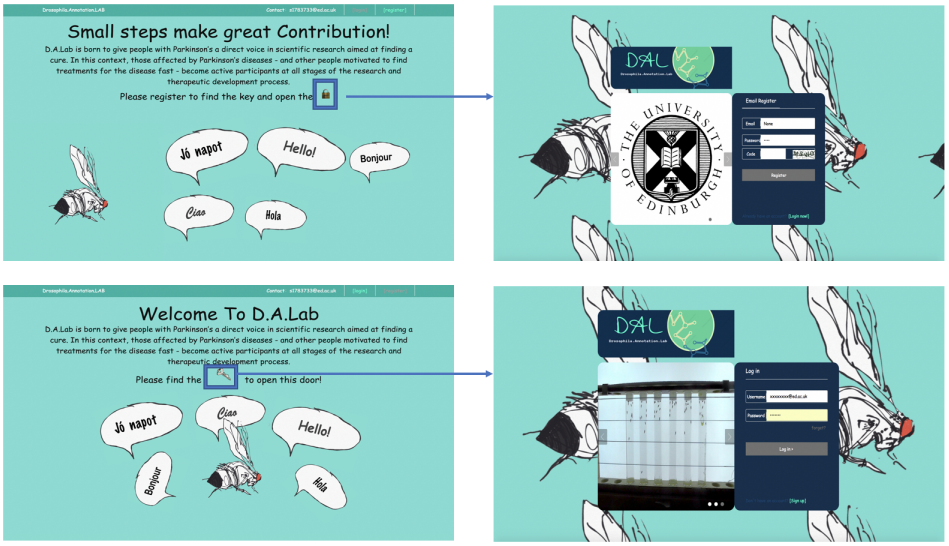
It helps add instruction and order to design schemes that may be visually too simple to provide enough direction for users. Animation, in this instance, is a lively blend of stripped-down simplicity and usability.

3.3.3 Page Display and functional description

- Login and Registration page

Login and registration pages operates on click logic, with triggered objects (key and lock icon) that motivate users to click on the icon and proceed to a hypertext reference related to login and registration function.

Figure 3.5: Login page and registration page



- Homepage

The homepage design is simple and straightforward. I removal the unnecessary messy or dazzling pictures, videos, photos and colors of the papers, stuck to the color scheme, and keep the page simple and well structured. Only two buttons, ”Starts!” and ”Tutorial Guide” are kept on the page. In combination with the shake animation and cartoon visual elements, the core information is emphasized, and the user reaction time and doubts are reduced in the most direct way, so that they can quickly obtained the information. Users could choose to watch the tutorial guide or start annotation work directly.

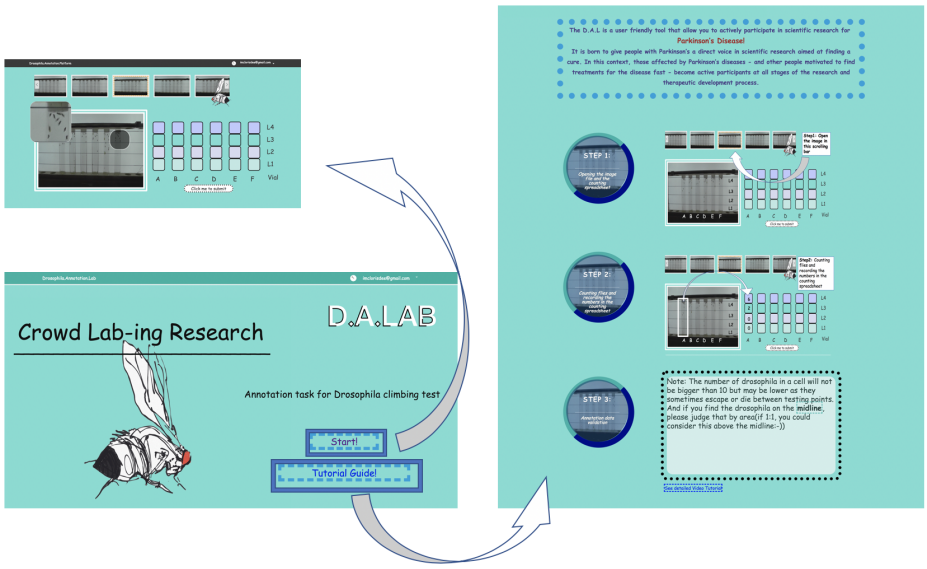


Figure 3.6: Homepage interface with the function of buttons and their related links

- Annotation task interface

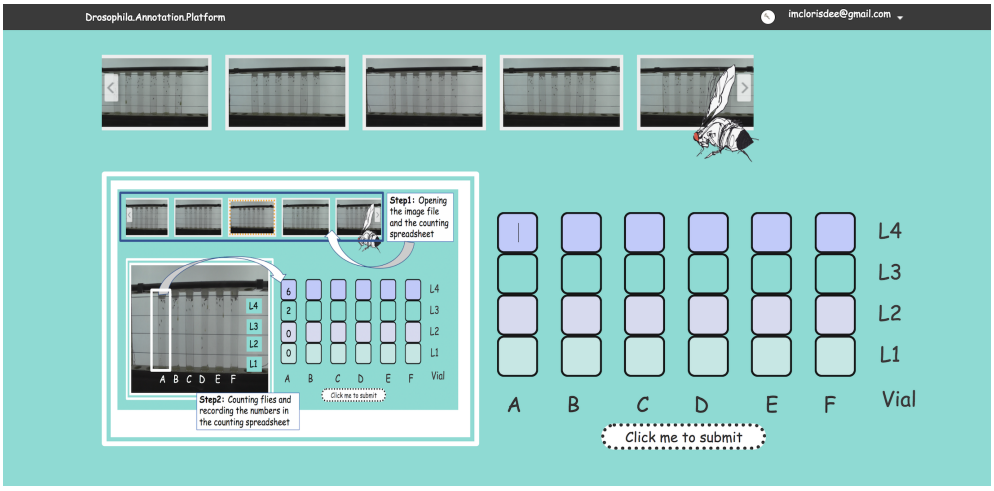


Figure 3.7: The initial annotation interface when users firstly click into

To handle the short-term limitation of participants, a summary step-by-step annotation guide reappears in the task interface; allowing users a quick and accurate review of the tutorial, without having to constantly return to the guide page.

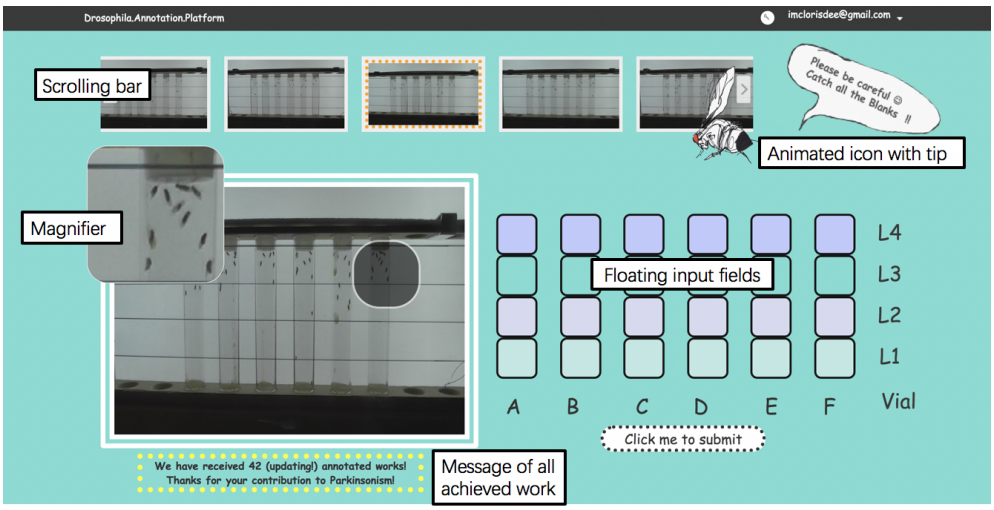


Figure 3.8: Annotation task interface

Considering the target population (Parkinson's disease patients) and the universality of the platform, annotation tasks would be accomplished with simple operations, as soon as possible. Participants only need to input integers into floating fields on the right-hand side, and then click the submit button to upload correspondent data with every image files to the back-end database. Several character could be introduced:

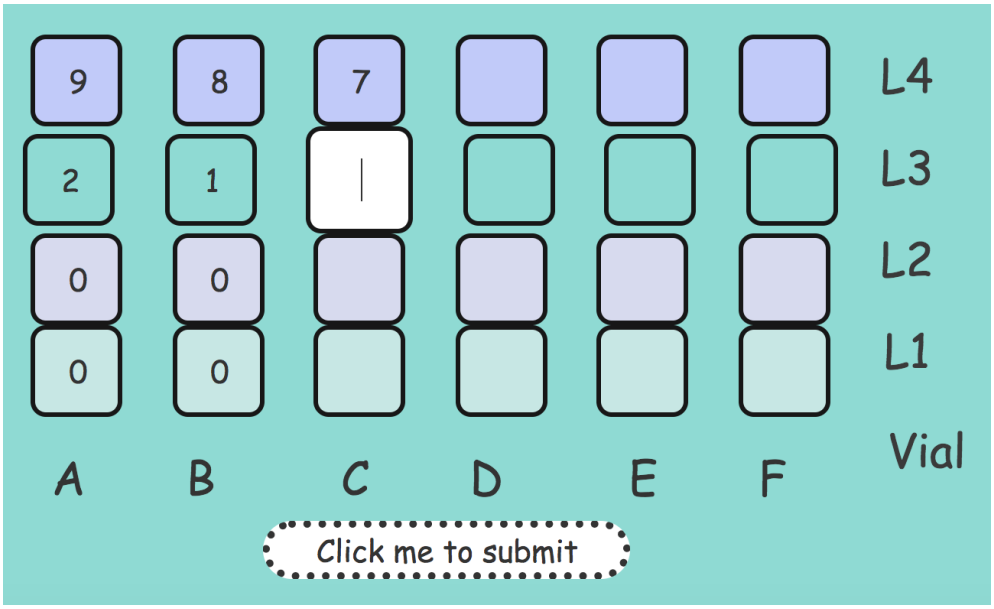


Figure 3.9: On-focus input field for drawing users attention on the their current input

– **Floating input field**

When the users mouse hovers over the current input box, it automatically magnifies and is emphasized with a white background to display that it is on focus. This setting is applied as soon as possible to increase user concentration, and avoid omission.

- **Scrolling bar**

  Used to freely select images.

- **Magnifier**

  Especially for near-sighted and elderly users, magnifier is aimed at making Drosophilas in image files that are easier to see and recognize.

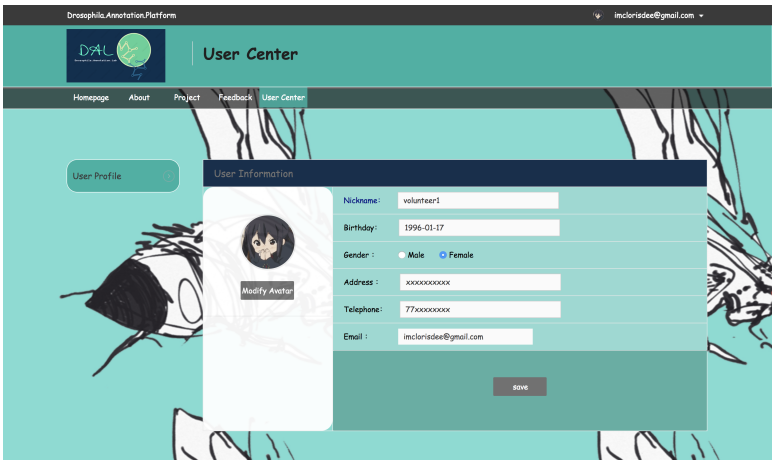
- **Animated drosophila icon with the slogan tip** An animated Drosophila icon was added to the pages for interactivity and optimized user experience, such that when users inadvertently hovered over the icon, the animated Drosophila icon bubble window popped up as a guided reminder that actually said: Catch all the blanks! It is designed to prevent user negligence in annotating work, and add interactivity and comic relief, especially for attracting the youth. A more rigorous manual-input validation algorithm will of course, be mentioned in the next section.

- **Message of work achieved**

  The message would appear together with the slogan tip, as a result of the hovering cursor. It turned out to be a seemingly inadvertent feature, as a first attempt to increase user motivation, showing the amount of work done, and informing participants as to ”how much was done” and ”their contribution” with a ”precise figure”. From a psychological perspective, when people are making positive changes, it always induces optimism. They may make only little progress, but once people are inspired to believe that small steps shape great contribution, their motivation is greatly enhanced. This design feature is simplistic and still only in proof of concept (to be optimized for future work). The impact it has on user feedback evaluation is discussed in the next chapter.

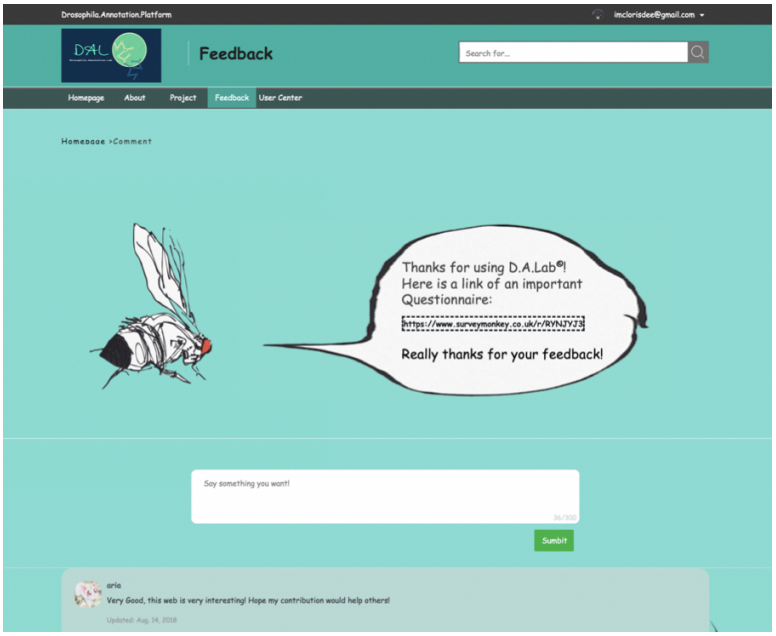
- User center page

Figure 3.10: User Center page



- User Feedback

Figure 3.11: User Feedbak page



3.4 Data collection

There are two indicator for project evaluation based on data collection, one is the data statistics comes from the back-end database, and the one is user feedback from the survey.

3.4.1 Manual-input validation

Scientists start the experiment with 10 Drosophila in each vial. The number of Drosophila in the test tube could be less than 10 due to the climbing test experimental design. We made an assumption that the flies might potentially escape from the vials during treatment, feeding or could die for any reason, perhaps when trying to get away from the test tubes or during breeding and treatment, and some of them may die during the recording of the climb status. In order to ensure the quality and integrity of data generated during annotation process, with minimum fuss or room for user errors, several constraint mechanisms were implemented in the front-end algorithm.

Thus the pattern validation method could include retrieving a value for a form-based manual-input field from a form defined in markup rendered in a content browser. The retrieved value along with a validation pattern for the form based input field could be passed on to a validation process disposed within a lightweight validation library coupled to the rendered markup. Client authentication occurs at the browser end before the submission of form data. As compared with server-side validation, it creates a better user experience through real-time verification process feedback on input results.

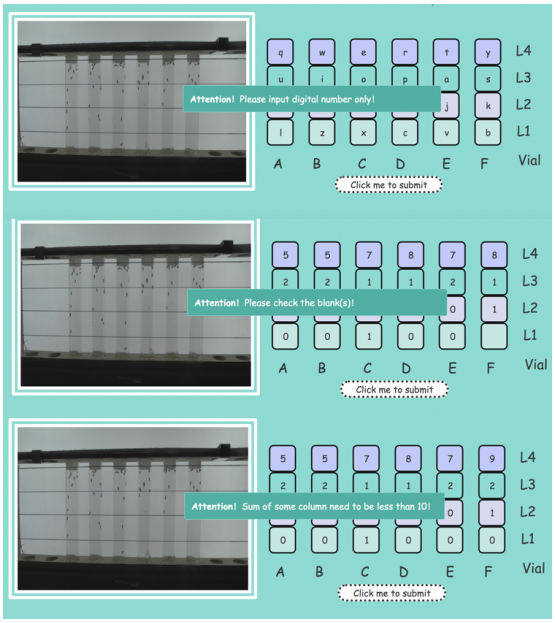


Figure 3.12: Attention tips for validation

The set constraints for user input include:

- Digital number only (two-digit number only)
- Sum of a column cannot be greater than 10

- Copy and paste are not allowed
- Non-empty validation

(See main Code for the HTML form submission attribute and JavaScript API achieve form validation behavior in D.A.Lab in appendix)

3.4.2 Survey Design

The first version of D.A.Lab is only aimed at investigating possible crowd sourcing solutions for research data annotation, as well as matching basic user requirement and expectation with design principle. So, there is still a significant room for further improvement. The following questionnaire was set up in order to analyze the practicability of the platform and subjective user sentiment and attitude. It aids evaluation of current endeavors, to improve platform architecture and further design principles.

The survey contains 15 questions, which analyze the user's psychology from two aspects: the user's actual experience, the user's engagement and the user's further expectation. (See it in appendix)

Chapter 4

Analysis and Evaluation

4.1 Self-Evaluation

In this article, our objectives were to (1) analyze users' requirement; (2) prototype and mockup design; (3) develop a web platform; (4)collect data from back-end database; (5) survey collection; (6)analyze and evaluate from the data results; (7)improve the platform.

The platform has achieved its basic functionality in the current phase. A normal, and ideal user operation logic sequence should be:

1. Click on the registration page, enter the email address and CAPTCHA(Completely Automated Public Turing test to tell Computers and Humans Apart) to register, after which an activation link will be sent to the user email address, so they are verified and can then log in directly to the homepage.

2. From the homepage of D.A.Lab, the user is given an option to see the annotation tutorial guide (and detailed video), or go directly to the annotation task interface.

3. Start annotation work. The user can start annotation work, by selecting an image from the scroll bar that is then displayed in a larger area to enable the annotation task. When the user cannot clearly distinguish the Drosophila and count them, the cursor can be placed over the picture area to enable the magnifier function.

4. Filling in the 24 input fields. The validation constraints are implemented, hence a reject alert box pops up for any empty value, non-integer number (maximum of two digits), or when the single-column sum is greater than 10. The data inputs will not be saved and submitted until the user has modified them correctly.

5. After the data is successfully submitted, the input fields cleaned automatically, with a successful submission status update, and the images will move to the next one .

6. The user could allocate their time freely to do the annotation tasks, with no constraints on the number of images they would like to annotate. Furthermore, they can selectively enter the feedback interface, click on the questionnaire link, or leave a comment. Besides, a user center page design feature allows user to upload an avatar and modify the profile, which performs as the basic function in every community web platform.

4.2 Usability Evaluation

The factors of usability are evaluated from users' experience, effectiveness [61] and accessibility. There could be two indicators including task completion rates/times and standardized questionnaires, could measure participants attitudes or preferences (Chin et al. 1988, Lewis 1995, Brooke 1996, Tullis and Stetson 2004, Pearson and Pearson 2008).

In this way, a psychological measurement is thus executed to.

4.2.1 Basic Profile of Participants

Volunteer participants from the University of Edinburgh included 14 students, 5 neuroscience researchers and 1 Parkinsons patient. At the beginning of the survey, a total of 21 participants were asked their age and experience in citizen science. The largest group was aged 18-24, accounting for 66.67%, 3 people between 24-33 years constituted 14.29% as the next group. 9.52% were aged 35-44 years, and that was 2 people.

The total number of participants were 21 and they were asked to give their age and experience of citizen science before at the beginning of the survey. The largest group were aged between 18-24, accounting for 66.67%. 24-33 age group included 3 people, accounting for 14.29%. 2 people aged 35-44 years, accounting for 9.52% and the elderly over 65 years old was only one individual. In the above cited data population, only two were involved in citizen science projects previously, but the remaining 19 had no awareness on the subject, and indicated they had never been exposed to any citizen science projects.

4.2.2 Exposure to Parkinsons Disease

In addition to the patient who was suffering from Parkinson's disease directly, 9.52% of those participators had relatives or friends that were suffering from the disease. And 80.95% of those surveyed concerned themselves with exploring how to support Parkinsons patients, even though they had no relatives or friends with the disease.

In the next related question set in the survey, users were asked whether they were informed about the disease, and the processes and challenges in drug research. 23.81% of users, mostly academics from related fields, said they followed the latest scientific research. However, 28.57% of non-professional users admitted to having read the information, but were not sure how much they really comprehended. And the remaining 47.62% of the participants stated: No, I am not very well informed, and I find scientific research too complicated to read. Due to the diversiform knowledge domain and level of education, gaining a comprehensive understanding of Parkinsons disease from a theoretical background, and seeking treatment, is not possible or something that is even accessible to every non-professional. Having said that, most people are aware of the disease to some degree, and at the least, are cognizant of the Parkinsons population through different channels(TV media, Internet, etc.). An empathy that drives their involvement in these citizen science projects serves as a huge positive factor, and they are more willing to contribute towards treatment and support initiatives to help manage the symptoms for Parkinsons disease patients.

This requires further work: if D.A.Lab evolves into a wildly-used online citizen science platform, it could be the crucial step to make Parkinsons, from a pathological level, clearer, simpler and more comprehensible to the public. With means to explain the causes of Parkinsons disease (e.g. simulating the central nervous system disorder through animations), and through relevant cutting-edge practices (such as visualizing latest research data, and reporting that on the homepage), public engagement could improve considerably.

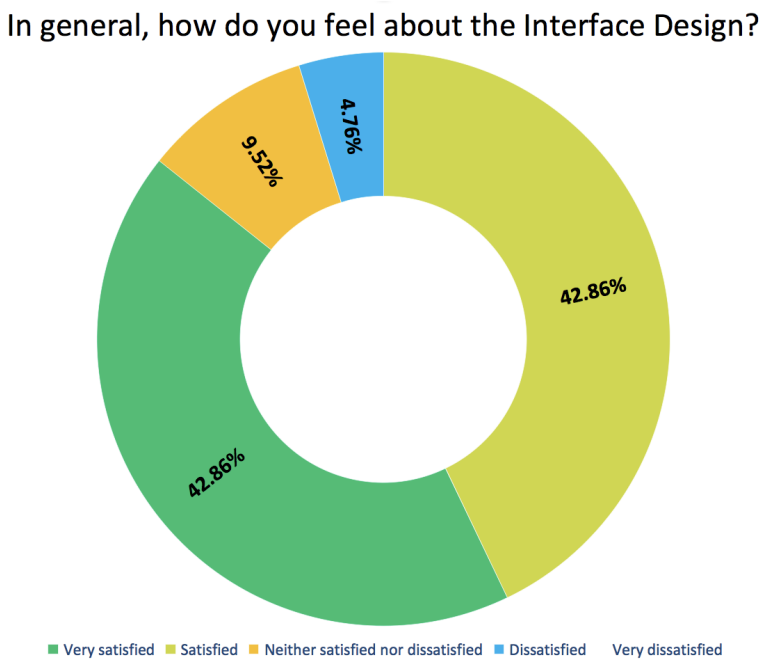
4.2.3 Task Completion Rates

Among the 21 participants, 70% annotated two or more images, with 5 as the highest image tally for one user. 30% of them worked on only one image. For users working on multiple images uninterruptedly, the average time spent on each annotation was 2:03 mins. This indicates that a user would spend an average of 5.125s on completing each input field.

4.2.4 User experience analysis

The user experience is mainly evaluated from these aspect:

Figure 4.1: Attention tips for validation



From the satisfaction degree of user feedback, in general, the overall D.A.Lab interface design has met basic expectations, with 85.72% users feeling satisfied or above, but there are 9.52% and 4.76% users as well, who are either 'neutral' or 'dissatisfied' respectively.

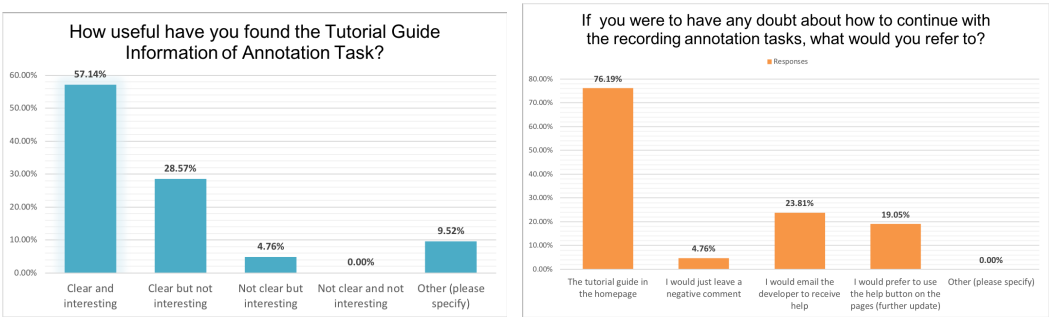
4.2.4.1 User Feedback on Tutorial Guide

Some questions were set in a more detailed format to collect the participants' experiences in operating an annotation task related interface: The tutorial guide is considered to be a factor of the efficiency and quality that directly affects the annotation task. If the users swiftly mastered the tutorial guide, that would greatly aid their average task completion rate.

User feedback on the survey questions highlights their dependency on the tutorial guide (fig 4.2(a)). 76.19% users chose to return to the homepage to watch the tutorial guide again, when they were unfamiliar with or forgot a specific annotation operation.

The next question set was relevant to evaluating the efficacy of the tutorial guide(fig 4.2(b)). 57.14% of users thought the tutorial guide was clear and interesting. This

implies that nearly half the users think the tutorial was either unclear or uninteresting. We received two specific Others as response, indicating some users didnt even look through it.



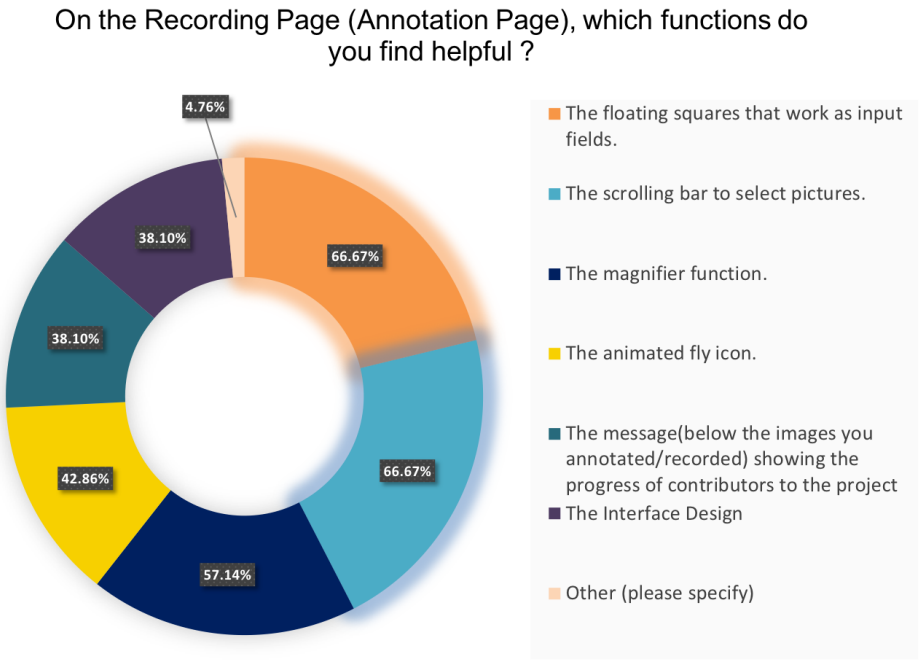
(a) question1 (b) question2

Figure 4.2: User engagement analysis

This design version allows users to get information quickly, but the simple static graphs applied to represent the tutorial guide make it uninteresting and unintuitive, in comparison with the Worm Watch Lab, a similar citizen science platform that boasts a large and stable user group. The tutorial guide interface displayed as a pop-up animation in the Worm Watch Lab (wor, 2016), when volunteers first click into the annotation work. Optionally, they could directly choose to watch or close the window. In possible optimization going further, a more step-by-step, easy-to-follow engaging procedure will be implemented.

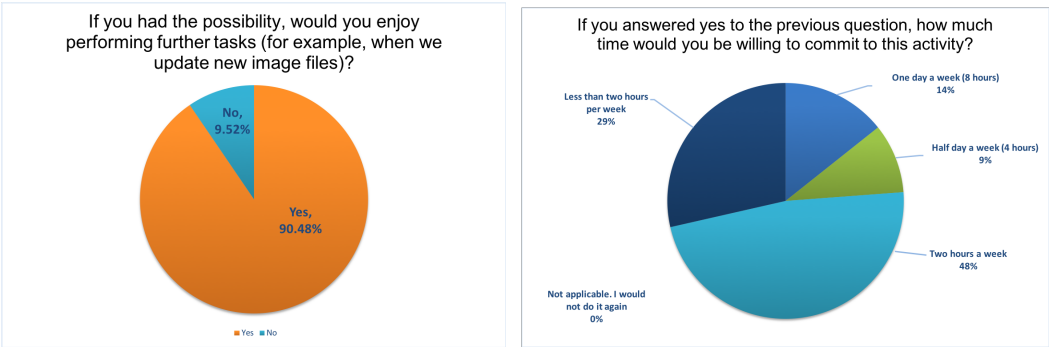
4.2.4.2 User Feedback on Annotation Interface

Figure 4.3: Users scoring for different aspects of D.A.Lab



The results indicate that the functional requirement of the user for annotation work assistance outweighs aesthetic requirement. The input fields autofocus function(66.7%), the scroll bar to browse images(66.7%) and the magnifier function(57.12%), are helpful, as per most users. Animated icons with interactive elements(42.86%), updates on project contributors progress that appear as messages below the images the user annotates(38.1%), and interface design (38.1%) were considered less important. Some users suggested adding an achievements list-style display of annotated images to the interface that showed what they completed successfully versus the total amount of unannotated images, which could serve as a clear indicator of the ultimate goal; plus a reward mechanism of points for each task accomplished. Furthermore, a competitive framework could be suitably considered and built into this project, possibly by ranking the points(or other substitution) that the user acquired on a weekly basis to boost their sense of accomplishment.

4.2.5 User Engagement Analysis



(a) SubfigureCaption (b) SubfigureCaption

Figure 4.4: MainfigureCaption

Table 4.1: Do you believe that CrowdLabbing and D.A.Lab projects can benefit people with Parkinson's in fighting the disease?

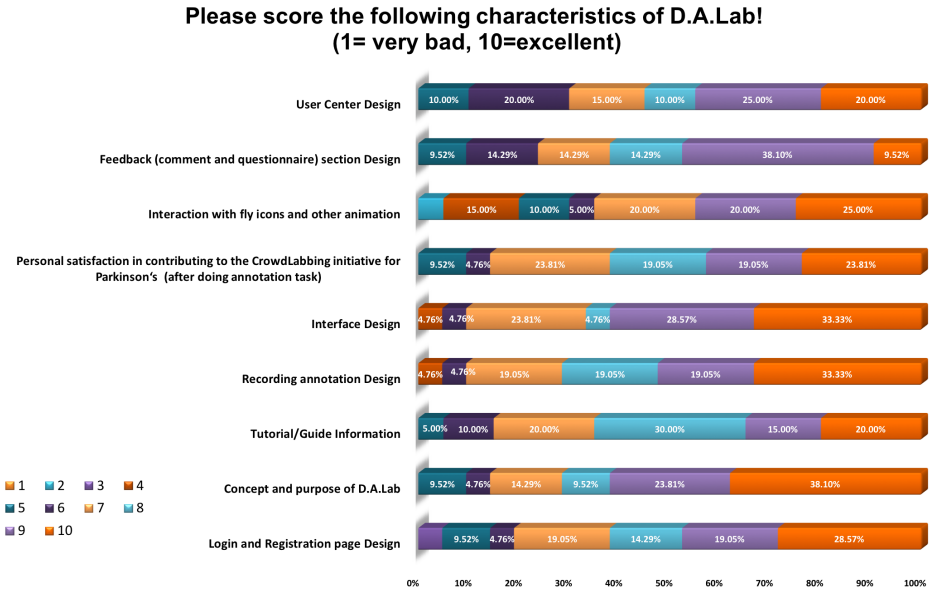
Answer Choice	Response
Yes. I participated and believe that it maybe really helpful and motivational to anyone interested in Parkinson's disease.	57.14%
Yes. I participated but not sure that it may be helpful.	19.05%
Maybe. I did not participate in it but I believe it could be beneficial and I would like to take part in the experience.	14.29%
Not sure, I do not know enough about Crowdlabbing and D.A.Lab.	4.76%
No, I believe that this work should only be done by professional scientists.	0.00%
Other (please specify)	4.76%

These findings demonstrate that 76.19% of participants who got involved in the annotation task gave positive feedback on the motivational and social value of D.A.Lab. 90.48% users enjoyed the experience and confirmed performing further annotation tasks through D.A.Lab. But when they were asked about the time they would you be willing to commit to this activity, responses were clustered around two hours (47.62%) or less (28.57%) per week. (The concept of D.A.Lab is to create a platform for volunteers who could allocate free time to participate in scientific research). With reference to the task completion rate, two hours a week is roughly equivalent to 8 annotated images per day that a single user could accomplish. User retention seems to be relatively impressive considering the current stage that the platform is at, but it is still a

challenge to meet the ultimate prospect, that is 'users spend over 8 hours a week in annotation tasks'. Annotation itself is a repetitive and tedious job, as seen with some users that quickly tire when in front of a screen for a length of time executing such simple tasks that involve only a swipe of their finger. It indicates the need to improve user commitment for the long-term development of D.A.Lab, creating perhaps a more appealing Drosophila annotating method in the future, which would be widely accepted and cover all user groups simultaneously.

4.2.6 Overall score of D.A.Lab

Figure 4.5: Users scoring for different aspects of D.A.Lab



4.3 Data Quality Evaluation

4.3.1 Data Integrity

Data integrity constraints are meant to prevent non-compliant data from entering the database. When a user inserts, modifies, deletes, etc., the DBMS automatically monitors the data based on certain constraints so that the non-compliant data cannot enter the database, and to ensure that the data is stored in the database correctly, efficiently, and is compatible. The analysis of the annotation data obtained from the database shows that the constraint validation settings performed well to ensure the integrity of the user input, which is as under:

- **Domain Integrity:**

  Domain Integrity is related to data type and additional constraints. In this project, each column in the table satisfies the constraints of the integer field, with no missing or blank values, and the sum of the values of each column is consistent with the rule of no more than 10 in the factual Drosophila Climbing experiment.

- **Referential Integrity:**

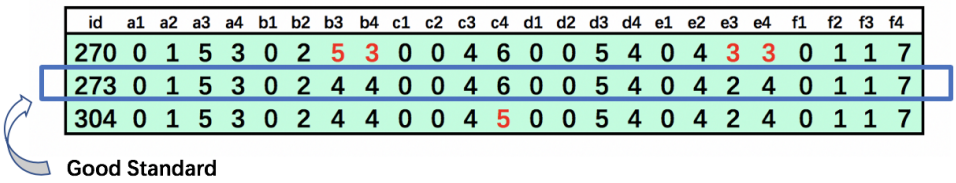
  Referential Integrity guarantees the consistency of the data regarding the logical connection between the tables so that the query can guarantee accessibility. We can get an easy access to find out all the pictures annotated by one user or the different users who annotated the same picture. It is necessary for preventing data loss or meaningless data from spreading in the database.

4.3.2 Data Consistency

This dataset consisted of a time-stamped data stream mainly containing user profiles, achievements, and for all users active during the 5 days from August 13, 2018 - August 17, 2018. In order to guarantee the coverage of annotation work towards all the images as much as possible, the images in the back-end database which refers to the front-end annotation interface are designed to be sorted randomly when users click into or refresh the page.

Since each user has the tendency to be subjective that causes biased perception, different participants are likely to have large or small deviations from the annotation data representing each test tube in the same picture. For example, they cannot distinguish the number of Drosophila that overlap together during data recording. Moreover, due to the climbing experiment of the Drosophila, it is difficult to avoid the adhesion of the drug on the inner wall of the test tube or the second stain. In this case, some users may even mistake the stain for a real fruit fly. Moreover, the adhesion of the drug to the inner walls of the test tube or the subsequent stain is inevitable. Some users may even mistake the stain for a real Drosophila. Considering the limited number of sample data and images in this project, the definition of a Good Standard(GS) is the same as annotated by most participants to one input field from the image which is consecutively annotated over three times or more.

Figure 4.6: This is the data from an annotated images which has been annotated for three times. We consider the row which has minimum dissimilarity to others as the Good Standard. a1 f4 are 24 input fields representing column_a1 to column_f4(see name field in database)



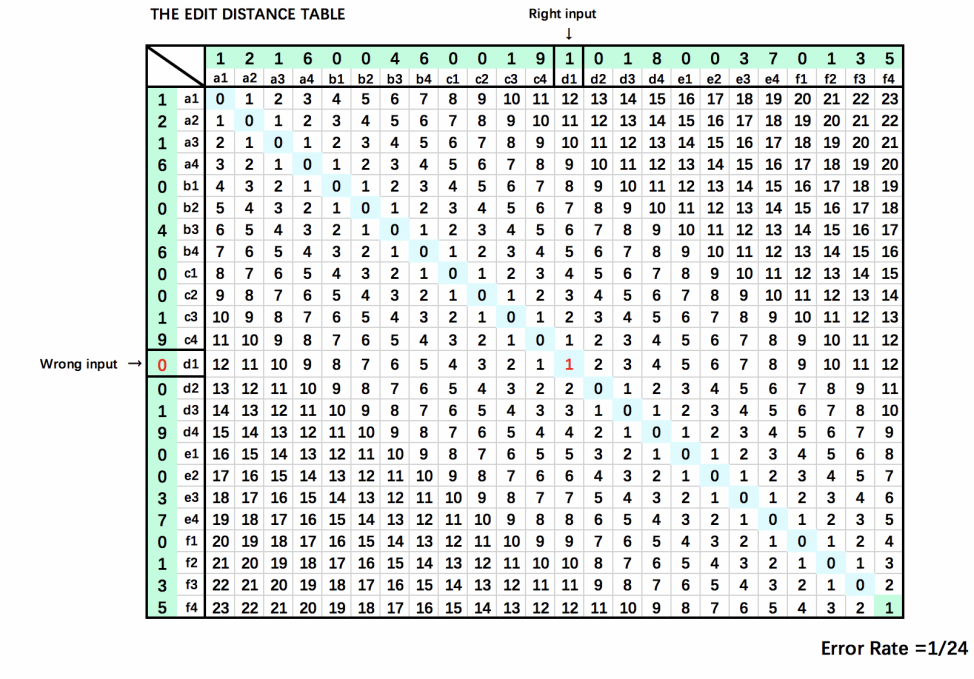
id	a1	a2	a3	a4	b1	b2	b3	b4	c1	c2	c3	c4	d1	d2	d3	d4	e1	e2	e3	e4	f1	f2	f3	f4
270	0	1	5	3	0	2	5	3	0	0	4	6	0	0	5	4	0	4	3	3	0	1	1	7
273	0	1	5	3	0	2	4	4	0	0	4	6	0	0	5	4	0	4	2	4	0	1	1	7
304	0	1	5	3	0	2	4	4	0	0	4	5	0	0	5	4	0	4	2	4	0	1	1	7

Good Standard

As mentioned in the Natural Language Processing field, an indicator for computing the word error rate is the **Minimum Edit Distance(MED)** between two strings[8]. As all pieces of the 2 data sources could be considered as long strings in the database, thus it becomes easy to calculate the error rate.

$$ErrorRate = \frac{Substitution + Insertion + Deletion}{ReferenceLength}$$

Figure 4.7: An Tabular Computation of MED



THE EDIT DISTANCE TABLE

Right input ↓

			1	2	1	6	0	0	4	6	0	0	1	9	1	0	1	8	0	0	3	7	0	1	3	5
			a1	a2	a3	a4	b1	b2	b3	b4	c1	c2	c3	c4	d1	d2	d3	d4	e1	e2	e3	e4	f1	f2	f3	f4
	1	a1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
	2	a2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
	1	a3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
	6	a4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	0	b1	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
	0	b2	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
	4	b3	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	6	b4	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	0	c1	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	0	c2	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	1	c3	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13
	9	c4	11	10	9	8	7	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	11	12
Wrong input →	0	d1	12	11	10	9	8	7	6	5	4	3	2	1	1	2	3	4	5	6	7	8	9	10	11	12
	0	d2	13	12	11	10	9	8	7	6	5	4	3	2	2	0	1	2	3	4	5	6	7	8	9	11
	1	d3	14	13	12	11	10	9	8	7	6	5	4	3	3	1	0	1	2	3	4	5	6	7	8	10
	9	d4	15	14	13	12	11	10	9	8	7	6	5	4	4	2	1	0	1	2	3	4	5	6	7	9
	0	e1	16	15	14	13	12	11	10	9	8	7	6	5	5	3	2	1	0	1	2	3	4	5	6	8
	0	e2	17	16	15	14	13	12	11	10	9	8	7	6	6	4	3	2	1	0	1	2	3	4	5	7
	3	e3	18	17	16	15	14	13	12	11	10	9	8	7	7	5	4	3	2	1	0	1	2	3	4	6
	7	e4	19	18	17	16	15	14	13	12	11	10	9	8	8	6	5	4	3	2	1	0	1	2	3	5
	0	f1	20	19	18	17	16	15	14	13	12	11	10	9	9	7	6	5	4	3	2	1	0	1	2	4
	1	f2	21	20	19	18	17	16	15	14	13	12	11	10	10	8	7	6	5	4	3	2	1	0	1	3
	3	f3	22	21	20	19	18	17	16	15	14	13	12	11	11	9	8	7	6	5	4	3	2	1	0	2
	5	f4	23	22	21	20	19	18	17	16	15	14	13	12	12	11	10	9	8	7	6	5	4	3	2	1

Error Rate =1/24

From the 48 existing records of the annotated data, only 4 images have been annotated over 3 times. Computing the error rate between wrong strings and Good Standard

respectively, we got an average error rate for annotating each image with at least made one mistake of 2.60%. It means that for each image, there is 97.4% chance of getting a piece of the same annotation data from different participants.

Chapter 5

Limitations and Further Work

So far, based on the the results and feedback of the tests of this project, D.A.Lab's prototype has been basically implemented, and users could provide accessible data for assisting neuronscience scientists to do further research on the treatment of Parkinson's disease.

Like some other mature citizen science web platform [44], D.A.Lab's pages tell about the project and investigators, registration and login systems, data annotation functionality, feedback box and detailed information about the participants.

This chapter contains some limitaions of the work and the further investigation including but is not limited to users' further expectations.

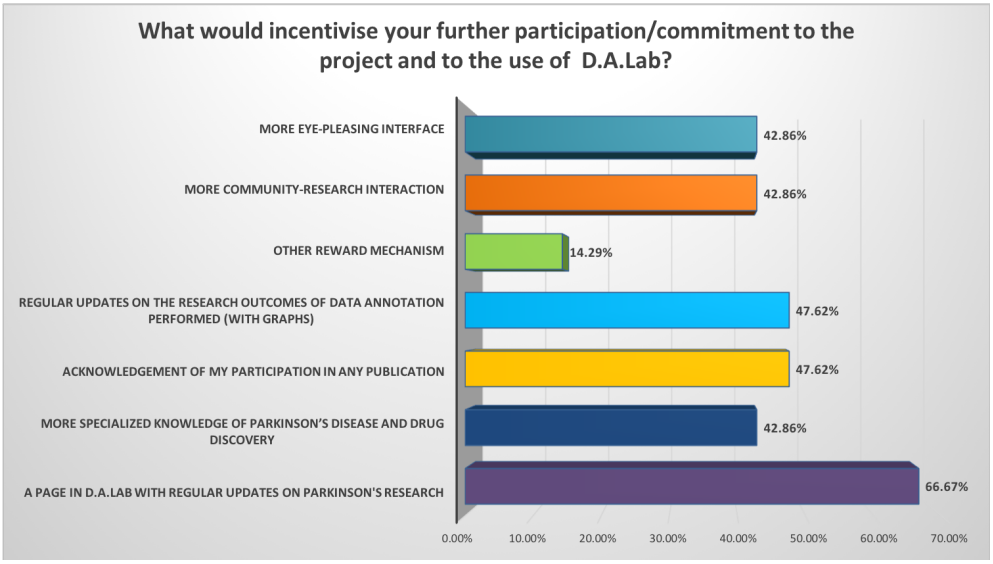


Figure 5.1: Further Expectations from users

5.1 User Stickiness and Motivation

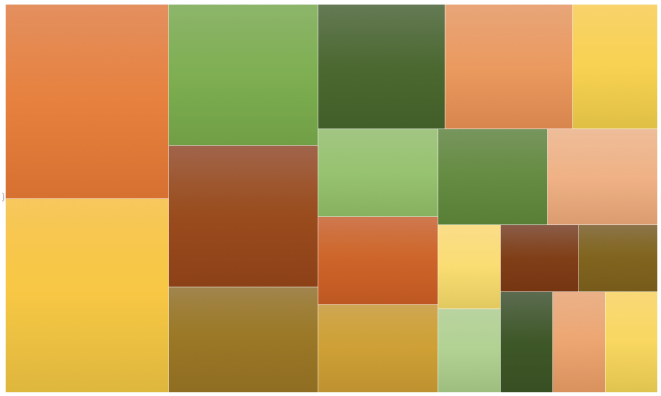


Figure 5.2: Distribution of users' contributions in D.A.Lab test process

As in every voluntary participation, it is common that most of users only tend to make infrequent contributions and log out quickly after registration. Fig5.2 illustrates the distribution of the number of annotated images completed by 21 participants in this project. Keeping a tab on user stickiness (An indicator for measuring the reusage of user's account to participate the project) of real users is worth considering. The key to user stickiness is how to give users a continuous perception of their achievement about what they are doing is helpful. To make the annotation tasks is widely-acceptable to the different participants, the work has be made relatively easy with just some simple operations, so that users may quickly tire of working through images. Collecting data is not always fun, and repeating the same thing brings about a 'diminishing marginal utility'. If there is no guarantee that the user perceives a sense of accomplishment and satisfaction from the continuous annotation behavior, or there is no good match with user's interests and attracts them as active users to continue logging in, the long-term development of the platform will be difficult and may also lead to failure in generating the desired results.

How to deal with the considerable improvement in motivation and engagement? From the users' feedback, gamification is a good option, but its accessibility of annotation task in this project is worth discussing in the further step.

5.2 Cognitive Friction

The conception of **Cognitive Friction** was put forward by Alan Cooper:

> **The resistance encountered by human intellect when it engages with a complex system of rules that change as the problem permutes.**

A bad UI always brings about frustration of a project. Eliminating cognitive friction is something need to be done. Designers should remain aware of established norms for eliminating user frustration.

This is the problem discovered from the user feedbacks which I had not considered in this version of the platform. Actually, the negative user experience is generated from cognitive friction to some degree which leads to their frustration and possibly the abandonment of the annotation tasks.

> "Tab on the data fields would be better if they went down the column rather than across the rows."

> "...I would also appreciate, for the same reason, the possibility to step from one annotation field to the next by using the navigation arrows on the keyboard rather than just the mouse (whether possible)."

Some users think that it is a bit troublesome to use the mouse for locating the input field. They prefer to locate their current input by arrow key navigation depending on their working habits. Actually, it is true. Especially for Parkinson's disease patients, manipulating the keyboard with a mouse at the same time is more complex than using only one tool. Thus, the elimination of cognitive friction and optimization on user operation are necessary to improve the user experience.

Also, some users insist that the auto-completion should be applied if they do not want to fill the blanks when the integers they need to input are '0'. It means that during the annotation process if there is no Drosophila in a certain slot of the test tube, users are prone to think that the default value which would be auto-completely filled as '0' in the input field. It is a way to minimize the requirement of annotations tasks, but whether the exclusion of some non-zero data omitted by users would have a significant impact on data quality need to be verified in the future studies.

5.3 Reward Mechanism

From the users feedbacks on Further Expectation, the Reward Mechanism is one of the aspects that needs consideration. Participants always expect to acquire a more straightforward motives and incentives which are normally not shown to them. In the future tasks, the implementation of competition events would be included in D.A.Lab,

as competitions always keep the contributors engaged [40]. A simple strategy is to qualify the reward into some cute virtual stamps which are a big motivation to Collectors [9**?**]. The collection rules could be that a participant receives a unique stamp with a different title based on the contributions they have made. For instance, the reward for each 10 successful annotation works could be represented as a new stamp with new title (e.g. 'Newbie', 'Enthusiasts', 'Super Scientist' etc.). The collection review could be displayed on their user profile pages allowing users to look at what they have achieved, with a weekly ranking on the Forum page. Besides, the impact of the competition-based reward mechanisms would be evaluated in the follow-on improvements.

5.4 Web Traffic Data

The PV(Page view) and UV(User visit), etc. statistic analysis could be implemented as two scalable indicator to all the pages for monitoring website traffic in real-time. For pages requested by a click on the website, the state of the visitor's browser is recorded and data relating to the path visitors take through the website is collected and studied.[23] It is an effective tool to understand the trend on users' behaviors and adapt the changing demands of users.

5.5 Community-based Character

Presently, the social benefit from the interaction was not represented as a satisfactory result from volunteers as the usage rate of "Feedback" and User Center was relatively lower than other functions. In general, it is significant to provide more social interaction among participants, developers and scientists. A discussion forums could be developed and divided into different blocks.

- More specialized knowledge of Parkinsons disease and drug discovery in this project.
- Regular updated on Parkinson's research and outcomes of data annotation performance in a visualized way.
- Blogs for registered account to post exchange idea, advices, comments or messages to the platform or interact with other users.

- Top 10 contributors list

It makes the fostering not just enjoyment from competition, but also a congenial and social element module.

Chapter 6

Conclusion

Overall, the D.A.Lab, is a usable web-based citizen science platform for scientific needs with technology-supported citizen participation. In the current stage, it is a proof of concept platform that allows a remote user to annotate simple behaviors of Drosophila in sample images. The result from data collection demonstrate that this project it is an accessible solution for crowdsourcing gathering with a guaranteed data quality. After evaluation, I learned the lessons from technologies(web development, project design, UI design etc.) and phenomenon of citizen science(concept of citizen science, user engagement and motivation, internal ethics of citizen science etc.) The long-term development of the platform requires more incentives and more user-friendly operation, and the strategies would be adjusted and improved from users feedback in the future test. Last but not least, not only to foster the process of scientific research, D.A.Lab, as a platform derived for enlisting people who are interested in treatment of Parkinson's disease and especially for encouraging suffered from Parkinson's disease have a direct voice in scientific research, also has more room for improvement in humanism and social responsibility. Enhancing the communication and collaboration with participants and improving their sense of meaningfulness will make the D.A.Lab to be a widely-acceptable useful platform. With this foundation in place and growing, the potential for D.A.Lab to take interesting scientific problems is nearly limitless.

Appendix A

Main code for Data Validaion

See main Code for the HTML form submission attribute and JavaScript API achieve form validation behavior in D.A.Lab

A.1 HTML form submission attribute

```
<input class="annotation" id="colunm_a1" name='a1'
       type="text" oncopy="return false"
     onpaste="return false" maxlength="2">
```

A.2 Javascript API

```
$("#submit-btn1").click(function () {
  a1 = document.getElementById("a1").value;
  ...
  f4 = document.getElementById("f4").value;
  if (a1 * 1 + a2 * 1 + a3 * 1 + a4 * 1 > 10 ||
      b1 * 1 + b2 * 1 + b3 * 1 + b4 * 1 > 10 ||
      ...
      f1 * 1 + f2 * 1 + f3 * 1 + f4 * 1 > 10 )
          { ErrorAdd(); }
   else{
      if(isNaN(a1)||isNaN(a2)||isNaN(a3)||isNaN(a4)||
         ...
```

```
        isNaN(f1)||isNaN(f2)||isNaN(f3)||isNaN(f4))
        {ErrorNum();}
    else {
        if (a1.replace(/(^\s*)|(\s*$)/g, "") == "" ||
            ...
            f4.replace(/(^\s*)|(\s*$)/g, "") == "")
            {ErrorEmpty();}
    else { UpLoadFrom();}
        }
    }
});
```

Appendix B

Survey Design

Table B.1: Questions asked in D.A.Lab survey

No.	Question
1	What's your age?
2	Have you ever participated in other Citizen Science Projects?
3	How are you affected by Parkinson's Disease?
4	Are you informed about the disease and the process and challenges involved in drug discovery?
5	Do you believe that CrowdLabbing and D.A.Lab projects can benefit people with Parkinson's in fighting the disease?
6	Regarding your visit to the D.A.Lab web pages (interface),which of the following affected your User Experience?
7	In general, how do you feel about the Interface Design?
8	How useful have you found the Tutorial Guide Information of Annotation Task?
9	On the Recording Page(Annotation Page), which functions do you find helpful?
10	If you were to have any doubt about how to continue with the recording annotation tasks,what would you refer to?
11	If you had the possibility, would you enjoy performing further tasks (for example, when we update new image files)?
12	If you answered yes to the previous question, how much time would you be willing to commit to this activity?
13	Please score the following characteristics of D.A.Lab!(1= very bad, 10=excellent)
14	What would incentivise your participation/commitment to the project and to the use of D.A.Lab?
15	Please feel free to add any other feedback.

Bibliography

[1] (2014). The web framework for perfectionists with deadlines. Website. https://www.djangoproject.com/.

[2] (2016). Worm watch lab. http://rspb.royalsocietypublishing.org/content/266/1437/2487.

[3] (2017). 6 web development stacks to try in 2017. https://webinerds.com/6-web-development-stacks-try-2017/.

[4] Alison Bryant, J., Sanders-Jackson, A., and Smallwood, A. M. K. (2006). Iming, text messaging, and adolescent social networks. *Journal of Computer-Mediated Communication*, 11(2):577–592.

[5] Amarilli, A., Amsterdamer, Y., and Milo, T. (2014). Uncertainty in crowd data sourcing under structural constraints. In Han, W.-S., Lee, M. L., Muliantara, A., Sanjaya, N. A., Thalheim, B., and Zhou, S., editors, *Database Systems for Advanced Applications*, pages 351–359, Berlin, Heidelberg. Springer Berlin Heidelberg.

[6] Amsterdamer, Y. and Milo, T. (2015). Foundations of crowd data sourcing. *SIGMOD Rec.*, 43(4):5–14.

[7] Anderson, D. P., Cobb, J., Korpela, E., Lebofsky, M., and Werthimer, D. (2002-11-01). Seti@home: an experiment in public-resource computing.(search for extraterrestrial intelligence). *Communications of the ACM*, 45(11):56(6).

[8] Arora, S., Bhattacharjee, D., Nasipuri, M., Basu, D. K., and Kundu, M. (2010). Recognition of non-compound handwritten devnagari characters using a combination of mlp and minimum edit distance. *arXiv preprint arXiv:1006.5908*.

[9] Blanchat, K. M. and Brooks, M. (2015). Collecting cats: Library lessons from neko atsume.

[10] Blanke, T. and Hedges, M. (2013). Scholarly primitives: Building institutional infrastructure for humanities e-science. *Future Generation Computer Systems*, 29(2):654 – 661. Special section: Recent advances in e-Science.

[11] Boff, K., Kaufman, L., and P. Thomas, J. (2018). Handbook of perception and human performance. volume 2. cognitive processes and performance. page 49.

[12] Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., and Shirk, J. (2009). Citizen science: A developing tool for expanding science knowledge and scientific literacy. 59(11):977–984. Exported from https://app.dimensions.ai on 2018/08/22.

[13] Bowser, A., Hansen, D., He, Y., Boston, C., Reid, M., Gunnell, L., and Preece, J. (2013). Using gamification to inspire new citizen science volunteers. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, Gamification '13, pages 18–25, New York, NY, USA. ACM.

[14] Crowston, K. and Prestopnik, N. R. (2013). Motivation and data quality in a citizen science game: A design science evaluation. In *2013 46th Hawaii International Conference on System Sciences*, pages 450–459.

[15] Danielsen, F., Mendoza, M. M., Tagtag, A., Alviola, P. A., Balete, D. S., Jensen, A. E., Enghoff, M., and Poulsen, M. K. (2007). Increasing conservation management action by involving local people in natural resource monitoring. *Ambio*, 36(7):566–570.

[16] DJ, G., E, O., and S, G. (1999). Diagnostic criteria for parkinson disease. *Archives of Neurology*, 56(1):33–39.

[17] Donley, S., McGregor, S., Wielinski, C., and Nance, M. (2018). Use and perceived effectiveness of complementary therapies in parkinson's disease. *Parkinsonism & Related Disorders*.

[18] Dunn, P. O. and Winkler, D. W. (1999). Climate change has affected the breeding date of tree swallows throughout north america. *Proceedings of the Royal Society of London B: Biological Sciences*, 266(1437):2487–2490.

[19] Follett R, S. V. (2015). An analysis of citizen science based research: Usage and publication patterns. *PLoS ONE*, DOI:10.1371.

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

[21] Ghose, R. (2002). Use of information technology for community empowerment: Transforming geographic information systems into community information systems. *Transactions in GIS*, 5(2):141–163.

[22] Gilbert, A. N., Fridlund, A. J., and Lucchina, L. A. (2016). The color of emotion: A metric for implicit color associations. *Food quality and preference*, 52:203–210.

[23] Glommen, C. and Barrelet, B. (2002). Internet website traffic flow analysis. US Patent 6,393,479.

[24] Goldstein, C. M. and Ford, W. H. (1978). The user cordial interface. *Online Review*, 2(3):269–275.

[25] Greenhill, A., Holmes, K., Woodcock, J., Lintott, C., Simmons, B. D., Graham, G., Cox, J., Oh, E. Y., and Masters, K. (2016). Playing with science: Exploring how game activity motivates users participation on an online citizen science platform. *Aslib Journal of Information Management*, 68(3):306–325.

[26] Haklay, M. (2013). *Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation*, pages 105–122. Springer Netherlands, Dordrecht.

[27] Hopkins, M. J. G. (2007). Modelling the known and unknown plant biodiversity of the amazon basin. *Journal of Biogeography*, 34(8):1400–1411.

[28] Jackson, C., Osterlund, C., Maidel, V., Crowston, K., and Mugar, G. (2016). Which way did they go?: Newcomer movement through the zooniverse. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pages 624–635, New York, NY, USA. ACM.

[29] Kurniawan, S., King, A., Evans, D., and Blenkhorn, P. (2006). Personalising web page presentation for older people. *Interacting with Computers*, 18(3):457 – 477. Human Factors in Personalised Systems and Services.

[30] Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M. J., Nichol, R. C., Szalay, A., Andreescu, D., et al. (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.

[31] Manmohan (2016). Why django framework continues to reign the python development world. Website. `http://timmurphy.org/2009/07/22/line-spacing-in-latex-documents/`.

[32] Molinillo, S., Aguilar-Illescas, R., Anaya-Snchez, R., and Vallespn-Arn, M. (2018). Exploring the impacts of interactions, social presence and emotional engagement on active collaborative learning in a social web-based environment. *Computers & Education*, 123:41 – 52.

[33] Nijkrake, M., Keus, S., Kalf, J., Sturkenboom, I., Munneke, M., Kappelle, A., and Bloem, B. (2007). Allied health care interventions and complementary therapies in parkinson's disease. *Parkinsonism & Related Disorders*, 13:S488 – S494. Proceedings of the XVII WFN World Congress on Parkinson's Disease and Related Disorders.

[34] Nov, O., Arazy, O., Lotts, K., and Naberhaus, T. (2013). Motivation-targeted personalized ui design: A novel approach to enhancing citizen science participation. In Bertelsen, O. W., Ciolfi, L., Grasso, M. A., and Papadopoulos, G. A., editors, *ECSCW 2013: Proceedings of the 13th European Conference on Computer Supported Cooperative Work, 21-25 September 2013, Paphos, Cyprus*, pages 287–297, London. Springer London.

[35] of Health, N. I. et al. (2010). Ninds parkinsons disease information page.

[36] Portas, A. M., Barnard, L., Scott, C., and Harrison, R. G. (2016). The national eclipse weather experiment: use and evaluation of a citizen science tool for schools outreach. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 374(2077).

[37] Prestopnik, N. R. (2012). Citizen science case study: Galaxy zoo/zooniverse.

[38] Prpi, J., Taeihagh, A., and Melton, J. (2012). The fundamentals of policy crowdsourcing. *Policy & Internet*, 7(3):340–361.

[39] Ravindran, A. (2015). *Django Design Patterns and Best Practices*. Packt Publishing.

[40] Reeves, N., West, P., and Simperl, E. (2018). "a game without competition is hardly a game": The impact of competitions on player activity in a human compu-

tation game. In *Proceedings of the Sixth AAAI Conference on Human Computation and Crowdsourcing (HCOMP-18)*. AAAI.

[41] Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C., Lewis, D., and Jacobs, D. W. (2012). Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, pages 217 – 226, New York, NY, USA. ACM, ACM.

[42] Shuupcom (2018). 25 of the most popular python and django websites. `http://sumo.ly/408S`.

[43] Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology & Evolution*, 24(9):467 – 471.

[44] Simpson, R. J., Page, K. R., and Roure, D. D. (2014). Zooniverse: observing the world's largest citizen science platform. In *WWW*.

[45] Stebbins, R. A. (1982). Serious leisure: A conceptual statement. *The Pacific Sociological Review*, 25(2):251–272.

[46] Sui, D., Elwood, S., and Goodchild, M. (2012). *Crowdsourcing geographic knowledge: volunteered geographic information (VGI) in theory and practice*. Springer Science & Business Media.

[47] Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., and Kelling, S. (2009). ebird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142(10):2282 – 2292.

[48] Sutherland, W. J., Pullin, A. S., Dolman, P. M., and Knight, T. M. (2004). The need for evidence-based conservation. *Trends in Ecology & Evolution*, 19(6):305 – 308.

[49] Sutton, T. and Whelan, B. M. (2017). *The Complete Color Harmony, Pantone Edition: Expert Color Information for Professional Color Results*. Rockport Publishers.

[Sveinbjornsdottir] Sveinbjornsdottir, S. The clinical symptoms of parkinson's disease. *Journal of Neurochemistry*, 139(S1):318–324.

[51] Tian, H., Stige, L. C., Cazelles, B., Kausrud, K. L., Svarverud, R., Stenseth, N. C., and Zhang, Z. (2011). Reconstruction of a 1,910-y-long locust series reveals consistent associations with climate fluctuations in china. *Proceedings of the National Academy of Sciences*, 108(35):14521–14526.

[52] Tinati, R., Simperl, E., Luczak-Roesch, M., Van Kleek, M., and Shadbolt, N. (2014). Collective intelligence in citizen science–a study of performers and talkers. *arXiv preprint arXiv:1406.7551*.

[53] Tinati, R., Van Kleek, M., Simperl, E., Luczak-Rösch, M., Simpson, R., and Shadbolt, N. (2015). Designing for citizen data analysis: A cross-sectional case study of a multi-domain citizen science platform. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 4069–4078, New York, NY, USA. ACM.

[54] Vainikka, J. (2018). Full-stack web development using django rest framework and react.

[55] Wang, L., Liu, W., Jing, N., and Mao, X. (2015). Simultaneous navigation and pathway mapping with participating sensing. *Wireless Networks*, 21(8):2727–2745.

[56] Wiggins, A. and Crowston, K. (2011). From conservation to crowdsourcing: A typology of citizen science. In *System Sciences (HICSS), 2011 44th Hawaii international conference on*, pages 1–10. IEEE.

[57] Wiggins, A. and He, Y. (2016). Community-based data validation practices in citizen science. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pages 1548–1559, New York, NY, USA. ACM.

[58] Wilderman, C. and Shirk, J. (2010). From citizen science to volunteer monitoring: seeking hybridization of models for community science.

[59] Woodcock, J. and Greenhill, A. (2016). Crowdsourcing citizen science: exploring the tensions between paid professionals and users.

[60] Wright, D. R., Underhill, L. G., Keene, M., and Knight, A. T. (2015). Understanding the motivations and satisfactions of volunteers to improve the effectiveness of citizen science programs. *Society & Natural Resources*, 28(9):1013–1029.

[61] ltekin, A., Heil, B., Garlandini, S., and Fabrikant, S. I. (2009). Evaluating the effectiveness of interactive map interface designs: A case study integrating usability metrics with eye-movement analysis. *Cartography and Geographic Information Science*, 36(1):5–17.