



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

INSTITUT FÜR INFORMATIK
LEHR- UND FORSCHUNGSEINHEIT FÜR
PROGRAMMIER- UND MODELLIERUNGSSPRACHEN



Conception and Implementation of a Collaborative Data Science Platform

Martin Bogner

Masterarbeit

Beginn der Arbeit: 25.2.2016
Abgabe der Arbeit: 8.9.2016
Betreuer: Prof. Dr. François Bry
Dr. Clemens Schefels

Erklärung

Hiermit versichere ich, dass ich diese Masterarbeit selbständig verfasst habe. Ich habe dazu keine anderen als die angegebenen Quellen und Hilfsmittel verwendet.

München, den 08.09.2016

Martin Bogner

Zusammenfassung

Vorliegende Arbeit entwirft und implementiert eine kollaborative Citizen Science Plattform zur quantitativen Datenanalyse im Anwendungsbereich Kunstgeschichtsforschung. Aufgrund ihrer quantitativen Herangehensweise an Kunstgeschichte ist die Plattform von Grund auf interdisziplinär zwischen der Kunstgeschichte und Data Science angesiedelt. Als Datengrundlage dient die durch das ARTigo-Projekt generierte Schlagwörterdatenbank für Kunstwerke, wobei bereits existierende Datenanalyse-Tools in die Plattform integriert werden. Die Förderung sinnvoller Zusammenarbeit durch hierarchisch geleitete Forschungsprojekte, das Motivieren freiwilliger Beteiligung durch die Anwendung von Methoden aus sowohl Motivationstheorie als auch aus dem Bereich der Gamification und auch die Ermöglichung einfacher und flexibler Integration von Datenanalysetools spielen als die drei Hauptanforderungen, die an die neue Plattform gestellt werden, eine zentrale Rolle in dieser Arbeit.

Abstract

This thesis designs and implements a collaborative citizen science platform for quantitative data analysis applied to art history research. Because of its quantitative data science approach to art history, the platform is in its core interdisciplinary in nature, bridging between art history and data science. As underlying dataset serves the keyword-based artwork database generated by the ARTigo project, integrating preexisting data analysis tools into the platform. Facilitating meaningful collaboration through hierarchical led research projects, motivating volunteer participation by utilizing methods from both motivation theory and gamification as well as allowing easy and flexible integration of data analysis tools into the platform are the three main requirements for the new platform and are addressed in the course of this thesis.

Acknowledgements

I am using this opportunity to express my gratitude to everyone who supported me throughout the course of this thesis. I am thankful for their aspiring guidance, invaluable constructive criticism and friendly advice.

First of all, I am deeply grateful to *Dr. Clemens Schefels* for his outstanding supervision of this thesis. He always found time in case of arising questions or problems and greatly helped me with his feedback for new ideas.

Also, I give my deepest thanks to *Prof. Dr. François Bry* for his continuous enthusiasm and support, and for our inspirational meetings together with *Dr. Schefels*. Especially the long and fruitful discussions on a greater vision for the platform, beyond just providing a hub for data analysis tools, gave this thesis its current direction.

The numerous ideas suggested from both *Prof. Bry* and *Dr. Schefels* during these meetings have shaped the character of the platform developed for this thesis decisively, for example by limiting the platform's focus to the European art of the long 19th century or by empowering citizen scientists to initiate research goals on their own.

Furthermore, I would like to thank all past, present and future members of the Teaching and Research Unit Programming and Modelling Languages who have in the past, currently do or will work on the ARTigo project, not to forget the ARTigo players themselves. Without them, this thesis would have been literally unthinkable.

My thanks also goes to my fellow student *Giulianna Dehn*, whose upcoming Bachelor's thesis continues the work on the platform presented here, for her inspirations and ideas for further improvement.

Finally, my deep and sincere gratitude goes to my girlfriend *Sabrina*, who profoundly supported me throughout this thesis, endured my technical ramblings and has helped me on innumerable levels along the way. Thank you.

Contents

1. Introduction	9
2. Related Work	11
2.1. Public Participation in Scientific Research	11
2.1.1. History and Overview	11
2.1.2. Disambiguation	12
2.1.3. Typology	14
2.2. User Motivation and Engagement	15
2.2.1. Attention and Engagement Economics	15
2.2.2. Motivation Types	16
2.2.3. Research on Motivation in Citizen Science Projects	17
2.3. Gamification in Crowdsourcing and Online Citizen Science Environments	20
2.4. Quality of Citizen Science Data	23
2.5. The ARTigo Project	24
2.5.1. ARTigo Gaming Ecosystem	24
2.5.2. ARTigo Artwork Search	26
2.5.3. ARTigo Analytics Center	28
3. Conceptualizing the Platform	31
3.1. Scope and Language	31
3.2. Target Audience	33
3.3. Research Questions	33
3.4. Engaging the Citizen Scientists	35
3.4.1. Attracting New Volunteers	35
3.4.2. Engaging Existing Users	36
3.5. User Collaboration	39
3.5.1. Structure of Research Topics	39
3.5.2. Creating New Projects	41
3.5.3. Permission Management and User Hierarchy	42
3.5.4. Collaborative Work on Research Questions	43
3.6. Publication of Contributions	46
3.7. Integration of Data Analysis Tools	46
4. Implementation	48
4.1. Architecture of the Platform	48
4.1.1. Adapting a Prebuilt System for the Platform	48
4.1.2. Introduction to the Wordpress Content Management System	49
4.1.3. Extending Wordpress	50
4.2. Collaboration	51
4.2.1. Projects and Community	52
4.2.2. Article-Based Contributions	53
4.2.3. Requesting Assistance	55
4.2.4. Commenting	57
4.2.5. Project Suggestions and Meta-Discussions	58

4.3.	Integration of Data Analysis Tools	58
4.3.1.	Communication Between Tools and the Citizen Science Platform	58
4.3.2.	Changes to ARTigo Search and Analytics Center	60
4.4.	User Motivation and Engagement	61
4.4.1.	User Profile Area	61
4.4.2.	Contribution Rating System	61
4.4.3.	Try-out Area for Unregistered Visitors	63
4.4.4.	Home Page	64
5.	Outlook and Conclusion	66
5.1.	Conclusion	66
5.2.	Outlook	66
5.2.1.	User Study	66
5.2.2.	Improvements to Existing and Addition of New Data Analysis Tools	67
5.2.3.	Additional Features for the Platform	67
5.2.4.	Adapting the Platform for Other Citizen Science Purposes	68
5.2.5.	Upcoming Usage in Teaching	68
A.	Appendix	69
A.1.	Adapting the Citizen Science Platform for Other Applications	69
A.2.	Maintenance and Development Instructions for the Citizen Science Platform	69
A.2.1.	Prerequisites for Server / Development Environment	69
A.2.2.	Common Issues after Updating Wordpress or its Plugins	69
A.2.3.	Changing the Content of the Home Page	70
A.2.4.	Adding a Group with Mandatory Membership	70
A.2.5.	Moving the Platform to a Different Server and URL	70
A.3.	Adding, Moving and Removing Data Analysis Tools	70
A.4.	Used Wordpress Plugins and Their Functionality	71
A.5.	Amount of Source Code Written	73
A.5.1.	Citizen Science Platform	73
A.5.2.	ARTigo Search	73
A.5.3.	Analytics Center	74
	References	75

1. Introduction

The fine arts, be it painting, music or poetry, are seen by many as the pinnacle of humanity's passion and creative power, surpassing any attempt to measure or quantify them. Art and art history, in turn, are two of the great academic disciplines in the humanities. In contrast, the sciences are seen as disciplines of cold logic and numbers, trying to discover universal laws and principles free of any room for interpretation.

This long lived view, popularized among others by German philosopher Wilhelm Windelband [Win94] or physicist and novelist C.P. Snow [Sno59], sees a distinct and incompatible separation between both academic fields, both in language as well as in methods employed. At the same time, much efforts are currently undertaken to reconcile this separation, for example in the field of digital humanities.

A similar rift has been present for the past century between professional scientists on the one side and the general public on the other. Though there is a long tradition of amateur scientists making groundbreaking contributions to science, today, they are mostly marginalized, often seen as cheap manpower at best.

One attempt in breaking these strict separations is the ARTigo project,¹ an interdisciplinary cooperation between the Institute of Art History, the Center for Digital Humanities, the Center for Information and Language Processing and the Institute of Informatics at the University of Munich, a combined effort to create a keyword-based semantic search engine for artworks. As the generation of semantic keywords or tags for artworks is not feasible with today's image processing algorithms, they have to be created by humans, a generally expensive and time consuming process. ARTigo, following a human computation and gamification approach, generates the required keywords for its artworks not by paid labor, but by motivating volunteers to play games with a purpose where the tag creation is embedded in engaging and enjoyable gameplay.

Thinking beyond using the gathered data solely to power the search engine, it can also be utilized for further research in art history, quantitative in nature, via data analysis. First tools, applying well known data science algorithms to the tag data created by ARTigo, have already been implemented, but lack accessibility for art historians or art-interested amateurs untrained in the methods of statistics and data science. Statisticians and computer scientists, however, do possess the essential technical understanding, but generally do not exhibit in-depth knowledge of art history.

Therefore, the idea for a shared platform was born, where professional scientists and amateurs from both disciplines alike can work collaboratively, in order to examine research questions from art history through the perspective of data science. As neither exclusive knowledge of data science or art history is sufficient for the task, people from both disciplines have to actively work together and complement each other's skills. By providing an opportunity for interdisciplinary collaboration and learning from each other along the way, the platform tries to do its part in overcoming the gap between sciences and humanities. As amateur volunteers have proven their dedication and usefulness in the ARTigo project, creating millions of tags for the artworks so far, their potential is harnessed for the new

¹<http://www.artigo.org>

1 INTRODUCTION

platform as well. The platform enables them to participate actively in most steps of the scientific process, following the principles of citizen science.

Scope of this Thesis

This master's thesis' objective is to conceptualize and implement a collaborative citizen science platform for data analysis, based on the data gained through the ARTigo project. As the aim is on creating a system ready for productive use, practical considerations take priority for the implementation part, with a strong focus on utilizing existing software, both from the ARTigo project and external sources, in the process.

This thesis is organized as follows: After this introduction, Chapter 2 will give an outline on related scientific research, with focus on public participation in scientific research, user motivation and gamification as well as provide an introduction to the ARTigo project.

Afterwards, Chapter 3 will present a general overview of the platform and elaborate certain design decisions and concepts employed in order to attract and engage volunteer users, facilitate a productive collaboration environment, keep administrative effort low while preventing abuse at the same time and integrate existing data analysis tools into the platform.

Finally, Chapter 5 provides a conclusion and outlook, also proposing future additions and improvements as well as additional applications for the platform.

2. Related Work

2.1. Public Participation in Scientific Research

2.1.1. History and Overview

Today, most science takes place at universities or academic research facilities and is done by paid, professional scientists as their main occupation. Contributions by non-scientists are in contrast often seen with aversion and regarded as of inferior quality [LW79, RP13].

While this type of science has been dominant in the last century and is now regarded as “traditional science” [FS14], this has not always been the case. Before the advent of predecessors of today’s research-focused universities in the course of the 19th century, academic teaching and research at most universities was stagnant and extremely conservative [Pal07]. This enabled people who would be regarded as amateurs today to make significant and fundamental contributions to science, their income depending on either rich patrons, personal wealth or work in other professions. Charles Darwin, for example, underwent his famous trip to the Galapagos Islands as a self-funded companion to the survey expedition, not as a paid scientist [Sil09].

Such solitary “gentleman scientists” have made considerable contributions to science, but doing science in such a way requires a deep understanding of the research subject and high amounts of time and money from one person. It excludes the majority of people from the scientific process who are interested in science, but lack the preexisting knowledge or resources.

As traditional academic science often lacks the funding or manpower to undertake labor-intensive large-scale monitoring or classification tasks, the scientific community searched for ways to make members of the general public participate in scientific research. For instance, without the help of amateur astronomers and their organizations, observing the full area of the night sky would not be feasible with the relatively low number of astronomical observatories available to professional astronomers [Bol00].

In order to allow people of heterogeneous backgrounds to participate in scientific research, collaborative citizen science projects have emerged, where amateurs under the guidance and leadership of professional scientists collaboratively work on a common research project.

While the term *citizen science* was only recently introduced in the 1990s, the basic concept has already been adopted in the beginning of the 20th century in research fields like phenology, weather observation and astronomy [BBJ⁺09].

One early example for such a project is the annual *Christmas Bird Count* of the *National Audubon Society* in the USA, first run in 1900 and still active, with thousands of participants in recent years [Sil09, Nat]. The bird census itself is done by volunteers, the resulting data then used by professional scientists to estimate the health and state of local bird populations.

Citizen science is therefore by no means a recent phenomenon, but the advent of the Internet and related technologies sparked an unprecedented rise for citizen science. For “classical” citizen science projects, which often require participants to gather data in the field (like the mentioned *Christmas Bird Count*), the Internet makes it easier to appeal to a large group

2 RELATED WORK

of potential volunteers, simplifies communication and aids in the aggregation and quality control of data [NWC⁺12].

Also, previously limiting factors for citizen science projects, such as the confinement of research to physical places or the general feasibility of only low-scale collaboration in small teams are no longer valid [WC11]: For many Internet-based citizen science projects, Internet access and spare time are the only requirements, enabling participants from all over the world to collaborate on a large scale.

An example is the web-based citizen science project *Galaxy Zoo*:² Volunteer citizen scientists can participate by simply visiting the project’s website. After a short introduction, users are presented with telescope photographs of the night sky and can start classifying galaxies by their morphology. The gathered data is then in turn further analyzed by professional scientists and has been the basis for many scientific publications [FS14]. Other projects span from transcribing Egyptian papyri in *Ancient Lives*³ to optimizing the molecular structure of proteins in *Foldit*,⁴ covering a large variety of topics.

2.1.2. Disambiguation

The variety of projects incorporating similar or related public collaboration models for science makes it difficult to subsume them under a single term. For example, what is now commonly referred to as “citizen science” has also been referred in one of its variations as “crowd science”, “networked science”[FS14] or “massively-collaborative science”[FS14, Bra13]. A common term describing any kind of public participation in scientific research has yet to emerge and may only yield low informative value given the variety of projects already grouped under the label of citizen science.

While the listing below gives a summary of the most common terms used when referring to collaborative science with public participation, it makes no claim to completeness. The boundaries between some of the terms are often ambiguous, so projects may match several at once.

Citizen Science as a term was introduced in the mid-1990s by Rick Bonney and Alan Irwin independently from each other, a generally accepted definition of what exactly comprises citizen science has not yet been achieved because of the great variety of self-proclaimed citizen science projects [RP13]. In recent years, the term citizen science has evolved into the direction of an umbrella label describing public participation in scientific research in general, making an exact distinction difficult.

In comparison to other collaborative models, projects labeled citizen science are mostly characterized by a hierarchical power structure between the professional scientists and contributing volunteers [WC11, BBJ⁺09]. Typically, scientists formulate the research topic and invite volunteers to contribute to the project. Tasks executed by the volunteers may be of varying complexity depending on the project, but involve actual scientific work in contrast to just providing resources or participating as a test subject.

²<http://www.galaxyzoo.org>

³<http://www.ancientlives.org>

⁴<http://fold.it>

Citizen science itself does not provide any assumption on the technologies used, many projects are limited to local communities and are conducted offline.

Commons-based peer production has been originally coined by Yochai Benkler [BN06]. He describes it as a model of social production, characterized by its decentralized and non-proprietary nature, enabled by the networked environment of the Internet [Ben06]. Its principles are not limited to a specific kind of work, they can be applied to scientific research, open-source software development or text production alike.

Compared to citizen science, commons-based peer production has no central authority, its organization is strictly bottom-up by equal *peers*. *Commons-based* refers to its second founding principle: Resources are publicly available and outputs have to be shared freely or under certain conditions (e.g., under a copyleft license). One prime example for commons-based peer production is *Wikipedia*: All content is user generated and openly available under a Creative Commons license. The associated *Wikimedia Foundation* only provides the necessary infrastructure but does not interfere with the content production itself.

Crowdsourcing as introduced by the Jeff Howe [How06] in Wired magazine in 2006, is often used as a generic term for projects involving distributed production models with open participation [WC11]. This definition would include open-source software development like *Mozilla Firefox* or commons-based peer production.

Another definition offers a more differentiated view on crowdsourcing, focusing on its unique traits compared to other collaborative models [Bra13, EAGLdG12]: Here, crowdsourcing projects are characterized by a top-down management where an organization asks for participation on a task via an open call to a large, heterogeneous group of volunteers. The participation itself is defined as an online activity and being mutually beneficial for both the organization and the volunteers. The term crowdsourcing does not make any implications on the nature of the tasks, both commercial as well as scientific projects can be described by it.

Crowd science can be thought of as applying the principles of crowdsourcing on citizen science [Daw12, FS14, You10] and is in some publications also called **online** or **digital citizen science** [NAA11b, NAA11a]. Compared to other forms of online participation on scientific projects, crowd science is characterized by the active engagement of the volunteers in the scientific work itself, not just by providing resources for distributed-computing [Bra13]). One often quoted example for this kind of projects is the aforementioned *Galaxy Zoo* project.

Citizen Cyberscience coined by François Grey, the term describes any type of web-based citizen science [Gre09, Hak13]. In contrast to most other definitions of citizen science, it also includes *volunteer computing* as one of its three subcategories. A prominent example is the project *SETI@Home*,⁵ where large datasets from radio telescopes are distributed among PCs of volunteers and automatically processed via algorithms in the search for extra-terrestrial life [Ber]. The other two subcategories include *volunteer thinking*, where participants contribute their time and mental “resources” to manually

⁵<http://setiathome.berkeley.edu>

2 RELATED WORK

classifying data, and *participatory sensing*, which utilizes the sensors of volunteers' mobile phones to collect data of their surroundings.

2.1.3. Typology

As seen in the previous chapter, scientific projects with public participation can take manifold forms. It is therefore reasonable to develop taxonomies for projects employing some of the mentioned concepts. The following section will summarize common frameworks for describing the characteristics of collaborative science projects.

Scientific research is no monolithic activity, most of the time it can be broken down into distinct steps, among them formulating a research question, developing hypotheses or collecting, analyzing or interpreting data. Depending on the involvement of the public in each step, Bonney et al. [BBJ⁺09] categorize public participation in scientific research into three major types of projects: *Contributory*, *collaborative* and *co-created*. In this context, the “public” consists of anyone participating in a project except for the professional researchers in charge.

Contributory projects are exclusively researcher-driven. These projects generally require the retrieval of large amounts of data, which would be too time consuming and expensive to be generated by the researchers themselves. Citizen scientists collect this data according to predefined protocols. Most projects labeled as citizen science fall into this category.

Collaborative projects allow participants, additional to data collection tasks, also to assist the professional scientists in the research process by analyzing the data and refining the project design.

Co-created projects involve citizen scientists in most or all steps of the scientific process. Often the research question itself is proposed by members of the public and the project has a bottom-up organization.

Based on this categorization, Shirk et al. [SBW⁺12] propose two additional categories on the edges of the spectrum:

Contractional projects are started by a community which commissions professional scientists to perform research on a specific topic and then report their findings back to the community. The scientific work itself is done exclusively by professionals. Unlike research initiated solely by scientists, which often tends to be of an academic and theoretical nature, this model directly orients research towards practical questions and issues relevant to the community.

Collegial projects are conducted independently by amateur scientist, much like in the days of the “gentleman scientist”. Collaboration between amateurs and professional scientists is limited to the final stages of the scientific process, i.e., peer review and publication.

Another typology for citizen science has been conceived by Wiggins and Crowston, focusing on organizational and macrostructural properties of projects [WC11]. They group citizen

science projects into five mutually exclusive categories, based on a project's primary goal and the degree participants have to interact with the physical environment:

Action Projects of this type focus on a local community and try to encourage the public in participating in local concerns. The scientific research is aimed at supporting a specific agenda, for example protecting local water quality; hence the term *action*. They are commonly organized bottom-up by the affected community, with the scientists as consultants rather than initiators. Often these projects are labeled *community-based participatory research* [WBI05] or *participatory action research* [WC11].

Conservation Primarily ecological and strongly regional focused, these projects collect data for long-term monitoring and resource management. In comparison to action-oriented projects, they are typically organized top-down by scientists or rely on public funding and organization.

Investigation The driving motivation in establishing an investigation project is the need for data by scientists: One or multiple research goals require gathering of data from the physical environment. This category encompasses most of the “classical” citizen science.

Virtual These kinds of projects share their motivation with the investigation type projects, but in comparison, the citizen scientists' activities lack any physical elements. Instead, any interaction is computer-mediated, which makes this kind of projects synonymous to *crowd science*.

Education Rather than focusing on research, the primary objective of education type projects is outreach and public education. Tasks for participants are typically designed in a way that supports a gradual learning experience.

2.2. User Motivation and Engagement

At the core of each citizen science project stand the participating volunteers. It is of critical importance for a successful citizen science project to gain a large enough group of volunteers who contribute regularly to the project. In order to achieve this, one must understand the motives behind participation in citizen science both for the volunteers as well as the professional scientists. A deeper understanding of these motivations can help designing an environment that caters to the need of both groups, making the collaboration more successful and sustainable. The following chapter will give an overview of user engagement and motivation in general as well as of current research about motivation in citizen science in particular.

2.2.1. Attention and Engagement Economics

Every project which relies on volunteer participation needs to address two basic questions, irrespective of its commercial, scientific or recreational background: First, how can one attract the attention of a sufficiently large group of potential volunteers and motivate them to join in the first place and second, how to keep them engaged with the project and motivate continuous contributions for a long period of time.

2 RELATED WORK

Many crowdsourcing or citizen science projects fail to achieve these two objectives and never grow to a sufficiently large user base or get abandoned in a short time. The Internet is full of examples of failed calls for crowd participation [McG11, RHP⁺14].

Since the early days of the Web, this fact has spawned numerous research. According to Michael Goldhaber [Gol97], the Internet can be seen as an **attention economy**. In this type of economy, attention is seen as a scarce resource and becomes the primary currency. The fundamental assumption of this model: Any person has only a limited amount of attention to give. As there is much more information and content on the Internet than a single person could process, attention becomes the limiting factor. Platforms depending on a sufficient user base have to compete against each other for the users' attention.

Adapting this model to online collaborative projects, Jane McGonigal [McG08] devised the idea of an **engagement economy**. While at first glance similar to attention economy, engagement economy also takes the user's level of involvement into account and makes a clear distinction between a crowd's attention, which is more passive by its nature (e.g., rating a proposal), and active engagement (e.g., submitting said proposal). McGonigal argues that attention alone won't make a collaborative platform successful. For example, the early crowdsourcing website *Cambrian House* failed after two years, despite good media coverage and acquiring over 50,000 members, because members remained mostly passive instead of participating actively [McG08].

According to McGonigal, any project relying on voluntary participation competes for each user's brain cycles or *participatory bandwidth*, "our individual and collective capacity to contribute to one or more participatory networks" [McG11].

With the abundance of calls for participation and collaboration on the Internet, a successful crowdsourcing project must secure as much participation bandwidth as possible, often at the expense of similar other projects. In a voluntary setting, this can only be accomplished by supplying an environment that is as motivating and rewarding as possible.

2.2.2. Motivation Types

Motivation is the driving force behind most human actions. Work that does not cater a person's motivation is unlikely to be done voluntarily, motivating work on the contrary provides pleasure to anyone just by doing it. In order to gain an active user base for citizen science projects, it is therefore critical to understand the different kinds of motivation and how they can be kindled and sustained. Modern psychology distinguishes motivation between *extrinsic* and *intrinsic* motivation:

Extrinsic motivation originates outside of ourselves. It applies to activities that are done in order to achieve external rewards like money, social status, self-use or career benefits. Most citizen science projects cannot offer extrinsic rewards for volunteers [SF14]: Overall, there is no self-use for the volunteers' contributions, there is not enough funding to pay the volunteers, and peer recognition is marginal because of the low visibility to the general public of an individual's contributions.

Intrinsic motivation in contrast has its origin inside of ourselves, it is defined as "the doing of an activity for its inherent satisfactions rather than for some separable consequence"

[RD00]. It is deeply emotional because doing work perceived as pleasurable or satisfying can motivate people to continue doing it despite the lack of any external rewards.

Further subclassifying intrinsic motivation, McGonigal [McG11] gives four general categories for intrinsic rewards that contribute to a person’s happiness:

Satisfaction People crave for satisfying work. While the exact outlines of what is perceived as “satisfying” vary from person to person, it is characterized as “clearly defined, demanding activities” that provide both a clear goal and actionable next step towards it, while one’s efforts result in directly visible effects.

Success The pursuit of the experience or prospect of being successful also factors into a person’s motivation. People in general want to feel powerful and amount to something. This includes the feeling of improvement over time and getting recognition for one’s skills and achievements.

Social Connection Humans, being social creatures, seek social connection. Social interactions like building bonds with other people or sharing experiences with them are basic human needs.

Meaning Lastly, people desire meaning, the feeling of being part of something meaningful and larger than oneself. Contributing to something of lasting significance can provide pleasure by itself.

Psychological research suggests that achieving these intrinsic goals is more sustainable and contributes more to a person’s general happiness than extrinsic motivation [KR96, LSS05, McG11]. Appealing to the intrinsic motivations of its users should therefore be a central focus for any collaborative venture, let alone citizen science projects.

2.2.3. Research on Motivation in Citizen Science Projects

While this chapter has provided a general overview on motivation so far, the following paragraphs will describe existing research on motivation specifically in the context of citizen science.

In a study in 2012, Rotman et al. [RPH⁺12] examined the motivation of both professional scientists and volunteers in citizen science projects and developed a motivational framework. Based on the results of an online survey and qualitative interviews with volunteers and involved scientists for several citizen science projects, they identified four types of motivation in collaborative science: **Egoism** (gaining benefit for oneself), **altruism** (increasing the welfare of others), **collectivism** (increasing the welfare of one’s in-group) and **principism** (upholding one’s important principles, e.g., justice).

While volunteers deemed each category similarly motivating, scientists rated altruism and principism the highest, followed by egoism. Collectivism was valued significantly less by scientists than by volunteers.

Another finding was that a volunteer’s initial interest in a project is mostly rooted in egoism, which typically stems from one or more of the following reasons: Personal curiosity about the research topic, previous engagement in scientific projects, preexisting hobby related to methods employed by the project (e.g., a photographer partaking in an insect documentation

2 RELATED WORK

project) or gaining experience for a future career in science. This is also supported by findings in user studies on other citizen science projects [RBG⁺13, RBG⁺10] which show that user motivation for participation often stems from the researched subject or the means or methods employed in the project.

In a later study, Rotman et al. [RHP⁺14] verified that the dominant motivations of volunteers differ between short-term participation (i.e., joining a project and initial contributions) and long-term participation for a prolonged period. They report that short-term participation is driven mainly by egoism, while motivation for long-term participation is more complex, factoring in other motivational types as well, and harder to effectuate. Additional to self-centered motivation, social interaction inside the projects, especially between scientists and volunteers, was identified to have a positive effect on long-term motivation.

Based on their initial findings, they developed a dynamic model for the engagement cycle of volunteers [RPH⁺12], i.e., which motivations come into play at different stages of the volunteers' engagement with a citizen science project, illustrated in Figure 1. According to the model, volunteers and scientists alike are mainly motivated by egoism at the initial stages of a project, the volunteers seeking opportunities to satisfy their personal interest, the scientists needing large-scale data for their research. This synergy between both groups facilitates the initial collaboration. Recognition by scientists and other contributors plays an important role in keeping the volunteers motivated after the first involvement. A pivotal point in the participation is at the end of a task: After a volunteer finishes a task, he reassesses his involvement in the project with collectivism and altruism now as additional motivational factors. Depending on the outcome, he may pick up a new task or leave the project.

Derived from this model, they propose several implications for designing collaborative tools, like the web platform for a citizen science project: The design should help identify points in a project's life cycle where participation is low or can decline (e.g., project initialization, the end of a task) and allow for responding with proper motivational measures. Usage of volunteers' contributions, for example in scientific publications, should be highlighted and volunteers notified about it to address their desire for recognition. Also, massive tasks should be broken down into smaller subtasks in order to provide starters with more tangible entry points to the project.

Nov et al. [NAA11a, NAA11b, NAA14] examined in several studies the relevance of different types of motivation to the actual contributions made by volunteers in citizen science projects. In their proposed motivational model (see Figure 2), different classes of motivation each influence the **participation intention** of the volunteers, which in turn correlates to the amount of actual contributions made.

They included five classes of motivation for volunteer participation in their study: **Collective motives** (level of importance attributed to the project's objectives), **norm-oriented motives** (expectations regarding the reaction of others deemed important, e.g., family, friends), **reward motives** (benefits of participating, e.g., increased reputation or social interaction), **identification** with the project and **intrinsic motives** (enjoyment associated with the participation).

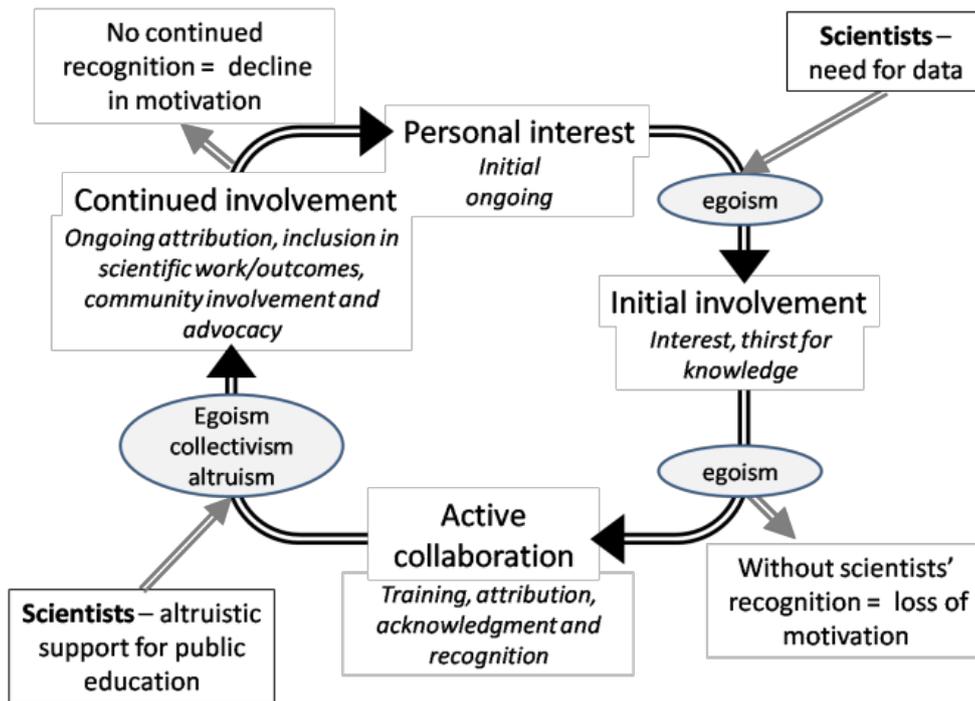


Figure 1: Process model of volunteer and scientist involvement in citizen science projects. [RPH⁺12]

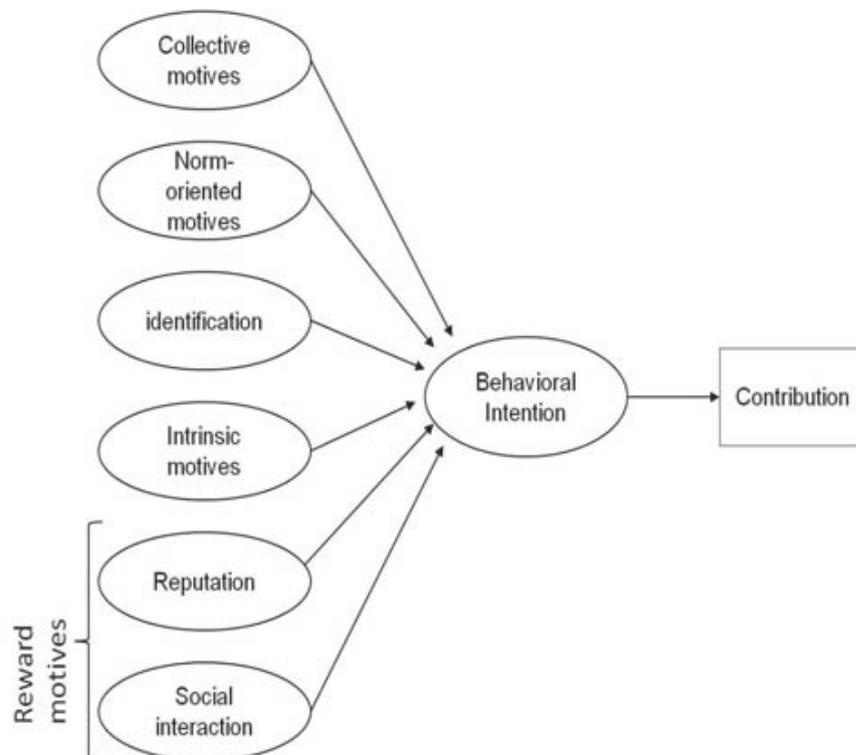


Figure 2: Motivation-intention-contribution framework of volunteer participation [NAA11b]

2 RELATED WORK

Examining the importance of each of the above classes, they identified intrinsic motives as the most relevant for contribution quantity, while its influence on the contribution quality was insignificant if not detrimental [NAA14]. According to them, enjoyable contribution mechanisms are consequently a highly effective way of keeping volunteers motivated, but additional mechanisms are needed to ensure contribution quality.

For all other motivations, only a moderate relevance to the contribution quantity was ascertained. A notable observation was made regarding collective motives: Though volunteers rated collective motives as their most important motivation, the actual correlation between collective motives and contribution quantity were only moderate. This finding has implications beyond this study, as it shows the limitations of similar research that solely relies on user questionnaires, like [RBG⁺10] or [RBG⁺13]. Nov et al. assume that the relevance of collective motives, though initially dominant when joining a project, fades over time. In their most recent study however, they identified collective motives (and to a lesser extend reputation) as the main drivers for high quality contributions [NAA14].

2.3. Gamification in Crowdsourcing and Online Citizen Science Environments

Gamification in general is an informal umbrella term for using selected “elements of game-design in a non-gaming context” [DDKN11]. Gamification pursues primarily two goals: Enhancing the user experience and increasing the user engagement. It is therefore a promising way to kindle and fuel the users’ motivation. Detering et al. [DDKN11] show, that while the term itself has only seen broader adoption starting 2010, underlying or related principles have been around much longer, predating the use of the Internet and computers. Gamification should not be viewed singularly, it is enclosed in the broader context of “ludification of culture” [Rae06], with related concepts like *serious games* (full-fledged games with a non-entertainment background) or *pervasive games* (games where the gaming experience is extended out in the real world [BML05], for example *augmented reality games* [McG11]).

Gamification does not necessarily aim to achieve a fully gamified experience. It is also possible to apply it only to certain aspects of a platform. In fact, many crowdsourcing or citizen science projects employ certain concepts also found in gamification, without labeling or perceiving themselves as “gamified”. For example, McGonigal [McG11] argues that the crowdsourced encyclopedia *Wikipedia* shares many characteristics with well-designed multiplayer games: First, according to her, the diversity and vastness of *Wikipedia* offers a “good game world” that encourages curiosity and exploration. Second, it features “good game mechanics”, defined by direct feedback for user actions (visible results after editing an article), wide choice of work opportunities of variable difficulty, and personal progression (number of contributions so far and the possibility to rise to privileged positions inside the organization). Last, it has a “good game community” that provides positive social interaction and a meaningful context for the collaborative effort. Her conclusion: Following these game-like structures when implementing a collaborative project improves its chances of success.

A study by Iacovides et al. [IJCTC13] points out some limitations of gamification in the citizen science context, though: Their results suggest that gamification of citizen science projects has little effect on attracting new gamers, rather it aids in motivating existing users and sustain their engagement. Users are typically not attracted initially by the gaming

aspect, they participate rather because of their interest in the research topic or citizen science in general. Other research by Bowser et al. [BHP13] even suggests that gamifying an experience too much may be in fact detrimental to the motivation of fully-pledged citizen scientists or viewed as distracting by them.

The following listing shows the most common approaches and concepts regarding gamification in online citizen science platforms:

Ranking Systems and Competition Eveleigh et al. [EJLC13] applied concepts of gamification in the citizen science project *Old Weather*,⁶ where users are asked to transcribe old handwritten ship logs with focus on past weather events. As the task proved repetitive and difficult, they introduced a ranking system to the project in order to sustain motivation and participation. They opted against a global leaderboard, instead partitioning contributors into teams or “crews” for the logs of each respective vessel. They also implemented three ranks for users depending on their amount of contributions, starting with “cadet”, progressing to “lieutenant” after a certain amount of contributions, finally leading to “captain”, the top contributor of a single ship.

Studying the effects the implemented ranking system had on contributors, they received mixed result: On the positive end, users reported the ranking system to provide validation, feeling appreciated for their past contributions. Also, it allowed users to track their personal progress and provided a challenging competition for the captain’s position. On the other hand, competitive aspects may lead to negative effects: First, they often incentivize low-effort, low-quality contributions. Additionally, high top scores may be viewed as unachievable by new volunteers, thus deterring them from participation. Lastly, the constant effort needed to keep a top position in the ranking may prove stressful and exhausting in the long run, ultimately leading to attrition. Their conclusion was to provide personal milestone targets to offset the negative aspects of competition and to provide more finely graduated progression levels for additional positive feedback.

Badges and Achievements Bowser et al. [BHH⁺13] implemented a badge and ranking system to gamify their citizen science mobile application *Biotracker*, where users tag and track single plants on a specific location (e.g, a maple tree in front of the university entrance) and take pictures of its different phenological states (e.g., budding or full flowering). Upon achieving the conditions for a badge, it is displayed on the user’s profile page. They created two general types of badges: Fixed-goal badges, which are awarded after performing a specific activity a certain amount of times, and competition-based badges, awarded when reaching the top spot on the leaderboard. Assessing the effects of the gamification elements in their application, Bowser et al. found that gamification can inspire new users, especially digital natives, who would not find the traditional motivations for citizen science rewarding.

However, the usage of ranking systems or badges can also have adverse effects on long-term motivation. The game design Professor Scott Nicholson [Nic15, Nic12] subsumes the introduction of badges, levels, leaderboards, achievements, points or ranking systems under the term **reward-based gamification**. He warns that an indiscriminate usage of them,

⁶<http://www.oldweather.org>

2 RELATED WORK

though having their merits in general, can lead to a decrease in user motivation: The extrinsic effects of rankings or badges, e.g., rising social status in the community, can replace already present intrinsic motivations. Since intrinsic motivations are more sustainable, the overall motivation decreases in the long run. As a counter measure, Nicholson proposes the use of meaningful, **user-centered gamification**, for example allowing for individual user-set goals and understanding badges more as “signposts” towards these goals instead of status symbols. As mentioned before, similar observations were also made by Eveleigh et al. [EJLC13], who explicitly advocate for using scoring mechanisms as personal milestones in addition to their competitive aspects.

User Profiles While not seen as gamification per se, allowing users to create accounts or profiles is a necessity for many other aspects of gamification which build upon these, like rankings or achievements. While many citizen science projects allow users to partake and explore the project on an anonymous level without registration, many functions are inaccessible for them. For example, anonymous users are not listed in the leaderboards of *Old Weather*; in *Galaxy Zoo*, anonymous users can contribute to classifying the galaxies, but are excluded from discussing findings any further and *Foldit* prevents contributions of non-logged-in users in the first place. However, enforcing user profiles with a mandatory registration and login comes with a trade-off, as newcomers in general avoid registration and rather leave the site [Wro08].

Teams, Groups and Social Community Creating teams, groups or social communities in order to supplement their platform’s core functionality has been done by many gamified citizen science projects. Several researchers have identified social interaction as an important aspect of gamification [McG11, GHL⁺14]. For example, *Foldit* allows users to form teams (which are also ranked in a global leaderboard), where they can communicate, organize and share their self-made macros for protein folding [CKT⁺10]. In *Old Weather*, contributors to a specific ship log have the possibility to start discussions revolving around the specific ship. Also *Galaxy Zoo*, though no example for gamification in citizen science itself, enables users to start discussions on all images with their fellow volunteers. Nevertheless, implementing community aspects on crowdsourcing platforms can bring problems of their own: The administration effort increases, as rules of conduct have to be enforced in order to ensure a productive collaboration environment. Additionally, the exposure to spam and illegal content increases if the communication of the community happens on a public level, e.g., via forums or public messages.

Narrative One often neglected aspect of gamification in citizen science is the narrative dimension a project can provide. Aside from providing a full-fledged game with an elaborate storyline like *Forgotten Island*,⁷ this also applies to more conventional projects. Eveleigh et al. [EJLC13] found when evaluating their ranking system for *Old Weather*, that some users were not motivated by the ranking system, but rather by the narrative appeal of following a specific ship’s journey through the ship logs. They call this aspect of the project, described by users as “fascinating” or “absorbing”, *narrative* or *imaginative immersion*. They see the

⁷<http://www.citizensort.org/web.php/forgottenisland>

narrative appeal of many projects as an opportunity for new features that can carve out narrative stories already present in the dataset and actively promote the immersion into those stories. However, as with gamification in general, the enforcement of narratives can impair the experience for the science-oriented audience.

2.4. Quality of Citizen Science Data

Even in today’s academic community, determining the quality of scientific research is not a simple task. The reputation of the researcher, past peer-reviews, a high citation count or the publication in established journals may be reference points, but the issue is so complex that it has spawned numerous research works. Determining the quality of contributions in crowdsourcing projects in general and citizen science in specific is even more complicated: Because of the usually non-uniform background of contributors, their varying expertise and – in many projects – the provided anonymity, the danger that some contributions are of poor quality or even consciously falsified is not neglectable [AH10, FSE03].

Projects purely focused on data-collection can remedy this problem by verifying any user input by comparing it automatically to input of other users or by running outlier-detection algorithms. For example, in ARTigo, a tag for an artwork is only considered validated if it has been provided by at least two players [WBBL13].

For projects where the focus lies on analysis and interpretation of data with free text discussion of the findings, like the platform presented in this thesis, the solution is not as obvious. Still, there is a need for estimating and ranking the quality of contributions. Otherwise, finding high quality contributions amongst mediocre ones would, after a high enough number of contributions, deteriorate towards finding the proverbial needle in a haystack. In a general way, contributions of citizen scientists to this kind of projects can be seen as a special form of user generated content and handled with similar concepts. The following section will show some approaches to this issue.

Measuring popularity of user generated content

As the quality of text-, image- or video-based user generated content is not automatically discernible and often very subjective, many platforms hosting user generated content rely on several metrics that try to estimate the popularity of contributions based on the response and reaction of its user base.

As every metric has its own merits and drawbacks, the user is often presented with several rankings based on different metrics. Common simple metrics are “views”, i.e., how often a certain contribution was viewed by other users, explicit rating of the contribution by other users and “most discussed”. Additionally, many sites rely on more sophisticated metrics, whose details and algorithms are not disclosed to the public [VD09]. *Reddit* and *Hacker News*, for example, two social news aggregators, also factor in the age of contributions, favoring newer ones, for calculating their popularity rankings [AS13].

It is important to note that these ranking metrics therefore regulate the visibility of contributions and influence how often specific contributions are read. This makes them susceptible for manipulation, either by the site operators themselves or by the users [VD09].

2 RELATED WORK

Another important observation is that the popularity of user contributions does not necessarily reflect their quality [SDW06, Sto15]. Factors like visibility or social bias may distort this relationship, which can lead to a self-reinforcing effect: Already popular contributions of low quality receive better visibility, leading to more popularity in turn at the expense of high-quality contributions.

2.5. The ARTigo Project

The citizen science platform presented in this thesis is build around the data produced by the ARTigo project. The foundation of ARTigo is its large artwork database, which encompasses currently over 58,000 accessible artworks. The works of art included in the database are not limited to any specific styles or time periods and range from the antique to contemporary art, including all different kinds of paintings, sculptures, photographs, architecture, etc.

ARTigo was initially created to address the problem of creating semantic tags for this large number of artworks in an effective and economical way in order to build a semantic artwork search engine based on this data. In contrast to expensive manual indexing of the artworks by paid manpower, ARTigo follows the idea of *games with a purpose* (GWAP), as described by Luis von Ahn [VAD04, VA06]. GWAPs in general are a human-based computation technique where parts of a computational process are delegated to humans in an entertaining and game-like way.

ARTigo adopted this concept and provides different kinds of GWAPs in order to generate meaningful semantic tags for its large artwork database. The games have been designed to be complementary, i.e., each game excels at creating different kinds of tags and data [WBBL13]. The data created from these GWAPs is in turn used as basis for a semantic search engine on the artworks of this database.

ARTigo currently consists of three distinct parts: The *gaming ecosystem*, the *semantic artwork search engine* and a not yet mentioned component for data analysis of the tag data, the so-called *Analytics Center*. In the following sections, in addition to the ARTigo games, the ARTigo search and the Analytics Center will be described in more detail, also with regard to integrating them as data analysis tools into the citizen science platform presented in this thesis.

2.5.1. ARTigo Gaming Ecosystem

As of now, the ARTigo platform offers several different games for producing a diverse set of tags. The games are designed complementary, i.e., the different games cooperate in order to generate different kinds of tags. Wieser et al. [WBBL13] use the term “ecosystem” to describe this interdependence of the games.

They categorize tag-related games on the ARTigo platform into four groups: Description games, diversification games, dissemination games and integration games. Description games are simple games where users enter tags describing anything related to the artwork. They primarily generate surface tags [BW12], i.e., tags that describe what is explicitly represented in an artwork, for example, “man”, “sword” or “tree”. Dissemination games allow

existing tags to be propagated to different artworks or to translate them to different languages. Diversification games create more specific and precise tags and/or deep semantic tags. They depend on already collected description tags as input for their own game mechanics. Finally, integration games take tags collected in the other game types as input to create semantically rich tag clusters or combinations, i.e., they establish relationships between tags for a single artwork. In the following, a short listing will describe the current games of the ARTigo platform, based on the aforementioned classification of Wieser et al. [WBBL13] when applicable.

ARTigo Game is an adaption of the *ESP game* described by von Ahn and Dabbish [VAD04].

It was the first game of the ecosystem and served as an eponym for the whole ARTigo platform. Two randomly paired players are presented with the same artwork and have to enter tags describing it. It is a description game and an output-agreement game [VAD08], as players only receive points if a tag has also been entered by other players. In order to prevent cheating, communication between players is not possible.

ARTigo Taboo is a diversification game and offers a similar gameplay to the ARTigo Game.

The main difference – as the allusion to the similarly named party game suggests – is that the game prohibits the usage of some tags entered previously by other users. Therefore, the user is forced to describe the artwork with less common or obvious tags, i.e., tags with a deeper semantic meaning. With its use of taboo words, it is a direct implementation of the original *ESP game*.

Sentiment is also based on the *ESP Game*, and currently only available for selected users [Wie14]. Users are asked what sentiments or feelings an artwork conveys in their opinion. It’s both a description as well as an integration game, as it creates new tags without input tags from other games, but all tags created belong to the deep semantic cluster of “feelings conveyed”.

Karido as described in more detail in [Ste10], is another diversification game. The game’s purpose is to generate more specific, deep-semantic tags. It’s a two player game where one player has the role of the describer, the other is the guesser. The players are presented with nine similarly described artworks. The describer picks the artworks one by one and describes it with tags that set it apart from the other artworks, meanwhile the guesser has to identify the matching artwork according to the given description of his partner. It’s cooperative in nature, as both players receive the same points and try to achieve the maximum possible score.

Combino is a two-player integration game with the fundamental mechanics of the *ESP game*. The players are presented with an artwork and a list of tags it had been previously tagged with. Using these tags, the players have to create semantically linked tag-pairs that describe a single facet of the artwork in more detail [Stöl12].

Tag-A-Tag is a single-player integration game, doubling also as diversification game. The player is presented with an artwork and a single tag related to the artwork (generated in other games). The player has to enter additional tags describing the relationship between the given tag and the artwork.

ARTime is a new two-player game for the ARTigo platform, currently undergoing testing and not publicly available. In contrast to the aforementioned tag-based games, it

2 RELATED WORK



ADVANCED SEARCH

Tags	<input type="text" value="e.g. sky -clouds"/>
Artist	<input type="text" value="e.g. da Vinci"/>
Title	<input type="text" value="e.g. Mona Lisa"/>
Location	<input type="text" value="e.g. Paris"/>
Year	<input type="text" value="e.g. 1701-1900"/>

Figure 3: The input form of the advanced search, artworks can be searched by tags, artist, location and year or range of years.

serves both the purpose of generating estimates for the date of origin of artworks (information often missing in the ARTigo database), as well as detecting discrepancies between known dates of origin and players’ perceptions. The player is presented with three artworks which he has to order chronologically. Since artworks with known dates of origin are also presented by the ARTime game – additional to those with missing dates of origin – an estimation of an artwork’s missing date of origin can be computed after a sufficient number of gaming rounds including that artwork.

2.5.2. ARTigo Artwork Search

The original purpose of the ARTigo games was to gather enough tags to make a semantic tag-based artwork search engine on the ARTigo artwork database feasible. Since its first implementation, the search engine and its user interface have undergone several iterations. The following section will give a short overview of its current state and implemented functionalities as well as the missing requirements necessary for integration into the citizen science platform. Technical details will be skipped and are described for the interested reader in detail in [WBBL13, Wie14, Sig15].

The ARTigo artwork search implements two kinds of searches, the so-called simple search and an advanced search. As all features of the simple search are included in the advanced search, the following section will focus on the advanced search. A screenshot of the advanced-search form presented to the user can be seen in Figure 3.

The search also implements a simple query syntax for tags: Artworks matching all of the searched tags are matched by entering “tag1 tag2 tag3 ...”. The “-” operator can be used to exclude artworks from the search that are described with a certain tag, e.g., “horse -rider” only matches artworks of horses without rider. Artworks tagged with combined tags (created in the games *Combino* or *Tag-A-Tag*) can be searched by the syntax “tag1+tag2”. For example, a search for “brown+horse” matches only artworks including brown horses, in

contrast to “brown horse”, which would also match an artwork featuring a grey horse in front of a brown log cabin.

The matching artworks for a search query are presented as a list with describing metadata and tags for each image, as seen in Figure 4. Displaying the metadata and tag cloud can also be dynamically disabled to get a more concise overview of the search results.

The screenshot displays the ARTigo search results interface. At the top, there is a search bar with 16 search results, 25 results per page, and options to display search form, tags, and data. Below the search bar, there are tags for 'cat' and 'Year 1750-1914'. The main content area shows a grid of search results, each consisting of a tag cloud, an artwork image, and its metadata.

Tag Cloud	Artwork Image	Artist	Title	Location	Date
BED WOMAN GREEN SERVANT LADY CAT SHOES NAKED MAID NUDE		Edouard Manet	Ohymnia	Paris / Musée d'Orsay	1863
SKETCH KITCHEN WOMAN BIRD PICTURES DOG CLOCK SHELF CAT ANIMAL POT CHILD MOUSE CATS ROOM DRAWING ANIMALS		Gabriel von Max	Hexenküche mit Tieren	München / Städtische Galerie im Lenbachhaus	1870/1882
POT CARROTS APRON PANS POTS STEAM BROWN CAT COOKING DOOR WINDOW WOMAN HOUSE DRESS VEGETABLES KITCHEN BASKET OLD DOG FLOOR		Edward Lamson Henry	Kitchen Interior, Oberammergau	Amherst (Massachusetts, USA) / Mead Art Museum, Amherst College	1871
TABLE DRESS RED CAT PLANT CARPET ROOM SUIT RUG WINDOW VASE GREEN WOMAN CURTAIN FIRE PLACE MAN SHOES CHAIR FLOWER MUSTACHE		Faust Signac	Un dimanche	Location (Privatsammlung)	1888-1890
FOREST WHITE GREEN TREE BED TREES CAT GIRL FACE WOOD PAINTING DRESS NOSE EYES PEOPLE CHILD SKY HAIR GRASS WOMAN		Felis Modersohn-Becker	Mädchen mit Katze im Birkenwald	Brämen / Sammlung Ludwig Roselius	1904
PANTHER STRIPES CAT INK PAINTING PAPER EARS WHISKERS PAWS DARK WHITE CLAWS DRAWING ANIMAL NOSE BLACK BLACK AND WHITE EYES TIGER TAIL		Franz Marc	Tiger	Lübeck / Museum für Kunst und Kultursgeschichte	1911
BLACK PINK RED GRAY FUR EARS 1819 EYE HAIR PROFILE WHISKERS HEAD PENCIL STUDY NOSE ANIMAL WHITE DRAWING MOUTH CAT		Jean Bernard	Kop. van een kat, naar links	Amsterdam / Rijksmuseum	1819
WOOD CAT HATS HANDS BROWN ROOM PLATE INTERIOR FLOOR HAT WINDOW DRESS KIDS FAMILY FEET SAUCEUR BASKET CHILDREN MEN BENCH		Ferdinand Georg Waldmüller	Der Abschied des Konkskribleren	Wien / Sammlung Leopold	1854

Figure 4: Search result presentation of the ARTigo artwork search.

Additionally, the search features an aggregated metadata visualization for the search results, as seen in Figure 5: The distribution of artwork creation dates (Figure 5a), tag clouds of the 50 most common tags and combined tags for the matched artworks (Figure 5b) and lists of the most frequent artists and current artwork locations (Figure 5a).

Regarding the integration of the search into the citizen science platform presented in this thesis, the current version has several limitations that need to be addressed: So far, search queries cannot be saved in any way, thus a search cannot be rerun at a later point in time without manually reentering all parameters. Also, results of a search cannot be saved for later use for comparison with a future state of the tag database.

As the ARTigo games are played on average 150 times every day and therefore continuously produce new tags, the ARTigo dataset is not static, but changes over time. Therefore, search results may differ notably for queries with the same parameters if run at different points in time. Without the possibility of saving a query and its current results, there is no easy way to rerun it with updated data and compare changes in the results.

2 RELATED WORK

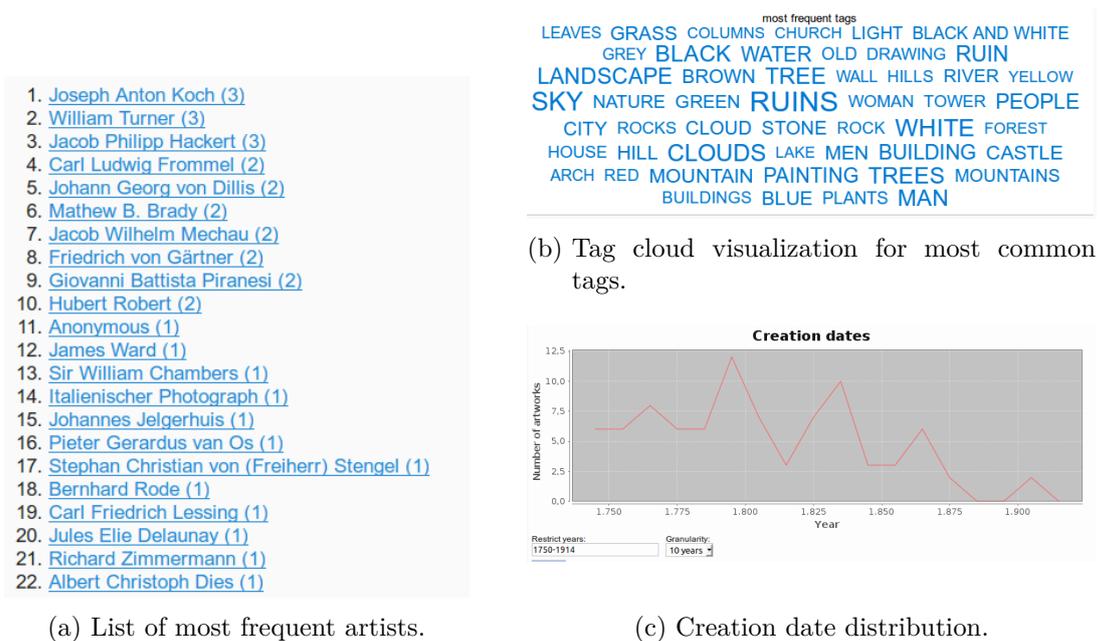


Figure 5: ARTigo Search tag and metadata visualization types.

Also, though citizen science relies on sharing one’s findings with others, there is no possibility to send a link for a certain search to another user so he can reproduce the query; the same goes for linking to single artworks, which was disabled in order to prevent external linking to the images but hinders collaboration. These issues need to be addressed to make the search reasonable easy to use for the citizen science platform.

2.5.3. ARTigo Analytics Center

The ARTigo Analytics Center is a first attempt to enable data science on the ARTigo tag data. It was created with the intention to “*help ARTigo users to spot thematic trends within certain artistic epochs, determine how specific [...] their tags are, and discover clusters and interesting associations between the motives*” [Hoi14].

It implements a query interface for several data science algorithms on the ARTigo tags and metadata, namely tag frequency over time, tag specificity computation overall and over time, tag association rule computation, tag similarity and tag clustering. In order to allow elaborate queries, the Analytics Center features a sophisticated query language, explained in Table 1. Queries for combined tags are currently not implemented. It is important to note that the current version of the Analytics Center focuses on tags and taggings,⁸ not the tagged artworks.

The individual components of the Analytics Center’s features are described in the following:

⁸The action of linking a tag, e.g., TREE with a specific artwork in the course of a game round; This can occur multiple times per (artwork, tag) tuple and can be seen as a measure of dominance for the feature described by the tag in the artwork.

Syntax	Explanation
tag1	Basic query for selected tag1
!tag1	Matches all taggings except tag1
tag1,tag2	Run two independent queries, one for tag1, one for tag2
tag1 tag2	Matches when artworks were tagged with either tag1 or tag2 or both
tag1&tag2	Matches when artworks were tagged with both tag1 and tag2
tag1~tag2	Matches when artwork was tagged with tag1, but not tag2
tag1=>tag2	Computes selected algorithm for tag2 on a basis consisting of artworks tagged with tag1, for example <code>Franz Marc => Horse</code> runs the query for “Horse” limited to paintings tagged with “Franz Marc”
*x	Matches the x most frequent tags

Table 1: ARTigo Analytics Center query syntax.

Tag Frequency Plot Inspired by the Google Ngram Viewer,⁹ the tag frequency plot visualizes the relative frequency of taggings for a given tag over time. The frequency for a single year is calculated by taking the number of times all artworks of this year have been tagged with the given tag, i.e., number of tagging actions, in relation to the number of times these artworks have been tagged overall.

Poisson Overdispersion The Poisson overdispersion is a statistical value which characterizes the specificity and salience of a given data point. The Analytics Center uses this to determine how specific and at the same time easily recognizable given tags are. The values are either presented on a time series plot or as the overall value. The result for an example query can be seen in Figure 6.

Association Rule Probability In the context of the ARTigo tag data, association rules measure the probability of how likely an artwork, tagged with a certain set of given tags X, is also tagged with another set of given tags Y.

Tag Similarity Ranking This component can be used to identify similar tags to the queried tags. Internally, artworks are represented in the tag vector space, i.e., vectors that hold the information how often an artwork was tagged by users for each tag. Using this as basis, the component computes the euclidean or cosine distance between tags in this vector space and generates a similarity ranking for the given query from its nearest neighbors. It also features a PCA-Visualization.

Cluster Expansion This module of the Analytics Center implements an algorithm that expands a single cluster around given tags. Its purpose is therefore not to detect all clusters in the data, rather it identifies tags that belong to the same cluster as the given tag. Explained intuitively, it is suited for detecting overarching themes or relationships to tags not immediately similar and only indirectly linked to the queried ones.

The current version of the Analytics Center has some limitations that make a collaborative usage difficult: Queries cannot be bookmarked in order to rerun or tweak them by other

⁹<http://books.google.com/ngrams>

2 RELATED WORK

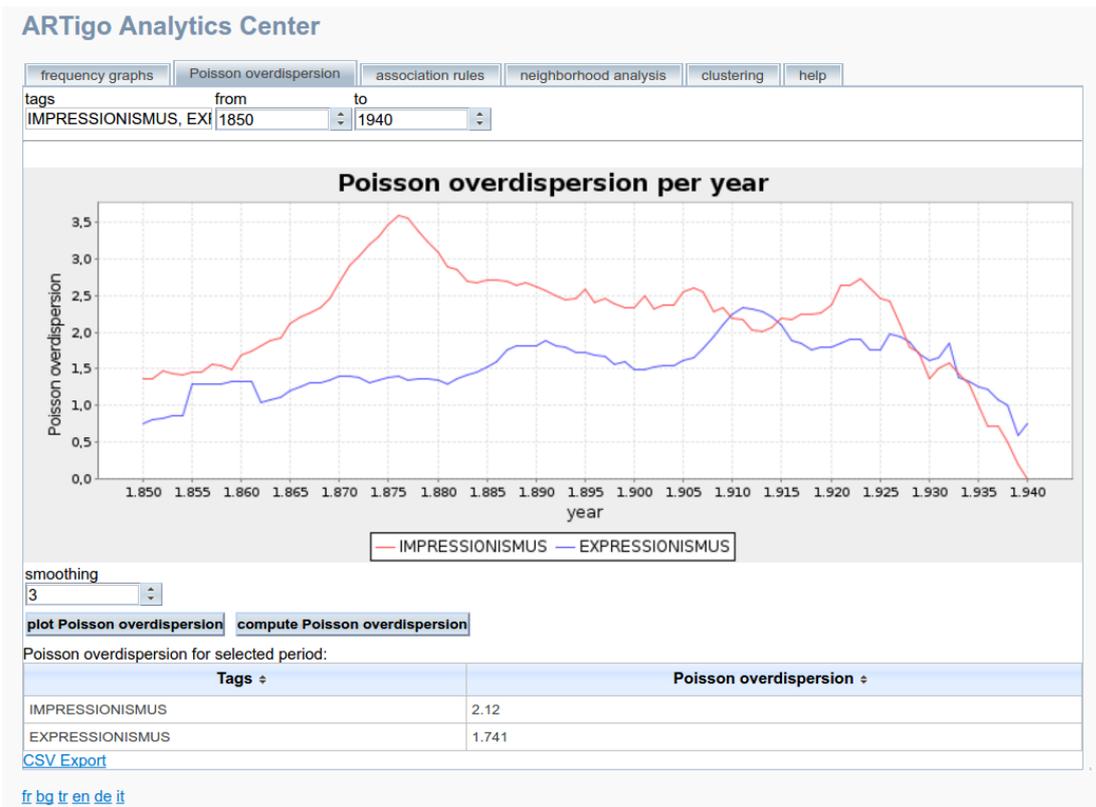


Figure 6: The Poisson overdispersion component of the ARTigo Analytics Center, query for the German tags **IMPRESSIONISMUS** (Impressionism) and **EXPRESSIONISMUS** (Expressionism).

users. Also, query results, especially the plotted graphs, are not cached in any way, though some components rely on costly computations and therefore take a long time to process or rerun a query. These issues will be addressed later on in this thesis (Chapters 3.7 and 4.3) and solutions on how to achieve a productive integration into the citizen science platform presented.

3. Conceptualizing the Platform

As seen in Chapter 2.5, the ARTigo ecosystem already features tools like the artwork search or the Analytics Center for exploring and researching the generated data. Currently, the ARTigo ecosystem covers generation of data in a crowdsourced way (classification of the artworks with tags) but lacks a central hub for starting and coordinating collaborative research efforts on this data. While most citizen science projects harness volunteers in the early stages of research for data collection and classification, they generally concede no influence on the direction of the research to the volunteers and leave analysis and interpretation of the obtained data exclusively to professional scientists [BBJ⁺09].

In contrast, the citizen science platform implemented in the scope of this thesis tries to transfer the citizen science spirit already present in the tag creation process of the ARTigo games to the analysis and interpretation of the generated tag data. Based on the quantitative analysis of the ARTigo tag data, the platform should give anyone interested in art history the possibility to collaboratively analyze and interpret the data, share and discuss their findings with others, publish them to the general public and ultimately gain a deeper understanding of art history overall. Volunteers participate in all steps of the scientific process, including the formulation of research questions to be investigated. This is achieved by users collaboratively reaching interesting conclusions from the ARTigo data with help of the provided data analysis tools and then sharing their findings in article-style contributions. This format, compared to simply presenting the output of the data-analysis, allows to embed the results inside text providing both context, further explanation on the data analysis queries performed and possible interpretations of the results. Imposing further restrictions, like limiting the number of data analysis queries per contribution or enforcing a strict format, would hinder the creativity needed for making novel research findings.

Thus, the platform needs to meet several requirements: To begin with, it has to facilitate a meaningful collaboration between its users while at the same time keeping the administrative effort manageable for the platform operators. Secondly, every effort has to be taken to ensure a motivating and rewarding environment in order to keep users engaged with the project. And finally, the existing and future data analysis tools need to be integrated in a both simple-to-use and easy-to-integrate way.

This chapter will provide a general overview of core concepts behind the citizen science platform, including the reasoning that led to certain design decisions. The scope and objective of the platform, mechanics for user engagement and collaboration as well as the integration of data analysis tools into the platform will be discussed in detail.

3.1. Scope and Language

One of the core truths of data analysis or statistics in general is that findings made from analyzing any given dataset are only as good as the quality of the underlying dataset. Otherwise, conclusions coherent for a specific dataset are meaningless in the greater scheme. Therefore, it needs to be ensured for the new citizen science platform that the data selected from ARTigo is as detailed and representative as possible in order to allow any findings to be of actual relevance.

3 CONCEPTUALIZING THE PLATFORM

As seen in Chapter 2.5, the artworks featured in the ARTigo dataset span across the globe from the antique until today. Additionally, the ARTigo games can be played in multiple languages (German, English and French), resulting in a different tag dataset per language. These datasets are independent from each other, tags created in one language are not automatically translated to other languages.

Looking at the distribution of artworks included in ARTigo in respect to their date (see Figure 7) and place of origin reveals that the majority of artworks in the ARTigo database originate from the time period roughly between 1750 and 1920 and are from Europe. Many historians refer to the time period starting around 1750–1800 (depending on the historian in question) and ending with the beginning World War 1 in 1914, with the memorable term “long 19th century” [ALH⁺00, Hob10].

It was therefore decided to limit the focus of the platform to this specific period, European art of the long 19th century, providing both the most representative cross-section of artworks and data in the ARTigo dataset as well as a memorable boundary for the platform. The decision was made to use the more lenient definition of 1750–1914 in order to allow more diversification for the research.

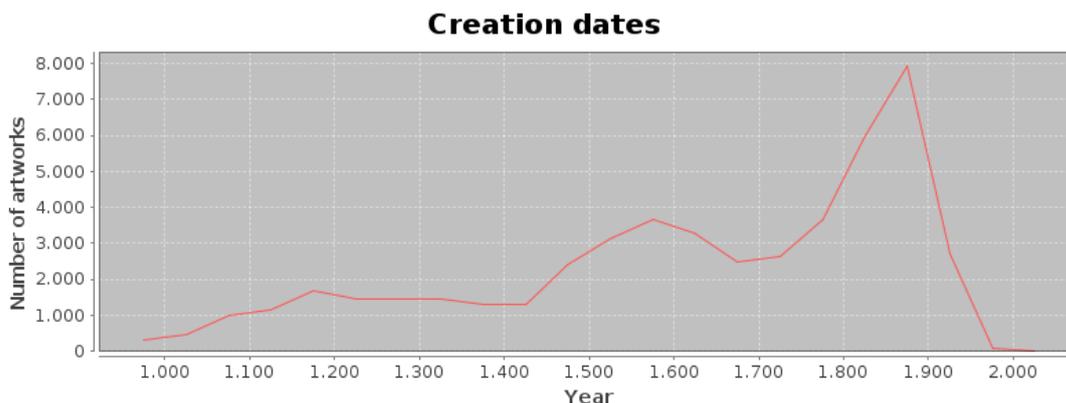


Figure 7: Distribution of date of origin of artworks included in the ARTigo dataset (artworks with unknown date of origin excluded).

Another important decision was made regarding the language used on the platform. As working on tag data different from the language of the platform seems dubious – How should an English speaking art historian work with an artwork tagged in German? – the platform and the selected tag dataset have to be of the same language. As English is the dominant international language both in science and international communication per se, using English as a citizen science project’s language is preferable in general if no restricting circumstances exist. Because of several characteristics stemming from the ARTigo project, the platform’s primary source of data, the decision was made to initially use German as the platform’s language. This decision stems from two considered aspects: First, the quantity of tags and second, the possible user base. While ARTigo is available in different languages, German is by far the most used language and in consequence thereof, the majority of tags are German. Additionally, since every citizen science projects has to establish a sufficient and

sustainable user base, the hope is to recruit some of ARTigo’s German speaking members and use the synergy as both projects are closely related.

Still, in case of a future larger English or French user base of ARTigo in conjunction with a rise in the respective number of tags, an international version of the platform should not be ruled out categorically. Therefore, the platform is prepared for easy translation from the start.

3.2. Target Audience

When designing any collaborative platform that relies on volunteers, it is of critical importance to identify early on the types of people who are most likely to participate and target them accordingly. As seen in Chapter 2.2, getting involved into a citizen science project is triggered by an initial interest in the project, which depends the user’s affinity for the project’s goals and characteristics.

Considering the goal and distinct characteristics of the citizen science platform presented in this thesis, its main goal is bringing together art history and data science in order to gain a novel insight on art history in general. It is bridging between the usually not related fields of art history and data science, a contrast often emphasized in the past as the nomothetic–idiographic separation of academic disciplines into the sciences and the humanities [Win94, Sch15].

Because of this quantitative and statistical approach to art history pursued on the platform, its targeted users groups are, additional to the art-interested in general, people with a background in computer science or statistics. Catering to the primary target audience of computer scientists, the platform provides them with the opportunity to gain experience with common data science practices and algorithms.

In addition to serving the interest in the subjects of art history, citizen science or data science, it addresses the egoism of users with a background in computer science or similar fields – especially students – on another level: It offers them the possibility to apply and get a deeper understanding of concepts learned in lectures on real world data, an aspect often neglected in today’s teaching at universities. In exchange for their participation, they can gain practical experience with common data science and knowledge discovery techniques and profit later on in studies or work.

3.3. Research Questions

As previous chapters have focused on more abstract ideas and theoretical background of the citizen science platform, this section will provide some practical examples for research questions that can be tackled on the platform. With the ARTigo data as the foundation, many interesting topics of art history can be explored. Some possible categories for research question are shown below:

Changes in Society and Politics: How are certain changes in society and politics, e.g., rise of the urban bourgeois and decline of nobility, reflected in the art of the era?

3 CONCEPTUALIZING THE PLATFORM

Impact of Historical Events: How were single historical events, e.g., the 1755 Lisbon earthquake or the French revolution, depicted in art and which lasting effects did they have?

Impact of art-related technical inventions: Many technical inventions relevant for art and painting were made in the 19th century, e.g., synthetic dyes or electric light. How did this influence subsequent art?

Artist's themes and influences: Art does not happen in a vacuum. Artists are often influenced by other artists and their surroundings. What are an artist's dominant themes and his influences?

Art movement and geographical comparisons: What are the measurable differences between selected art movements? Are there distinct differences inside a single art movement, e.g., Impressionism, across several countries?

This listing does not claim to be exhaustive, also making clear distinctions between the categories may be difficult in practice for some of the questions. Still, it gives an impression on the diversity of possible research that can be achieved on the platform. Often, these questions have been previously researched by classical art history beforehand. The novel approach of the platform is to find quantitative data that supports these claims.

Formulating a universal approach for the analysis of the ARTigo data in the context of exploring a research question is difficult due to the diversity of possible questions, but some fundamental considerations are described in the following:

- As the ARTigo tag data comprises primarily of surface-tags [Sch14], researching highly abstract tags as “nation”, “democracy” or “country” will yield only few results. A more promising approach is to find correlating concrete terms for the abstract ones, e.g., “flag” for “nation” or “parliament” for “democracy”.
- Most research has to start with identifying tags related to the research question, both abstract and concrete, and a suitable time frame and then proceeding from there. The most relevant tags can be singled out by computing the Poisson overdispersion, further analysis focusing on them.
- Though ARTigo does not bring a tool that actually analyzes the color spectrum of artworks, the dominant colors are some of the most common tags. Therefore, the usage and frequency of specific colors in artworks and changes to the color spectrum over time can also be questions of interest for the platform.
- The ARTigo tag data is human created and reflects today's users' perception of artworks and is not necessarily an objective, accurate description. Instead of considering this as faults in the data, this can also be a chance for research: Analyzing typical “mistakes” made by users when tagging an artwork, e.g., repeatedly confusing two artists, can lead to interesting results. In this example, the confusion could be founded in their similar painting style or choice of subjects and indicate a connection between the artists.

3.4. Engaging the Citizen Scientists

As already shown in Section 2.2, any citizen science platform that fails to attract the attention of potential volunteers and engage them for long-term participation will fail overall. Additional to the considerations already presented in Chapter 3.2, the following chapter describes concepts implemented in the citizen science platform both to attract new users as well as to encourage continuous engagement of existing volunteers.

3.4.1. Attracting New Volunteers

Any platform on the Internet has to compete for the initial attention of users. For new users, unnecessary hurdles and frustration must be avoided as good as possible: As the Internet offers an abundant number of choices for activities, it is of critical importance to not deter potentially interested users by confusing them with a complicated user interface on first visit or forcing them to register before seeing most features of the platform. A once turned-away user is unlikely to come back, so the platform often has only one single opportunity to peak a potential volunteer's interest when he first visits the platform. Concepts employed by the platform to gain new volunteers are presented in the following:

Home Page The home page of the platform is the first impression any new users sees when he first visits. It is therefore critical to communicate the essence of the citizen science platform and catch the interest of passing by users. The home page of the platform employs several features to achieve this.

As the platform is focused on art history, the visual representation has to reflect this accordingly, i.e., artworks have a prominent role on the home page and are preferred to long descriptive text paragraphs. The goal is to captivate a user's interest on first glance.

A call-to-action message is displayed prominently on the home page. A technique employed often in marketing, the purpose is to both communicate unique selling-points of the platform, bringing together art history and data analysis, as well as emphasize the need for active contribution.

Also, several examples for research questions and already made findings are presented, the purpose is to show off the potential and possible impact of the platform for art history. Visitors get direct example on what work is performed on the platform and what they themselves could do if they participate in the platform.

The home page also has to presents the organization behind the citizen science platform and its underlying goals. This communicates both the relevance of the platform as well as the trustworthiness of the organization. Intransparency in this regard would otherwise discourage volunteers from dedicating their time and effort to the platform.

Access without Registration On the one hand, registration on any new Internet site is a big commitment for most users, many opt to rather leave a site than to suffer through an elongated registration process [Wro08]. On the other hand, user registration is necessary for certain features of the platform, for example attributing contributions to users or facil-

3 CONCEPTUALIZING THE PLATFORM

itating its community aspects. For instance, without registration and user accounts, there would be no way of ensuring the identity of users, for example to locate contributions of high-experienced senior members of the platform. Additionally, the danger of spam and vandalism increases. Both sides have to be considered carefully.

The decision was made to allow unregistered users access to most areas and features of the platform, in order to enable users to get an impression of the platform early on, but to restrict access to some of the collaborative aspects of the platform. The hope is to convince new users interested in the platform to participate occasionally and get gradually more engaged, finally signing up and becoming contributing member on the platform.

Additionally, search bots receive the same access as unregistered human users in order to ensure a good coverage of the platform's content by search engines. This aids in increasing the visibility and exposure of the platform, in turn attracting more newcomers.

To keep the initial entry barrier low, unregistered users are allowed to view all public contributions, showing them interesting findings already made by other users and piquing their further curiosity. They can participate in discussions surrounding the contributions, giving praise, helpful advice or new ideas. Also, they have limited access to the data analysis tools (alongside introductory examples and tutorials) in order to allow them to make their first steps in data analysis and to get accustomed to the work done on the platform.

However, some features of the site, like membership in projects and accessing their protected areas, rely on the existence of user accounts and cannot be provided to unregistered users.

Additionally, unregistered users are restricted from creating and editing the article-styled contributions or from accessing the protected areas of projects. There are several reasons for this decision: First, this is a consequence of the project-centered workflow enforced by the platform, as explained in detail later in Chapter 3.5. Second, it limits the potential for vandalism on the website, lowering the platform operators' administrative load. This is an important requirement in the academic environment the platform will be operated in, where resource and manpower needed for administration is often scarce. Finally, by restricting some of the more advanced features of the platform, a strong incentive for registration is provided. Similar concepts of withholding certain features to registered or premium members is a common marketing strategy across the web.

3.4.2. Engaging Existing Users

After getting volunteers interested in the platform, they need to become active participants. The platform, as any citizen science project, relies on contributions of volunteers, only having passive readers as users is not sufficient. As seen in Chapter 2.2.1, a high, but passive number of users still means failure in the crowdsourcing and citizen science context. Therefore, the first goal is to motivate visitors to becoming active participants. The first steps employed by the platform, providing low entry barriers and incentivising users to register and create accounts, have been already presented above.

This chapter will present considerations towards two aspects of engaging users: First, activating newly registered users or those who show a long-term interest in the platform, in order to make them contribute on a recurring basis while avoiding to repel them.

Second, keeping up the level of participation of already active contributors by providing a continuously motivating environment, in order to avoid user fatigue and resignation after elongated participation, a common occurrence on crowdsourcing platforms [HGMR12, SF14, McG11].

As participation in the platform is voluntary and unpaid, no extrinsic incentives are offered by the platform, making intrinsic motivation the main driving force behind participating in the citizen science platform. To account for that, the participation experience needs to be as rewarding as possible and tailored to providing intrinsic motivational factors.

The four general categories for intrinsic rewards defined by McGonigal [McG11] (*meaningfulness, social connection, prospect of success* and *satisfying work*), presented earlier in Chapter 2.2.2, are addressed by the citizen science platform in the following manner:

Meaningfulness of participation is offered by allowing users to make novel contributions to art history and ideally receiving recognition for their findings by professional scientists. *Social connection* is achieved via the collaboration facilitated by project structure. *Prospect* and *experience of success* is produced by giving contributors measurable feedback in the form of positive ratings, comments and acknowledgment as well as by showing them their contribution amount.

Finally, in combination with giving each project a clear goal, formulating concrete, well circumscribed tasks can provide actionable next steps during work for most users, preventing them from idling and making the work more *satisfying* in general. These tasks are not limited exclusively to data analysis in a narrow sense, instead they can include formulating research questions or proof-reading as well, in order to address a broader audience.

To further increase the enjoyability and engagement potential of the platform, its design borrows several concepts from gamification. The gamification aspect is not apparent at first glance, as no gameplay per se is present. Also, user rankings were consciously avoided in order to underline the collaborative nature of the platform in contrast to the competitiveness rankings imply, additionally avoiding incentives detrimental to contribution quality as seen in [EJLC13].

Instead, the platform implements McGonigal's [McG11] more general aspects of games found in successful crowdsourcing platforms (as seen in Chapter 2.3) in the following manner: First, it provides the *good game world* by offering access to the data analysis of ARTigo database with its currently over 55,000 tagged artworks and data analysis tools, inviting exploration and discovery by the volunteers.

Second, and of equal importance is enforcing a *good game community*, characterized by being benevolent and including for new users. The damage done by an unwelcoming and restrictive community can be severe: For instance, in *Wikipedia* frequent rejection of newcomers' contributions and an often harsh and dismissive tone of communication by high-ranking members has led to a serious decline in active participants [HGMR12, Sim13]. To counteract a similar development, the platform enforces a netiquette that forbids rejection of contributions without extensive feedback and providing ideas for improvement in order to lower the frustration potential for new users. Also, as the platform is to a certain degree compartmentalized into self-managed projects, a concept explained in detail later in Chapter 3.5.1, a negative development of the community inside one project is less likely to affect

3 CONCEPTUALIZING THE PLATFORM

other projects and leaves users the possibility to stay on the platform and participate in a different project.

Last, for the *good game mechanics* established by the platform: Each contribution has to go through a feedback and improvement cycle, providing constructive and motivating suggestions for improvements before being published, in contrast to frustrating users with rejecting contributions without giving reasons. In the same spirit of providing positive feedback to the contributors, the platform also implements a rating system for contributions as well as activity tracking and statistics for users, described in more details later. The possibility to be rewarded for continuous commitment to projects by being promoted to a leading role in it serves as another motivation factor.

As the citizen science platform has to deal with a heterogeneous user base, consisting primarily of art historians, computer scientists, statisticians and the art-interested in general with varying degrees of proficiency, different tasks of varying difficulties are available for each group. It's obvious that art historians are better at formulating interesting research questions or providing context and background information, while computer scientists and statisticians on the other hand are better at data analysis. Hence, formulating research questions and hypotheses is a domain primarily interesting to art-historians, while selecting the tools and running the data analysis is more relevant for computer scientists.

In the following, details of features implemented on the platform will be discussed mainly with regards to their effect on user motivation:

Contribution Rating System As mentioned above, the platform implements a rating system for contributions. Rating systems in general have two basic functions: First, providing users with information about the popularity of content, helping them in deciding which contribution to read or improve, i.e., where to spend their limited resources (for volunteers in citizen science: time and effort). Second, it provides feedback for the contributor, in case of positive ratings affirming him in the meaningfulness of his work, which in turn raises his motivation and keeps him engaged. The rating system is kept simple, only allowing giving single-level positive ratings to contributions, respective revoking ratings at a later point in time. Graded and negative ratings were rejected: Negative ratings provide no productive feedback for users on how to improve a contribution in order to receive a better rating, potentially leading to frustration and a decrease in motivation. Graded ratings, e.g., one to five stars as employed by various other online platforms, have the same drawback.

User Profile Each user has a user profile page where past contributions and personal information about the user is displayed. Also, users are allowed give information about their skills and academic background in text form. This allows users to showcase their contributions and build an online persona on the platform. The amount of their contributions to the platform are immediately apparent, increasing the volunteers sense of importance and improving his identification with the platform.

Additionally, the user profiles assist the community aspect of the platform, as they help in determining the contribution potential of volunteers, e.g., when recruiting additional members for a project. Displaying detailed statistics about the activity of a user on its

public profile is not supported. Though it would further assist in determining a volunteer’s potential, this might conflict with privacy concerns of volunteers. For the same reason, providing personal information in the profile, like real name, external contact information or academic background is optional.

Implementing a user ranking system, showing the “top” contributors globally or per project was rejected for several reasons: First, in comparison to data-collection-centered citizen science projects, the value of each contribution is not easily apparent or quantifiable on the platform. For instance, an insightful comment inspiring a turn in research direction can be rather short, but more valuable for the platform than extensive editorial reorganization of existing contributions without adding new insights. This would provide wrong incentives, favoring low-effort contributions, as seen by Eveleigh et al. [EJLC13]. Second, the competitive aspect a user ranking system invokes does not fit into the collaborative atmosphere the platform tries to achieve. A ranking of users by the number of positive ratings on their contributions was not implemented for the same reason.

Instead, users are presented with personal non-public statistics and graphs about their past activities for selected periods, similar to many sports activity trackers. As users do not receive pre-defined goals or badges, the statistics primary objective is to serve as signposts to each user, in the spirit of Nicholson’s “meaningful gamification” [Nic15]. By providing no external rewards like badges or recognition in the community, it rather addresses the user’s urge for self-improvement, a powerful intrinsic motivator also seen in multiplayer role-playing games [McG11].

3.5. User Collaboration

As already reflected in the title of this thesis, collaboration is one of the integral characteristics of the platform. In contrast to many other online citizen science projects, where users submit data isolated from each other and with little or no communication in the process, the platform presented in this thesis aims for a more collaborative approach. This is in part necessitated by the shift of the platform’s focus away from the classical domain of citizen science, data collection and classification, towards integrating volunteers into more steps of the scientific process, beginning with the formulation of research questions as well as data analysis and interpretation. The word “interpretation” already implicitly hints at the reason why more communication and active collaboration is needed in these steps: Many findings in data analysis are a matter of perspective, they need to be put into context, scrutinized and further cross examined. A simple valid/invalid duality as seen in data collection cannot reflect these nuances.

The following section will introduce the basic structuring of the platform, the collaboration specifics on a research topic, user hierarchy and permissions as well as the public presentation of research contributions.

3.5.1. Structure of Research Topics

In order to provide a productive collaborative environment on the platform, its fundamental organizational structure has to take the characteristics of the research to be performed into

3 CONCEPTUALIZING THE PLATFORM

consideration. As seen in Chapter 3.3 about possible research topics for the platform, several loosely defined higher-order *research topics* can be identified, which in turn spawn several single, specific and well-defined *research questions*. For example, a single research question on the general topic “influence of technical improvements on artworks”, could be: “How did the invention of synthetic ultramarine blue in 1824 [Dou08], which made the color affordable and its extensive use possible, affect the color spectrum of paintings?”

In order to reflect this inherent hierarchy, the platform is structured on a higher level into distinct *research projects*. Each project is exclusively represented by a single research topic, which is explored in several cohesive *contributions*, i.e., research questions. This organizational structure helps to keep the number of participating users and the size of a contribution manageable compared to performing all work on a research topic in one single contribution.

A project is driven by a team of volunteers (its members), lead by one or several project leaders with administrative rights within the project. Membership in a project is mandatory in order to contribute to it, but the number of projects a volunteer can join is not limited. Figure 8 exemplifies the concept.

Compared to more generic citizen science platforms also hosting multiple projects like *Zooniverse*¹⁰ or *CitSci.org*,¹¹ the projects of the platform presented in this thesis are more closely related: While each project on *Zooniverse* and *CitSci.org* comes with their own datasets (e.g., *Zooniverse* hosts projects spanning from transcribing ship logs to watching local wildlife), projects on the platform presented in this thesis share the same dataset and data analysis tools from ARTigo as well as the abstract mission, researching art history through the use of data science. Therefore, projects are closely related, which allows for a community feeling across the whole platform and fluid interchange between the projects. At the same time, the fixed project and team structure encourages a stronger identification with each project and the made research results for its members. Providing fixed projects with clearly delimited responsibilities and formal membership also enables an additional benefit: Administrative tasks can be delegated to the projects themselves, making them mostly self-managed.

As of now, all projects are publicly visible and can be joined and left freely. This design decision is open for later review, as projects lead by professional scientists may have for example the need to closely examine prospective members before allowing them to contribute.

For the beginning, though, this was deemed unessential or even harmful for the platform: To attract a sufficient initial user base, interesting content and projects must be openly available in order to make the platform attractive in the first place. Providing access-protected or hidden projects, where users do not need to share their findings with the public, would be detrimental to this initial goal. Like most projects reliant on crowdsourcing, the citizen science platform presented here faces a *cold-start* problem: Without a sufficient user base and content, the platform is unattractive to new users. If existing content would be withheld from other members and visitors, even less content would be available in the starting phase of the project. On a more ideal level, this also opposes the general spirit of reciprocity present on the platform: While providing the tools for collaborative research to the users,

¹⁰<http://www.zooniverse.org>

¹¹<http://citsci.org>

the platform expects open sharing of research findings and participation in the community in return.

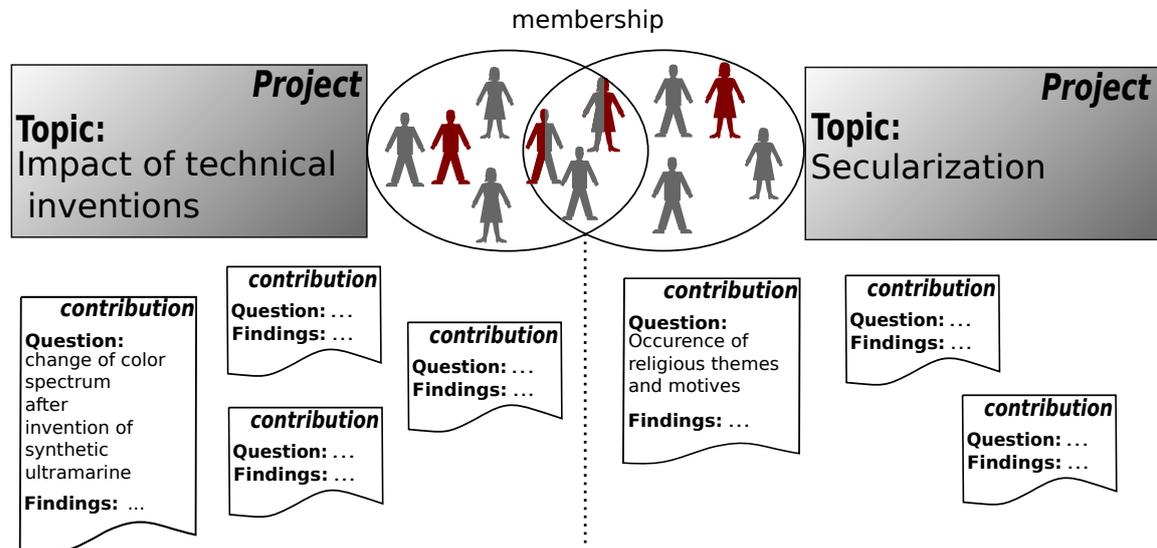


Figure 8: Project structure of the platform using two projects as an example. Project leaders are marked red. Note that users can be members of multiple project while membership status (leader or basic member) can vary for each project.

Other approaches, like loosely coupling research topics, projects and teams, were also taken into consideration but discarded in the design process, since clearly assigned responsibilities are a necessity for the intended productive collaboration on the platform. Such an approach, as implemented by *Wikipedia*, can produce competing or unclear responsibilities of several projects and teams within a single contribution. This has the potential to result in prolonged arguments between teams. Such disputes could be aggravated even more, if non project-members were allowed to edit a contribution.

3.5.2. Creating New Projects

In most collaborative science projects with public participation, the amateur volunteers are only included later in the research process and have no influence on its direction. Often, they are seen as little more than cheap manpower by scientists [RPH⁺12], any deeper involvement of them in the scientific process beyond execution of predefined tasks seen with reservation. In comparison, the platform presented in this thesis tries to embrace volunteer contributions on all levels. In order to achieve this, it enables volunteers to actively codetermine the direction of future research by giving feedback for planned projects or suggesting and initializing projects on their own.

The advantages are apparent: Through getting feedback from volunteers early on before creating a project or letting them create projects of their own, their interests and needs can be better addressed. As projects depend on the contribution of volunteers, who perform

3 CONCEPTUALIZING THE PLATFORM

the bulk of work, primarily realizing those with a good chance of high participation and contribution rates is sensible. Additionally, by allowing volunteers to take part in shaping a new project, their identification with the project in particular and the platform in general increases, which in turn positively effects their motivation.

However, some restrictions for the creation of research projects have to be enforced in order to ensure a thriving platform for the following reasons: First, even with a sufficient user base, the platform will be in constant shortage of working power. It is more sensible to concentrate the effort on a few, but meaningful and active projects with multiple members collaborating, than to fan out on a multitude of projects, each not having the manpower for significant research. Second, organizing and leading a research team requires experience and comes with responsibilities and privileges. Therefore, the platform operators have to check for every new project whether these conditions are met for the new project leaders.

To reflect these requirements, the platform implements a two-tiered process for the creation of new projects. First, any member of the platform can propose new research projects on a central hub established for that purpose. There, the proposal is discussed and receives repeated feedback and is improved. The platform operators then select in consultation with the community projects deemed promising. At this stage, it is crucial for the platform operators to give detailed reasoning on why some projects were chosen over others in order to avoid the notion of overruling the community will and frustrating users in the process.

Though the platform is therefore not exclusively limited to citizen science if defined narrowly as seen in 2.1.2, since projects can also be initialized bottom-up by non-scientist volunteers, the platform reflects the hierarchical nature of citizen science for each project.

After the platform succeeds in building a vivid and self-supporting community, active and experienced volunteers could be promoted to a co-operator role, both to further strengthen community involvement as well as relieve administrative work from the original platform operators. This way, a form of peer operation for the platform could be achieved.

3.5.3. Permission Management and User Hierarchy

The openness of a collaborative crowdsourcing platform is often in conflict with other requirements, mostly the relevance of produced content and administrative effort for the platform operators. *Wikipedia* for example has the – for the time of their creation radical – concept of publishing every contribution of any contributor instantly and without editorial supervision, also allowing changes later on in a similar fashion. This freedom is traded off for the inherent possibility of abuse, e.g., vandalization, propaganda or falsified information, making a constant and high editorial and administrative effort necessary. Also, this approach deliberately allows content of low quality or in a draft phase to become – at least for short times – publicly accessible. While these drawbacks are justifiable for an independent entity like the *Wikimedia Foundation*, the operator of *Wikipedia*, repeated abuse as described above can cause permanent damage to a platform's and its operators' reputation and credibility. For an academic department, like the future operators of the platform presented here, this risk is unacceptable.

On the other side of the spectrum, unlocking all submitted contributions and comments on the platform only after explicit approval by the platform operators has several downsides as well: First, it hinders dynamic collaboration; platform operators may not always be available, therefore long delays between submitting a contribution and it becoming available are common. Especially, collaboratively improving a contribution with many small edits in rapid succession is all but impossible. Second, this also has detrimental effects on contributor's satisfaction: As there is a waiting period between submission of a contribution and it becoming available to others, he won't receive immediate feedback for his actions. Third, by running all contributions through the bottleneck of available platform operators, a rapid increase in the platform's popularity and rising number of contributions would surmount the capabilities of an academic department, effectively suffocating the platform by its own success.

Therefore, the concept finally implemented tries to balance both requirements, not hindering collaboration as well as keeping the administrative effort manageable, while at the same time filtering out most abuse or low quality contributions before they become visible to the public. The basic idea is to delegate as much responsibility as possible to the project leaderships, who serve as an intermediate layer between the project members and platform operators in terms of moderation and administration.

It lies within a project's respectively their leadership's power and responsibility to autonomously decide which content will be visible to the public. While this requires a certain level of trust in project leaders, it also creates a clear hierarchy and better control over projects, since there is a clear contact responsible for a project's overall content. In case of repeated transgressions, i.e., spam or low quality contributions becoming repeatedly public, project leaders can be reprimanded or replaced by the platform operators or the project in question closed as a whole.

In order to allow unhindered collaboration, contributions to projects are visible instantly after submission to other project members, but not to non-members or unregistered visitors, for the sake of limiting the impact of abusive contributions.

As problems of scalability can effect large projects in the same ways as the platform as a whole, another user role below project leaders with moderative and unlocking power is present for each project. Project members can therefore either be *basic members*, *project moderators* or project leaders respective *project administrators*.

3.5.4. Collaborative Work on Research Questions

The following chapter will describe in what ways the platform presented in this thesis encourages collaboration on research questions, how communication between users during the collaboration is managed and what mechanisms are employed to ensure high-quality research results.

Resulting from the project structure and the permission management utilized by the platform, collaboration on the platform can happen in two ways: First, higher-order collaboration determining organizational and planning matters on a project-level, and second, collaboration on single contributions. While the first type of collaboration results from the

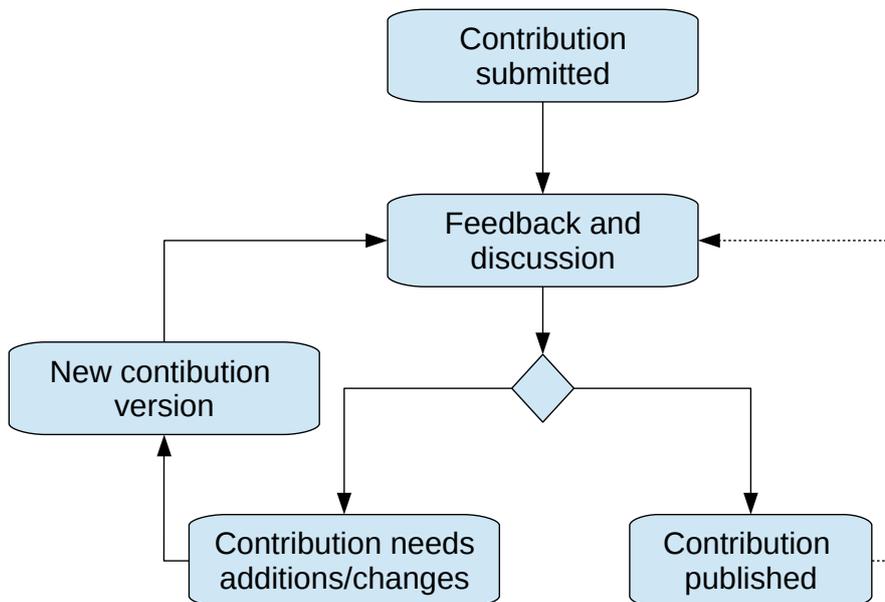


Figure 9: Workflow for new contributions to a project.

intended inclusion of volunteers in most steps of scientific research, as described previously in Chapter 3.5.2, single contributions could also be handled on a one-user-per-contribution basis, effectively disabling dynamic collaboration of other users on contributions. Ownership of a contribution would be clearly linked to a single user, collaboration only happening by the contributor actively accepting advice and input from other users. Many online citizen science platforms, for instance GalaxyZoo or ARTigo, employ this concept for data generation purposes. The validity of a contribution is not checked via other users' feedback, instead, the same question is posed to multiple users and the answers than cross-checked.

Though useful for crowdsourced classification of data in quantifiable ways, this methodology cannot be transferred to the hybrid data-analysis and text-production approach of the platform. As its goal is to ultimately produce article-styled contributions that contain both the research question posed, historical context, data analysis and interpretation, expertise from both art history and data analysis is needed on a single contribution, a hard to find combination of skills in a single researcher, even more so in a volunteer amateur. Therefore, collaboration of multiple users on a single contribution is beneficial in order to complement the different skills each user can present to the platform. To allow this collaboration on single contributions, a simple version control system to allow undoing harmful changes by other users has to be employed.

Nevertheless, the issue of filtering out or improving low-quality contributions is not sufficiently addressed by giving users the optional possibility to collaborate. In fact, collaborative improvement has to be encouraged or to some extent enforced inside the projects. Plainly rejecting premature contributions would be an easy way to ensure only high-quality contributions, but that could lead to a serious waste of potential, a beneficial resource for a

platform such as this, and does not take the continuous improvement aspect aimed for into consideration. Also, as explained in Chapter 3.4.2, instant positive feedback on a user’s action, in this case, him contributing, has positive effects on his motivation. In contrast, as seen by Halfaker et al. [HGMR12] in the case of *Wikipedia*, a dismissive atmosphere directed at (newcomers’) lower-quality contributions and rejection of them without advice for improvement is one of the main reasons in crowdsourcing environments for decline in actively participating users. As a consequence, the platform presented in this thesis puts emphasis on communication and positive feedback from the community during work on contributions instead of simple rejection. An obvious but effective way of putting more emphasis on the communicational aspect of collaboration is employed by the platform: Discussion and feedback is not displayed on a different location away the contribution, but right below it.

To reflect these considerations, the following workflow is encouraged and partly enforced by the platform: Because of the permission hierarchy employed by the platform as explained in Chapter 3.5.3, new contributions to projects can be made by any of its members and are instantly visible to all other project members, but not to outsiders or unregistered visitors. Additionally to one already presented advantage of this organization, i.e., preventing abuse from becoming public, it also facilitates a more playful and experimental atmosphere inside a project, as “failure”, e.g., inconclusive or contradictory findings, won’t be automatically seen by outsiders.

There is no restriction imposed on the extend of the initial version of a contribution submitted. Additional to submitting a complete article-styled contribution, users with less time or no experience in data analysis can for example contribute by formulating an open research question, which then in turn can be picked up by more data-science-savvy users.

After a contribution is submitted, it is passed to a review cycle of alternating feedback, supplied by comments, and additions or improvements made to the contribution. All members of the project can discuss the post, give feedback to the original author and collaboratively improve it. At this point, the original author loses his authority over the contribution, it becomes a project’s property, preventing him from deleting the contribution.

After a consensus is reached about the contribution sufficing the quality standards of the platform, project administrators or moderators publish the contribution, making it publicly visible for all visitors of the platform. At this point, the contribution can be frozen in a revocable manner by the project leaders, disabling further modifications. Figure 9 visualizes a simplified version of the process as a flow-chart. At this point, the contribution’s state can be frozen in a revocable manner, both preserve it and still allowing future modification if the explicit need arises.

Because of the flexible nature of the presented contribution system, it also fills the need for a project-wide communication independent from specific contributions, e.g., project-internal organizational discussions: New contributions inside a project can not only be created to investigate research question, but rather as “meta-contributions”, their content and comments serving as project-wide communication medium.

3.6. Publication of Contributions

As discussed in the previous chapters, the ultimate goal of the platform is to produce contributions publishable to a general audience. Where and how to present them is therefore also an important design decision for the platform.

Since all projects of the platform will share the same basic theme of researching art history through quantitative data science, only accessing published contributions through the projects would make for visitors not accustomed to the project structure finding interesting contributions cumbersome, leading to less visitors in general. Therefore, the platform implements, additional to contributions being presented in a project's area, a central space where all published contributions can be browsed after they have been approved by the project leaders. To further aid navigation, also a search, basic sort and filtering functionality is needed, e.g., for displaying only contributions of a specific volunteer or showing popular contributions with the highest rating count.

Also, since one important motivational factor in citizen science is acknowledgement by scientists, as discussed in 2.2.2, platform operators should have the possibility to highlight particularly noteworthy contributions.

3.7. Integration of Data Analysis Tools

As noted in Chapter 2.5, the most advantageous implementation of existing and future data analysis tools is just as important for the platform's success as the previously discussed concepts. The two fundamental ways for such an implementation would be either the seamless integration into the platform by building the tools into the platform itself or the loose embedding of independent tools into the platform. If focusing on providing a superior and frictionless user experience, the first option would be preferable. However, it comes with significant detriments. The whole ARTigo ecosystem is built on the outdated and discontinued Java framework SEAM. If the tools were to be tightly integrated in the platform, the platform itself would be required to also use SEAM and additionally would have to be built from scratch, or should the platform use another framework, the ARTigo data analysis tools would have to be ported onto the new framework. Either way results in a considerable rise of time and effort needed for implementation.

Therefore, the decision fell on the loose integration of the tools. Not only does this option generate lower implementation costs, it also simplifies the integration of external tools, i.e., tools can be created independent from the implementation of the platform and without access to the source code, making the platform overall more adaptable. For example, it could be possible that external partners, or even citizen scientists themselves, build their own tools to be integrated into the platform; allowing them to use the tools and frameworks best suited for their respective work instead of adapting to the platform's implementation specifics and limitations. The most basic method for such a loose embedding would be to simply add links to external tools, the users then tediously manually copying over the results to the platform when needed. Taking user experience into account, this is a rather undesirable solution. However, the problem can be addressed by implementing a simple

communication interface allowing external applications to forward results to the platform when creating a new contributory post.

Also, the current implementation of the ARTigo search and the Analytics Center do not meet the basic requirement for collaborative use, since both tools lack the capability for persisting their queries and results, preventing users from sharing with each other or publishing them, as already described in more detail at the end of both Chapters 2.5.2 and 2.5.3. In order to keep the amount of modifications on the tools to a minimum, no elaborate session-storage or query-saving mechanisms were implemented for the tools themselves. Instead, the tools send their results as described above to the platform, which then handles persisting them inside the contributions. By modifying the tools to store all needed query-parameters inside the URL, persisting additional information by the tools was avoided. A query can simply be reproduced with the same parameters by visiting the same URL, making sharing it with other users trivial. This also means that future tools do not need to implement a result-storage functionality, reducing their complexity and allowing for easier development. The implementation specifics are described in detail later in Chapter 3.7.

4. Implementation

4.1. Architecture of the Platform

The following sections will discuss the fundamental choices made at the beginning of the implementation phase and give a high-level introduction to the framework on top of which the platform was created. They highlight general development guidelines for the platform and can serve as a starting point for further extensions of the platform.

4.1.1. Adapting a Prebuilt System for the Platform

Implementing any citizen science project always poses one basic question: Should an existing system be extended and adapted for custom needs or should everything be implemented in-house from the ground up. This fundamental decision, which determines the nature of all following implementation work, is commonly referred to as “build versus buy” [PC12].

On the one side, *building* most components oneself brings the advantage of full control over functionality for each component, generally allowing to generate a system closely tailored to the posed challenge, with the trade-off of high development costs, that being either time or money, as even basic functionality has to be build by oneself.

On the other side, taking the *buy* approach, i.e., resorting to an off-the-rack pre-made system – typically *content management systems* (CMS) – by either buying one or using free alternatives, provides already most basic functionality out of the box, e.g., user management or administration. This allocates time and effort to the developer, who can start instantly with adapting and extending the existing system to the projects needs instead of reimplementing basic functionality first. However, this comes at the cost of loosing flexibility when implementing additional features for the platform, as any new feature has to fit into the preexisting framework or sometimes being plainly impossible to implement due to internal restrictions of the CMS.

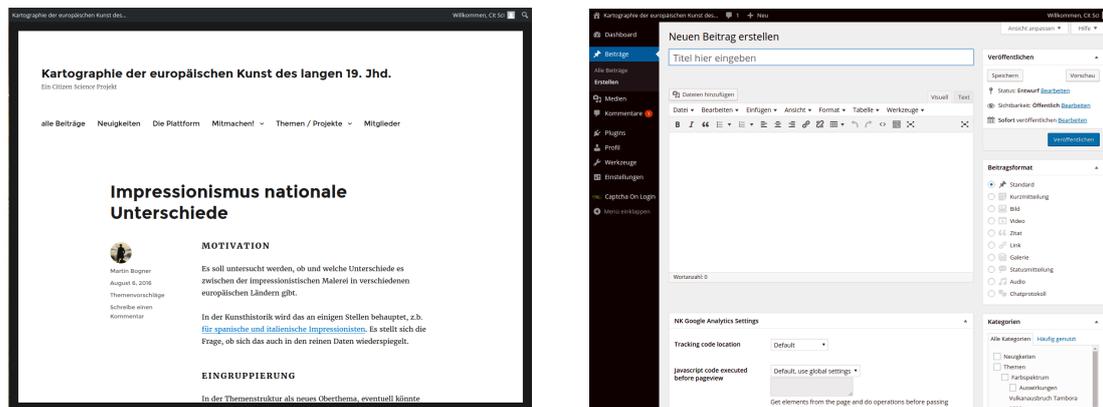
As this thesis is mostly a single-author effort and development time is the most restricting factor in such cases, higher priority was placed on creating a functional system than on utmost flexibility and adaptability. Therefore, this thesis follows the buy approach whenever possible. Because of this, a considerable portion of time used implementing the platform was not spend on developing new features, but also on evaluating preexisting functionalities or different third-party extensions for their potential to realize the concepts presented in Chapter 3. Sometimes, preexisting functionalities had to be deactivated to achieve this goal.

Searching for a both free and well supported system, offering both extendability, long support in the future and an active developer community, the *Wordpress* CMS was chosen as the starting point for further development. Additional to some technical aspects discussed later on, practical considerations made the choice of Wordpress sensible. As constant administration of the platform is needed, choosing a system well known to the prospective administrators is wise. At the Teaching and Research Unit of Programming and Modeling Languages, where the platform will be supervised, several Wordpress installations are already administrated. Therefore, instead of adding a new, foreign system accompanied by

additional operational overhead, the same infrastructure and preexisting know-how can be utilized.

4.1.2. Introduction to the Wordpress Content Management System

While most know Wordpress exclusively as a blogging software, at its core, it is an adaptable and extensible document-centered CMS written in PHP. In their self-representation, the Wordpress team calls it a “semantic publishing platform”.¹² Wordpress is centered around creating, managing and publishing textual content, the so-called *posts*. Most basic and some of the more advanced functionalities needed for the platform, i.e., user and document management, version control, commenting system and administration features, are already implemented in the Wordpress core system. Wordpress provides two distinct areas with different layouts and functionality: First, the area mainly used for presenting published content to end users, the so-called *frontend*, and second, the administration screens or *backend* for creating new content or administration and maintenance purposes. Screenshots of both in a basic Wordpress installation can be seen in Figure 10.



(a) Frontend: Blog entry listing.

(b) Backend: Post editor screen.

Figure 10: Screenshots of typical Wordpress frontend and backend screens.

Wordpress is the most widespread CMS in use on the Internet. Recent numbers from the Web technology survey group W3Techs¹³ suggest that Wordpress powers 59% of all CMS-based websites. It provides a large developer and support community as well as a rich extension ecosystem, which allows to further add many features without writing a single line of code and having pre-built solutions for the most common issues. Wordpress also allows to switch its look-and-feel with a single click in the administration screen by changing its current design package, called *themes* in the Wordpress jargon.

Regarding the implementation of modifications to Wordpress or adding additional features, the Wordpress core system employs a well-documented API, which makes extensible modifications and adaptations possible. The programmatic core architecture of Wordpress is event-driven. At certain points in the execution cycle, Wordpress provides selected points,

¹²https://codex.wordpress.org/WordPress_Features, accessed August 25, 2016.

¹³<https://w3techs.com/>, numbers as at August 25, 2016.

4 IMPLEMENTATION

so-called *hooks*, where custom functionality can be introduced by registering functions to them. Hooks can be either *actions*, whose main purpose it to introduce new functionality, or *filters*, which allow to manipulate specific data during the execution cycle. Using actions or filters, most basic functions of Wordpress can be overridden or modified. Figure 11 visualizes the concept: At an early point during the startup of a Wordpress-based application, extensions can register custom functions to the predefined hooks in the core system. During the life cycle of handling a request, Wordpress calls the registered functions at the given points. Extensions in turn can also define hooks in their own process cycle, in order to allow other extensions to modify their behavior, too.

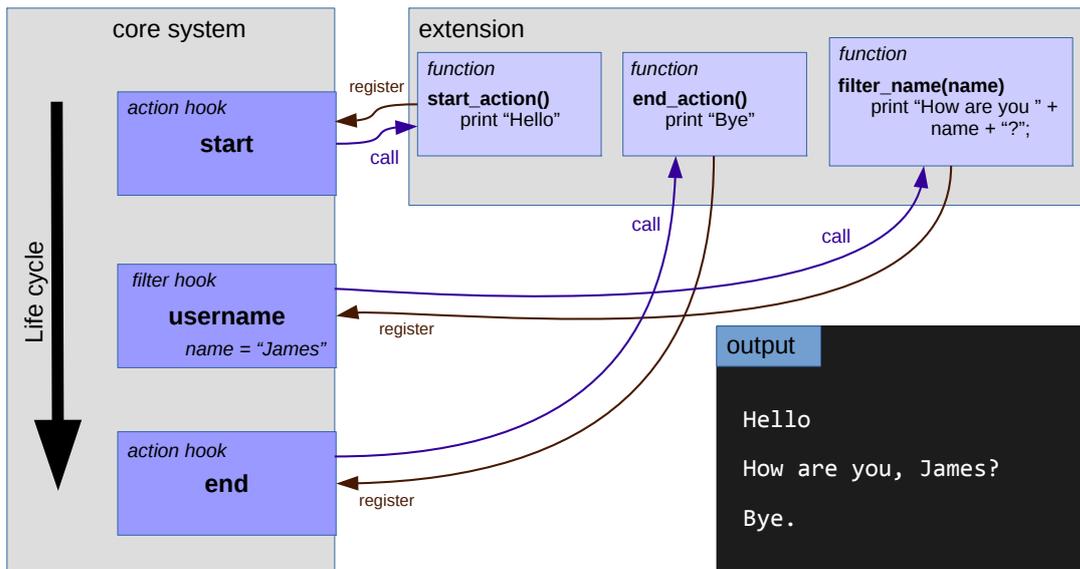


Figure 11: Abstract example for the event-driven architecture employed by Wordpress for extending core functionality.

A risk lies in the wide spread utilization of Wordpress, which does have its benefits as seen above, but comes with a single yet serious drawback: As several million Internet pages use Wordpress, finding security flaws in the Wordpress core and exploiting them is very lucrative. Therefore, Wordpress-based web applications have to always be kept up to date, both the core system, the plugins and the theme, in order to prevent being hacked and compromised. For this purpose, Wordpress already provides update checking and easy-to-use, automated update mechanisms, but it is also wise to install specific monitoring plugins that detect possible hacking attempts and close the most common attack vectors.

4.1.3. Extending Wordpress

In theory, it would be possible to alter the programming code of Wordpress core or other plugins directly without relying on their event-based APIs, but this would impede the possibility of automatically updating to newer versions of Wordpress or the plugins, as made

changes would be overridden with every update. This would either mean manually merging each update into the system, a time-consuming and error-prone process, or fewer updates accompanied by high exposure to the security risks an outdated Wordpress installation represents.

Instead, Wordpress provides in principle two ways for extending its functionality in a maintainable way: By implementing or adapting a theme, or by placing additional functionality into newly-created plugins. Though the distinction is not enforced by Wordpress in a grammatical manner, general Wordpress development guidelines recommend following the principle of separation of concerns, limiting themes to implementing the presentation layer, while adding additional functionality through plugins.

Though Wordpress specifically provides a single editable file for each theme, intended for small changes to a plain Wordpress installation, the extensive changes needed for the citizen science platform presented in this thesis dictate employing the plugin approach for structuring and maintainability reasons. Also, as plugins can be dynamically turned on or off by the site operators, they offer an obvious way to segment features into independent compartments which can be easily turned off if no longer needed without changes to the underlying source code.

Therefore, all modifications made to Wordpress during the course of implementing the platform were done in newly created, dedicated plugins (discernible by the name-prefix “CISI”) without altering the source code of Wordpress core or other 3rd party plugins, each plugin generally responsible for a single functionality. A full list of all employed plugins can be found in Appendix A.4.

To further decrease development effort and harness the design expertise of professionals, in contrast to implementing a new theme, a preexisting one with advanced customization options was selected, the free *Parabola* theme.¹⁴ Due to its popularity and professional backing, a good support and updates in case of backward-incompatible Wordpress core changes can be expected. To fit the specific needs of the platform, the theme was extended where necessary. As changing a theme’s source code would prevent easy updates, the platform employs a Wordpress mechanism called *child theme* for this purpose. The child theme, for obvious reasons named *parabola-child* in the platform, can declare its inheritance from a *parent theme* and overrides or adds specific layout templates and css-styles, without the need to copy the remaining code.

4.2. Collaboration

As seen in Chapter 3.5, collaboration and community functionality stand at the core of the platform. Therefore, much time and effort was spend on selecting an appropriate solution which approximated the envisioned collaboration concept as close as possible. Before settling on a final choice, several alternatives on how to achieve the conceptualized functionalities were explored.

The original plan was to utilize most of the built-in features of Wordpress to emulate the project and contribution structure conceived for the platform. Wordpress in its basic

¹⁴<https://www.cryoutcreations.eu/wordpress-themes/parabola>

4 IMPLEMENTATION

configuration provides article-styled blog posts, categories, a tagging system and a role-based access right management. In the first tested approach, building on top of the post and category system of Wordpress, blog posts should serve as contributions and categories as projects. The limitations soon became apparent and the approach was abandoned thereafter: The role-based access model could not reflect the membership aspect of projects. Users could only be assigned with global roles respective their corresponding global permissions, not on a per-project level. That meant no access-control on a per-project basis was feasible.

A second approach, utilizing Wordpress' so-called *multisite* functionality (multiple independent virtual blogs that share a single Wordpress installation, only linked by an administrative interface for the site operators) eliminated the access-control issue, but came with high administrative and maintenance overhead for the project leaders as each project was represented by a distinct blog with all its configuration options. Additionally, a platform-wide listing of contributions proved complex and imperfect.

Also, Wordpress in its basic installation does not allow to create or modify posts directly from the user frontend. Instead, users have to access the administration screen with its completely different and fixed design and manifold options, as seen previously in in Figure 10, a confusing experience especially for newcomers.

At this point, the decision fell to move away from most of the Wordpress core functionalities and to rely on plugins to provide the projects with access-control and linked contributions. In the following, for each of the required parts of the platform as conceptualized, the selected plugins, their functionality and made modifications will be discussed.

4.2.1. Projects and Community

A review of choices the Wordpress ecosystem presents for plugins enhancing it with further community features revealed the free and open-source plugin *Buddypress* to be the best match for the requirements posed by the platform. It is the most widespread and feature-rich plugin for social network functionalities in the Wordpress ecosystem, currently in user by over 200.000 Wordpress installations and continually maintained and enhanced by an active community. Buddypress consists of several components that allow to augment a Wordpress site with social network functionalities. Of its many optional features, the support of user groups with access-control accompanied by per-group user roles, elaborate user profiles, logging of user activities and direct messages between users are utilized for the platform. Additionally, Buddypress itself features an extension API, allowing plugins to add and modify functionality and layout.

The conceptualized project structure was implemented through the user groups component of Buddypress, “user groups” simply relabeled to “projects”. By default, Buddypress allows all users to create groups which can be joined by other users. Buddypress already implements a group listing and search for easy group navigation and overview, also each group has their own space where content and information of the group are displayed. It supports three distinct group user roles, *administrators* (full control over the project and its settings and members), *moderators* (supervising all group content, e.g., contributions) and *basic members*. These user roles replicate closely the conceptualized user hierarchy in Chapter

3.5.3 and were adopted as is, since further modification on this part was not possible without changes to the BuddyPress source code itself.

In the default configuration of BuddyPress, groups can have one out of three visibility levels: *Public* (group can be accessed and joined freely), *private* (group is visible in group listing, but only accessible by members, membership can only be explicitly granted by group administrators) and *hidden* (same as private, but group is invisible for non-members). As discussed in Chapter 3.5.1, all projects should be open for anyone to join in the initial phase of operating the platform. Therefore, creating projects of type *private* and *hidden* was deactivated, leaving only the *public* project type for the platform.

While BuddyPress provides a chronological list of last activities in the project and allows projects leaders to add a simple project description text, in order to give each project a central hub for presenting itself and also managing collaborative work, a new project overview page was implemented for the platform, shown in Figure 3.5.1: It provides projects with the ability for a more appealing self-representation. In comparison to the previous descriptive text, the new project description also supports more styling options and allows any project member to freely edit it. Additionally to serving as a description, this makes it also useful for in-project communication, e.g., for lists of upcoming or proposed tasks in text form. By giving volunteers clear and concrete actionable steps where to start working at this location, user engagement can be increased. Otherwise, if only vague project goals but no clear options how to contribute are provided, especially new or occasional users can hardly be convinced to participate. Internally, this description is represented as a Wordpress post of a newly defined post-type *groupstext*, each project being linked to one project description. Therefore, the advanced functionalities of Wordpress posts can be reused without reimplementing.

Also, a list of recently active contributions is displayed in the overview in order to provide members with a one-glance overview of recent activities. As new and active contributions are likely to require further assistance, this provides members with a starting point for active participation, similar to the free-form task list as described above.

Last, the overview page also features project statistics, e.g., how often contributions were read. This helps both volunteers to determine if a project is still active and worth their participation as well as provides a mechanism for members to feel part of a greater effort. A more detailed description on how statistics on user activities are displayed will be presented in Chapter 4.4.1.

4.2.2. Article-Based Contributions

As addressed at the beginning of Chapter 4.2, the off-the-rack functionalities of Wordpress did not offer a satisfying approach for implementing the contribution workflow as conceptualized, therefore plugin-based solutions were explored. The following requirements had to be met by any prospective solution: Additional to general ease-of-use, first, it had to provide the possibility of a seamless integration into the project structure as implemented in the previous chapter. Second, all relevant user actions (except for platform operators) had to be limited exclusively to the frontend in order to not alienate and confuse users by the

4 IMPLEMENTATION

Gesamtübersicht **Meine Projektstatistik**

Übersicht

Projektbeschreibung 1

Einführung:
Im Gegensatz zur Kunst der vorangehenden Epochen mit ihrer relativen Einheitlichkeit über längere Zeit hin, ist die Kunst des 19. Jahrhunderts widersprüchlich, schnelllebig und vielgestaltig wie nie zuvor. Das ist Folge der Aufklärung, der Französischen Revolution, der Entmachtung von Königen, Fürsten, Adel, Klerus, des Anspruchs jedes Einzelnen auf Bürger- und Menschenrechte, auf Freiheit.

Es ist Folge von Einsichten, Erfindungen, Verfahrensweisen in der Wissenschaft, der Technik, der Wirtschaft, der Geistesgeschichte, im Sozialen, wie es sie in solcher Fülle nie gegeben hat.

TODO:

- Erste Forschungsfragen formulieren
- Schlagwörter, die mit Nationalismus im Zusammenhang stehen, identifizieren
-

[Editieren][Versionsgeschichte]

Neue Beiträge

- Tambora 29.6.2016 21:07 Uhr
- META: Forschungsfragen für das Projekt 23.6.2016 9:52 Uhr
- Carpe Diem 20.5.2016 20:10 Uhr

Letzte Änderungen

- META: Forschungsfragen für das Projekt 3.9.2016 11:52 Uhr
- Tambora 3.9.2016 11:50 Uhr
- Carpe Diem 12.7.2016 9:57 Uhr

Zuletzt kommentiert

- Carpe Diem 23.5.2016 7:58 Uhr

Mitarbeit angefordert

- Tambora 3.9.2016 11:52 Uhr
- META: Forschungsfragen für das Projekt 3.9.2016 11:52 Uhr
- Carpe Diem 3.9.2016 11:52 Uhr

Projektstatistik aller Mitglieder 3

7 Tage **Monat** 6 Monate Gesamt

Aktivitätsverlauf

Tag	Beitrag gelesen	Beitrag editiert	Beitrag erstellt	erhaltene Bewertungen für Beiträge dieser Gruppe	Kommentar verfasst
So	0	0	0	0	0
Mo	0	0	0	0	0
Di	0	0	0	0	0
Mi	0	0	0	0	0
Do	1	0	0	0	0
Fr	0	0	0	0	0
Sa	2	1	1	0	0

- Beitrag gelesen
- Beitrag editiert
- Beitrag erstellt
- erhaltene Bewertungen für Beiträge dieser Gruppe
- Kommentar verfasst

Figure 12: Project overview page for a single project: 1) editable project description, 2) overview of recently active contributions, 3) project statistics.

complicated backend user interface. Third, it had to provide access-control with different settings for both visitors, project members and administrators.

Fortunately, the preexisting plugin *Buddypress Docs* met all these requirements. Also, as it is widely adopted in the Wordpress community and its author and maintainer is one of the Buddypress lead developers, it offers good documentation and high modifiability as well as assured future support and maintenance. Buddypress Docs allows each registered user of the platform to create text-documents, so-called *docs*, which are internally based on Wordpress posts. Optionally, docs can be linked to a project. Buddypress Docs supports a variety of access levels (everyone, members of linked project, administrators of linked project and author only) for each doc's properties, allowing to independently change permissions on who can read, edit or comment on a doc. Additionally, Buddypress Docs also features a version control system by default, a necessity in collaborative work. As all changes to contributions are logged, their current states can be compared or reset to earlier ones, making it possible to track changes and undo vandalism. Also, all user interactions are limited to the frontend. A screenshot of the frontend editor for docs can be seen in Figure 13.

Based on this features, the platform's contribution system was implemented. As a first step, docs were renamed to *contributions*. Due to the required hierarchic permission model for contributions, as conceptualized in Chapter 3.5, further changes to the default behavior of Buddypress Docs were needed. By default, Buddypress Docs leaves all privacy-level decisions to the original creator of a contribution. This behavior is not suitable for the platform presented in this thesis, as it would allow the original author to delete his contributions to a project at any time, even if most of the subsequent collaborative work on it was done by other users. This was modified so that after a contribution is linked respectively *submitted* to a project, the responsibility for managing the contribution is taken over by the project respectively its leaders. Once a contribution is submitted to a project, the original author cannot make any changes to the privacy settings, delete it or attach it to another project. Also, as contributions should only be publicly accessible after explicit unlocking by project leaders, permissions for basic project members were changed so that all contributions to projects are automatically set to being only visible and editable by the project's members. Buddypress also allows docs without project affiliation. This feature was adapted for the platform to allow users to create private contributions (only accessible by the author himself), useful for storing personal notes or creating drafts before submitting them to a project.

4.2.3. Requesting Assistance

For any collaboration workflow, getting assistance if needed is an important requirement. Therefore, the platform implements a simple feature which allows all members of a project to flag some of its contributions as in need of assistance or rework. While this could be achieved in the simplest form by placing a text notice inside the contribution content, this makes automatic detection and central listing of contributions in need of assistance difficult. For this reason, a single button was added to the contribution view, which toggles the request for assistance.

All articles flagged that way are programmatically added with a notice banner, reminding all readers of the contribution of this fact as well as added to a central listing on their projects' overview page, making it easy to spot articles in need of revision. The flag can be removed by any project member at any time, relying on users to remove it as soon as no further assistance on the contribution is needed anymore.

While this feature is primarily aimed to assist a more efficient collaboration, its effect on user engagement is also noteworthy: It lowers the barrier for project members to actually get involved as the listing of article in need of assistance immediately provides actionable next steps to perform for the volunteers, e.g., initially reading the contribution and assessing the kind of assistance needed. Providing users with several concrete predefined choices for participation results in higher engagement rates than relying on them to actively search for work as seen in Chapter 3.4.2.

4.2.4. Commenting

The comment system used by Wordpress comes with a few restrictions: By default it does not allow users to edit comments after committing them. Also, the comment section is not clearly arranged: Because all comments are on the same level, answers may not seem affiliated to the original comment.

Therefore, the commenting function was extended by introducing the free *wpDiscuz* plugin. In comparison to other advanced commenting plugins, wpDiscuz only extends the existing commenting system of Wordpress, rather than replacing it by one hosted on an external server, which could induce privacy issues. The plugin supports subsequent editing of comments and, as seen in Figure 14, comments answering previous ones are being displayed hierarchical, allowing for discussion among users in the comment section to be more apparent and easier to follow. Also, comments can be sorted by age both descending or ascending, making scrolling to the bottom for the newest comments, as with the default comment system, unnecessary.



Figure 14: Screenshot of the commenting function. Hierarchical comments and editing supported.

4.2.5. Project Suggestions and Meta-Discussions

As discussed in Chapter 3.5.2, the platform is not satisfied with being a tool for scientists to create independent citizen science projects, but tries to accomplish an overarching community and identification across the otherwise independent projects, with volunteers actively helping to improve the platform itself and proposing new, interesting research projects themselves.

Looked at from this perspective, improving the platform both with general feedback as well as new project suggestions is a project by itself. Therefore, the idea is not far to implement it as a specific project, the *project of all projects*, a central hub for making general suggestions to the platform operators and submitting project proposals the same way as any project on the platform. In comparison to implementing new full-blown functionalities for proposing new project suggestions or giving general feedback, the implementation can simply repurpose the project and contribution models and mechanics presented in the previous chapters. The decreased implementation and maintenance effort, a lower complexity of the platform overall and a more consistent user experience made this solution favorable.

As the participation of as much users as possible is desired, in order to form the overarching community as mentioned in the beginning, any user joining the platform also becomes a mandatory member of this meta-project. This way, users also have a way to directly interact with the platform operators as well as other project's contributors. Though BuddyPress itself does not provide the functionality for mandatory project memberships, the plugin *BuddyPress Auto Group Join* adds this feature in a flexible way, allowing for additional mandatory projects later on if necessary.

4.3. Integration of Data Analysis Tools

The following chapter describes the integration of external, preexisting data analysis tools into the platform. The first section shows a general mechanism how tools can be embedded into the platform and how communication between the platform and the tools are handled. The second section describes changes made to adapt the ARTigo search and the Analytics Center for use with the citizen science platform.

4.3.1. Communication Between Tools and the Citizen Science Platform

As outlined in Chapter 3.7, the requirement was to create a loose integration of existing tools while keeping necessary changes to the tools to a minimum. To facilitate getting the results of the data analysis tools into the contributions without manually copying them over, a communication method between tools and citizen science platform was devised.

The tools are reachable via a simple link in the sidebar of the contribution editor, as seen previously in Figure 13, and opened in a new tab. Integrating them via HTML-*iframes* directly into the contribution editor was considered, but ultimately dismissed, as opening them in additional browser tabs or windows allows for simultaneously working with multiple tools at the same time, delegating window management and multitasking simply to the user's browser.

To handle the communication between tools and citizen science platform, an exclusively client-sided approach was implemented, i.e., all necessary processing and communication is done in the user's browser via JavaScript, the tools sending their results through cross-tab communication to the contribution editor tab where the tools' results get appended to the contribution. As no elaborate server-to-server communication is needed, the implementation effort can be kept low both for the tools as well as for the citizen science platform.

Security policies in modern browsers prevent websites from directly modifying content from other domains (the so-called *same-origin policy*), i.e., the tools cannot directly add their results to the contribution editor. Instead, browsers allow communication through a predefined interface via the *window.postMessage* function to a website from another domain (e.g., embedded inside an *iframe*). The target website then can either discard the payload of the message or process it further. This way, data can be sent in the browser safely across domains, e.g., from `http://data-analytics.tool` to `http://citizen-science.platform`.

Therefore, to allow the tools to be hosted on arbitrary domains, a cross-domain cross-tab communication was implemented, relying on two features of HTML5, the previously mentioned *window.postMessage* and *local storage*, a key-value store in the browser where websites can store data for specific domains. As both HTML5 features are supported by all modern browsers, no substantial amount of users is excluded from using the platform. In order to enable the cross-tab communication part, a specific feature of local storage was utilized: Whenever a value is changed in the local storage, an event is sent to all browser tabs or windows from the same domain, notifying them of the changes.

As the only prerequisites, any data analysis tool has to embed a specific (hidden) *iframe* from the domain of the citizen science platform, which contains event handling and processing logic, and some user interface, e.g., a simple button that allows users to trigger the export of results to the citizen science platform. Adapting tools from external websites without the consent of their operators is therefore not possible.

Figure 15 shows the full communication flow as implemented for the platform. When results are exported into the platform, the tool sends a message containing the results in textual form to the *iframe* via the *postMessage* functionality (1). After processing of the event, the event listener inside the *iframe* then stores the payload inside the local storage (2). At this point, the local storage notifies all browser windows and tabs of the *iframe*'s domain of changes, including the contribution editor (3), the contribution editor, subscribed to these type of events, fetches the result data from the local storage (4) and appends it to the contribution content (5).

In the current implementation, the data analysis tools simply send over a representation of the results in HTML form, to keep the initial implementation effort low. Therefore, results do not need to be persisted by the tools themselves, but can be saved in the contributions if they can be represented in text form or are statically available under a specific URL, e.g., linkable images. For additional tools in the future, sending over data in JSON form and processing it further on the platform's side or encoding binary data, e.g., images, in Base64 is also possible in principle, further reducing the need for tools to persist their results themselves.

4 IMPLEMENTATION

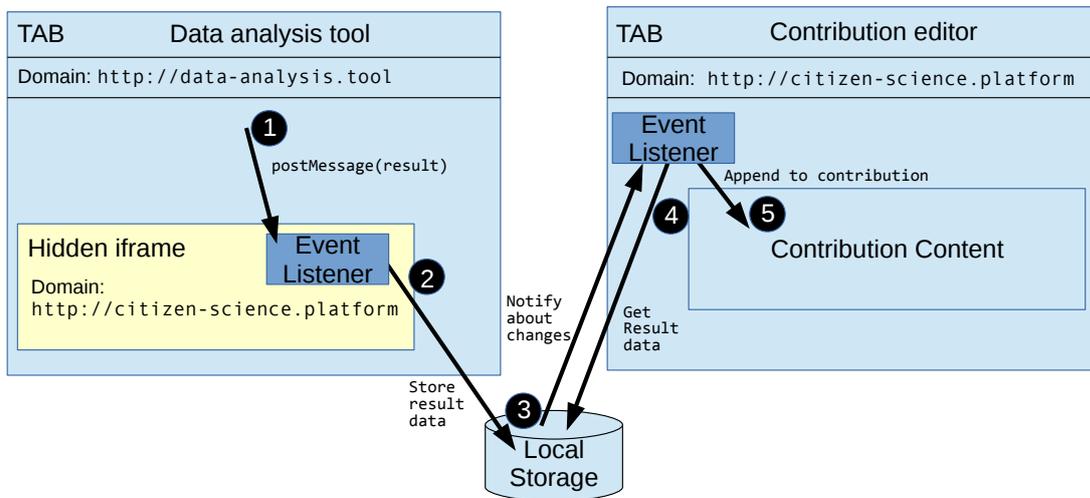


Figure 15: Cross-tab cross-domain communication as implemented by the platform.

4.3.2. Changes to ARTigo Search and Analytics Center

Additionally to implementing the mechanisms for communication as shown in the previous chapter, further changes had to be made in order to allow a meaningful collaborative use of both the ARTigo search and the Analytics Center. As shown in Chapter 2.5.2 and 2.5.3, both tools lacked the capability of providing an easy way to rerun or modify search respective data analysis queries.

To address this, search queries for ARTigo as well as for the Analytics Center were made bookmarkable, i.e., all relevant query parameter stored in the URL. That way, when the URL is copied over and opened on another device, a query with the same parameters will be run.

Since the Analytics Center and the ARTigo search share most of their implementation specifics (both being written in SEAM, a Java EE based web application framework), the approach was identical for both tools: Explicitly defining all query parameters in SEAM's `pages.xml` configuration file, which handles routing of URLs to specific pages, and then changing the pages' logic to retrieve them.

Additionally, the ARTigo search did not allow to link to the artwork images displayed in the search results in a permanent manner. Images received an expiring link and afterwards the images were no longer reachable under the given URL. This was previously implemented in order to prevent statically linking to the artworks from external sites. Exactly this feature was needed for integration of artworks from ARTigo into the contributions of the citizen science platform, i.e., embedding artwork images, tough. While the expiry of links to artworks was deactivated for a version of the search adapted for the citizen science platform, statically linking from external websites can still be prevented by blocking requests from websites not belonging to the search or the citizen science platform. This is achieved by

checking the HTTP referrer-information (i.e., the URL of the website an image is embedded in) by the server for every request.

Some further changes were made to the Center, improving its PCA-visualization by displaying labels for the queried and resulting tags inside the scatterplot. Also, a previously deactivated feature, limiting a query to certain artists (based on artwork-metadata, not tags), was revised and re-enabled to provide a better user interface.

4.4. User Motivation and Engagement

The following chapter will provide implementation details to functionalities previously conceptualized in Chapter 3.4, which are aimed primarily at providing an engaging and motivation environment both for registered users as well as new visitors.

4.4.1. User Profile Area

The user profile area of a single user serves both as a public presentation to other users as well as his own starting point when engaging with the platform. Users are allowed to add profile pictures in order to strengthen their identification and help building a “persona” for the platform. Most of the features of the profile area are provided by the plugins BuddyPress and BuddyPress Docs with little modifications, i.e., private messages between users, browsing contributions either created, edited or rated by the user, and seeing his project memberships. Also, a user’s recent activities, if publicly visible, can be listed chronologically.

Additionally to these built-in features, an overview page was implemented to provide a motivational environment for the user, displaying statistics on his past activities on the platform, as conceptualized in detail in Chapter 3.4.2.

A screenshot of the finalized version of the overview page can be seen in Figure 16. It provides a quantitative summary of his past activities for selectable time frames (past week, past month, past six month and overall) as bar charts. The statistical activity tracking functionality is not taken from BuddyPress but was developed specifically for the platform, as the activity logging of BuddyPress is focused on textual representation of the activities later on, not the generation of statistics. Therefore, a new plugin was implemented, called *CISI user stats*, to provide a simple way of logging user activity and generation of statistics. Activities are logged by type of activity, acting user and related project, allowing to also display activity statistics per project overall and per user per project. The charts are rendered in JavaScript using the free and open-source *chartist.js* library.

4.4.2. Contribution Rating System

The conceptualized rating functionality, as discussed in Chapter 3.4.2, is implemented by using a preexisting plugin, *WP ULike*, which provides a one-tiered rating system in the style of Facebook’s like/unlike functionality. By providing many options for modification of its given layout and a ready integration into BuddyPress, much development time was saved for the platform. As any allusions or resemblance to Facebook should be avoided

4 IMPLEMENTATION

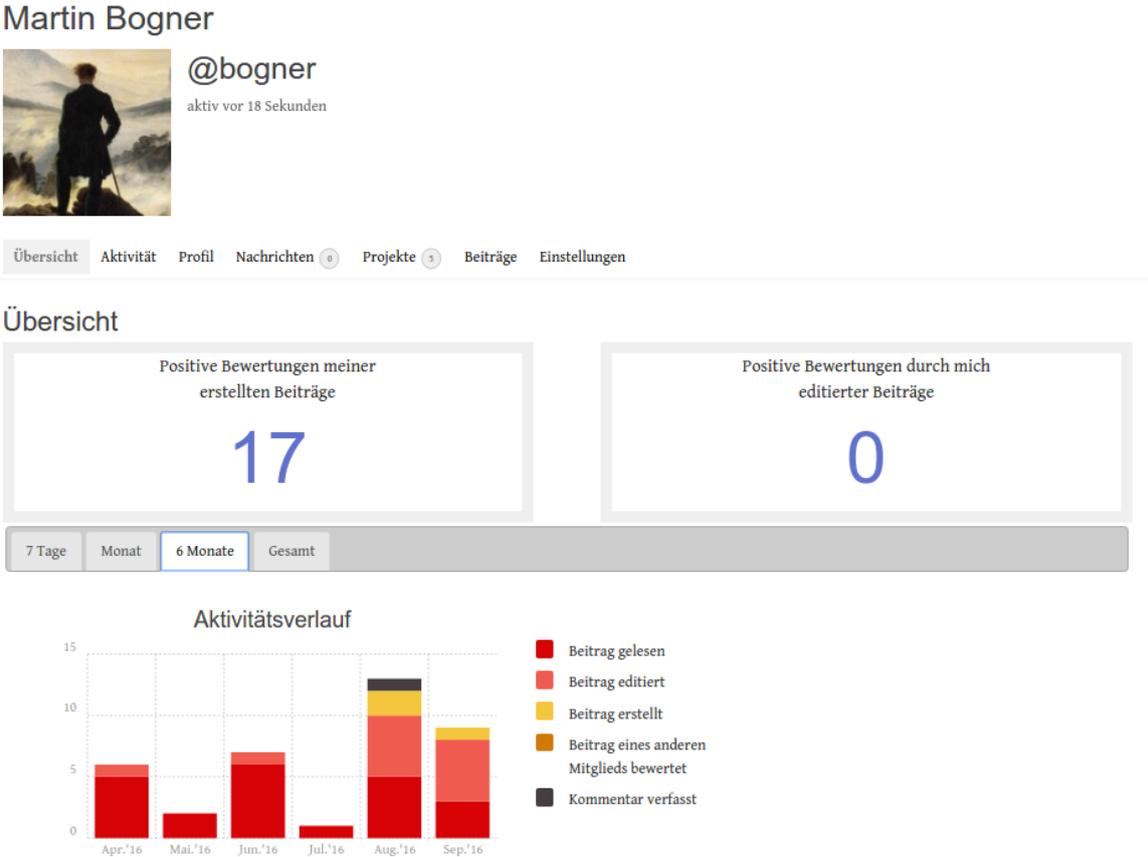


Figure 16: Screenshot of a user's profile area

for the platform, because of their privacy issues and unrelatedness to scientific works, the functionality was relabeled to “rate positive” and “remove rating”.

The plugin was integrated at three points into the platform: First, a rating button was placed on contribution pages, allowing to easily rate contributions when reading them. Also, a list of users who rated the contribution with links to their respective profiles is displayed below, aiding the community aspect of the platform. A screenshot of this can be seen in Figure 17. Second, contributions can be sorted in the contribution overview listings by their rating count, making popular contributions trivial to find. Though this does not necessarily reflect their inherent quality or scientific value, as seen in Chapter 2.4, it helps making interesting contributions better visible on the platform, which in turn attracts more visitors. Third, in the private overview section of each user profile (see Figure 16), the received ratings both for contributions created and edited by the user are displayed. They are kept private in order to avoid competition incentives as discussed in the conceptualization.



Figure 17: Rating box presented for each contribution after the user has rated the contribution. Translation: “Revoke rating”, “Users who like this contribution:”

4.4.3. Try-out Area for Unregistered Visitors

Registration is a big hurdle for most users, especially newcomers. Without obvious advantages or being convinced in the personal use a website can offer, most users cannot be motivated to register. As already discussed Chapter 3.4.1, offering unregistered users the opportunity to create contributions is not maintainable, but still, ways to try out the platform’s contribution and data analysis functionalities have to be given in order to convince users of the qualities of the platform. Therefore, the contribution system of BuddyPress Docs was modified in a way to allow unregistered users to access a special version of the contribution editor. Here, they can test all data analysis tools and import the results into a contribution. Only the last step, saving and submitting a contribution, is disabled. Instead, users who try to do so are redirected with additional notice to the registration page. The intention is to motivate users, who already invested time and want to persist their findings, into registering at this point, a technique known as “upselling” and often employed in marketing [Bus]. After registration, the temporarily stored contribution content is restored and can be saved permanently.

4.4.4. Home Page

According to the conceptualization in Chapter 3.4.1, the home page plays a significant role in activating new visitors for continuous participation and usually is the first page new visitors see. When composing a home page it is therefore important to consider how different designs may influence a user. For example, a text laden home page may be useful to convey all the important information in one step, but it may also intimidate new potential users. As the platform is aimed at art aficionados, a high priority was instead placed on impressing and captivating visitors, by using plenty of artworks throughout the page (as well as on the platform in general) to underline the messages. A screenshot of the home page can be seen in Figure 18.

The homepage consists of several distinct elements arranged under the common header of the platform, which displays a navigational menu. On the top is the call to action, the platform's slogan, telling the visitor the essence of what to experience on the platform. The first thing to catch a visitor's eye, however, is a slider where several (up to 5) featured projects or contributions are displayed. With it, potential users can see prime examples of what to expect, which hopefully intrigues them enough get further interested in the platform, and provides a dynamic to the page. The content is represented by pictures and short, descriptive text and can be changed manually by the platform operators. Below the slider are links to all available projects, to an overview for new users and to the about page for introduction on the platform's operators. The overview provides newcomers with instructions on how the platform can be used and how to participate, while the about page gives information on the background of the platform. Giving detailed information about the platform's goal and operators prominently demonstrates its relevance and builds trust to the platform early on, encouraging more volunteers to participate.

On a technical level, the homepage uses exclusively functions provided by the Wordpress theme in order to save implementation time. Though less flexible for customization, this comes with the benefit of an elaborated user interface in the backend, also enabling platform operators without knowledge in programming and HTML to edit the content displayed on the home page.

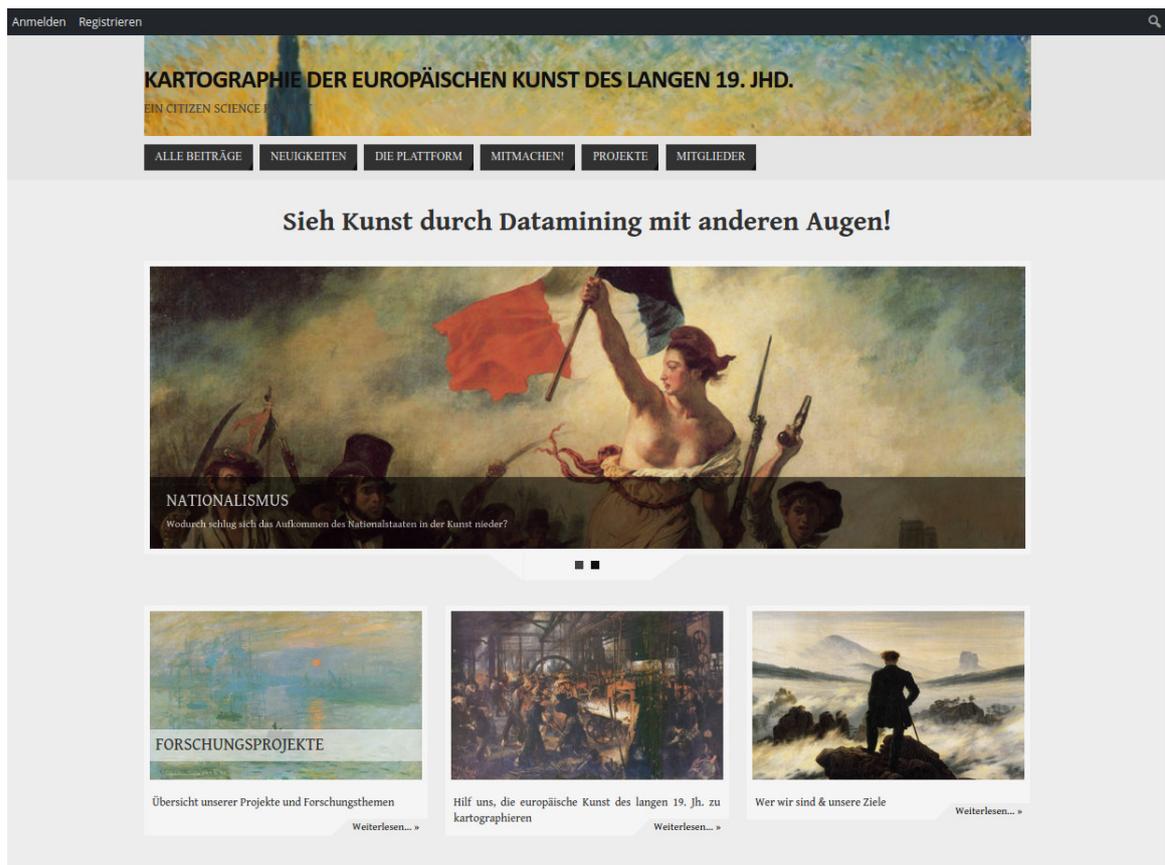


Figure 18: Screenshot of the platform's homepage.

5. Outlook and Conclusion

5.1. Conclusion

The goal of this thesis was to conceptualize and implement a collaborative citizen science platform for data science based on the ARTigo tag data and its pre-existing data-analysis tools. The three prominent requirements for the platform, meaningful collaboration, volunteer motivation and integration of data analysis tools have been given a theoretical background, addressed in a sophisticated conceptualization for the platform and finally implemented based on previous considerations.

The platform facilitates a meaningful collaboration between both professional researchers and amateur volunteers from the disciplines of art history or computer science and statistics, while keeping the required administrative effort low, through structuring the platform in independent projects with attached contributions and hierarchical user permissions inside the projects. It provides a motivating and engaging environment for both newcomers and experienced volunteers through the appliance of conclusions from motivation theory and gamification techniques. Finally, it provides a mechanism for easy integration of both existing or future data analysis tools, allowing for persisting the data analysis results inside the platform, while keeping necessary adjustments for existing tools to a minimum.

5.2. Outlook

When work on the platform started in the course of this thesis, a strong emphasis was placed on creating a system ready for productive use. Therefore, it should not be seen as a final product, but as a first iteration. In the same spirit, many other citizen science projects, instead of preliminary trying to produce an optimal system, rely on an iterative process with active community involvement [CKT⁺10]. Improvements are made in consultation with the community, while the project is in active use. The same is true for the citizen science platform presented in this thesis. The following sections provide an outlook on the future use of the platform as well as starting points for further work.

5.2.1. User Study

Preparing, running and evaluating a user study for the platform takes time. After completion of a first productive version and making it publicly available, at least one year would be needed to properly start the platform and grow a sufficient user base in order to make a meaningful usage analysis. Subsequently, a couple of months would be additionally needed for conducting the user study. In comparison, the time devoted for work on this thesis was six months, used for literary research, conceptualizing and finally implementing the platform. There was no choice but to not conduct a large-term user study.

In contrast, a small-term, small-scale user study would only yield results of limited significance. As the platform is currently still in preliminary testing and not publicly available, volunteers for a user study would have to be recruited via external incentives, for example money or simply as part of their working contract. Since many features of the platform

are centered around intrinsically motivating users to participate, these aspects cannot be tested at this stage. Also, to evaluate the collaborative aspects, a critical mass of users is needed for actually testing workflow and possible issues arising; A typical short-term user study with a small group of participants, viable in the scope of this thesis, would also not cover many administrative and scaling difficulties that encompass a large user base.

5.2.2. Improvements to Existing and Addition of New Data Analysis Tools

An important aspect for data analysis tools, especially if used by layman, is the graphical visualization. Numbers may be precise, but are hard to grasp for the non-expert and in many cases the expert as well. A whole field of study, information visualization, has evolved to research methods for visual representations of abstract data in order to reinforce human recognition of underlying patterns or trends. While the Analytics Center already features a basic PCA-based scatterplot visualization for nearest-neighbor and cluster-expansion queries, much more can be done in this respect.

Currently, all data analysis tools heavily rely on the ARTigo tag data, another possibility to analyze artwork is a computational generation of data for certain aspects of artworks, for example the color spectrum. Color spectrometrics have been already used in art history research, for instance in a study by Zerefos et al. [ZGB⁺07] evaluating the impact of atmospheric effects caused by volcanic eruptions on artists' depictions of sunsets. Such a tool would allow similar research to be performed on the platform.

As the estimated user base consists of many different types of users, many might want to participate or initiate a project but lack the analytic mind, imagination or inspiration needed to do so. In such cases it could prove useful to have the system itself pose research questions. For example, new tools could find significant patterns in the data through automatic data mining algorithms on their own, users then do not have to run the tests themselves but rather evaluate and/or interpret the given result.

In order to give the users a large scope of possibilities for their projects, additional tools may be needed. Finding the right tools may difficult for the site operators, who may not be involved in specific projects and therefore do not know the exact requirements and possibly do not have the time to do so for every request. When implementing the platform, a loose integration of tools was chosen. Therefore, tools written by the community could be integrated in a simple fashion. The only requirement missing is for the community to have open access to the underlying tag data in order to implement these tools. An open, well documented interface to the tag data would solve this problem while following the open science ideal.

5.2.3. Additional Features for the Platform

As previously mentioned, the platform is yet to be made available for the general public. Once active, additional features or changes to existing ones of the platform should only be made in close consultation with the community. Concepts presented in this thesis are a product of consideration between in parts conflicting requirements and have to pass the trial of elongated productive usage, being adapted if necessary.

5 OUTLOOK AND CONCLUSION

Based on the implemented activity tracking feature, an elaborate motivation system can be employed. It is the bedrock for many gamification mechanisms, like unlocking achievements or earning badges, giving the users a quantified evidence for their work. As stated in chapter 2.3, it is however important to not facilitate competition among or ostracism of users. Therefore, said achievements and badges should be private. While this solves one problem, it creates another: How can users searching for support assess the future value of a user's participation in an easy manner, without reviewing all his past contributions in detail? For that purpose user scores could be introduced, i.e., a computed score for the value of his participation per project. These scores could be visualized as a soft color scale on a user's profile page, omitting mention of the exact score altogether. The eschewal of hard numbers would also serve in avoiding competition and counterproductive arguments. Additionally it could be possible for users to set custom goals and reminders, similar to most sports activity trackers. That way, a user can decide for himself what motivates him best and may even assist in keeping up the workflow.

Furthermore, the author of this thesis is aware of his limited experience in user interface design. As of this moment, a complementary bachelor's thesis by Giulianna Dehn is in progress at the same university department, focusing on the user interface aspects of the platform.

5.2.4. Adapting the Platform for Other Citizen Science Purposes

Thinking beyond the scope this thesis, the loose integration of data analysis tools allows them to be easily removed or replaced by different tools. As most of the concepts for user engagement and collaboration implemented for the platform are of a generic nature and not limited specifically to ARTigo, art history or data analysis in general, adapting the platform for other purposes in citizen science or crowdsourcing is also a promising future field of application for the platform. For this reason, Appendix A.1 supplies a list of instructions on how to adapt the platform for other purposes.

5.2.5. Upcoming Usage in Teaching

Most crowdsourcing project face the already mentioned *cold-start* problem: In order to be of any interest for visitors and future contributors, they require interesting and engaging content, as well as a vibrant community. But as content is provided primarily by the user base themselves, none is available at the beginning, which in turn makes the platform unalluring for contribution. To address this problem, by creating some interesting projects and research findings beforehand, and to collect first experiences with the platform, its use in a practical course for graduate students during the next teaching interval is planned.

A. Appendix

A.1. Adapting the Citizen Science Platform for Other Applications

The platform can be adapted for different purposes in a simple way. This guide acts on the assumption that a current installation of the citizen science platform including the database dump provided with this thesis is the starting point. Starting with a pre-filled database is important as Wordpress stores much of its configuration inside the database.

Basic Setup See `CISI_SETUP_AND_DEVELOPMENT.md` located in the root directory of the platform source code directory; see also Appendix A.2.5 for changing the platform's URL

Data Analysis Tools See Appendix A.3 for changing or removing data analysis tools

Home Page See Appendix A.2.3 for changing the home page

Sites Add/edit/remove the descriptive pages, e.g., the about or tutorial pages, to reflect the new purpose of the platform

Projects Start creating new projects, edit existing ones or delete them

A.2. Maintenance and Development Instructions for the Citizen Science Platform

Generally, as the platform is based on Wordpress, all rules from the Wordpress documentation¹⁵ also apply for this platform. Above all, regular updates of Wordpress and the installed plugins is mandatory, as Wordpress is due to its wide adoption in constant focus of hackers and exploit developers.

Major updates to Wordpress and the installed plugins, especially BuddyPress and BuddyPress-Docs, should be tested beforehand on a staging system. Otherwise, they may introduce non-backward compatible changes, breaking some the platform's functionality or layout.

A.2.1. Prerequisites for Server / Development Environment

- PHP version 5.6 or higher
- MySQL version 5.6 or higher, alternatively, MariaDB version 10.0 or higher
- Recent version of Apache Webserver with `mod_rewrite` module enabled

Specific instructions on setting up the citizen science platform are located in `CISI_SETUP_AND_DEVELOPMENT.md` in the source code of the platform

A.2.2. Common Issues after Updating Wordpress or its Plugins

“Projects” are originally called “groups” by buddyPress, this was changed manually by overriding the BuddyPress translation inside the `cisi-custom-plugin-settings` plugin. Therefore,

¹⁵<https://codex.wordpress.org>

A APPENDIX

on any larger version changes of BuddyPress, the chances are that some translations will break. If so, the best solution is to use a diff viewer, import all new changes and replace all mentioning of “group” by “project”. Afterwards, the `.mo` language-file has to be recompiled:

```
msgfmt -o buddypress-de_DE.mo buddypress-de_DE.po
```

A.2.3. Changing the Content of the Home Page

The homepage settings are located in the Design settings in the Wordpress admin area, found in the administration area in:

(**Appearance** | **Parabola Settings** | **Presentation Page**)

A.2.4. Adding a Group with Mandatory Membership

Adding a mandatory project can be achieved by first creating a new project (or taking an existing one) and adding it to the “auto join” list. The settings-option is located in the administration area in:

(**BuddyPress** | **Auto Group Join**)

A.2.5. Moving the Platform to a Different Server and URL

For moving the platform to a different server and/or its URL, all steps from the general Wordpress documentation¹⁶ apply as well. Performing a backup of the Wordpress installation and its database is strongly advised before undertaking any further steps. As Wordpress stores many of its configuration options in the database, replacing all references to the old URL is a must.

Additionally, as tools have to integrate a specific page from the platform for communication purposes, as described in Chapter 4.3.1, the referencing URL of the hidden iframe has to be changed as well.

- The iframe for the ARTigo search is embedded in:
`view/artigo/cisi/cisiShareLogic.xhtml`
- For the Analytics Center in:
`view/AnalyticsCenterComponentViews/cisiShareLogic.xhtml`

A.3. Adding, Moving and Removing Data Analysis Tools

The logic for adding the Analytics Center and the ARTigo search integration to the platform is handled by the *CISI Data Analysis Tool Communication* plugin. To add a new data analysis tool, the following steps have to be taken:

- Add export-logic analog to the Analytics Center or the ARTigo search to your data analysis tool: Embed `cisi-message-iframe.html` as a hidden iframe into the tool

¹⁶https://codex.wordpress.org/Moving_WordPress

A.4 Used Wordpress Plugins and Their Functionality

and send the results via `window.postMessage()` to it; see the files named in A.2.5 for reference

- Add message-event handling logic to:
`tools-communication/cisi-message-iframe.html` located in the application root directory, and store data to local storage
- Add storage-event handling logic to:
`cisi-data-analysis-tool-communication/js/data_analysis_tools_communication.js`
- Add the link to the tool to the editor sidebar, examples can be found in:
`cisi-data-analysis-tool-communication/includes/editor_sidebar_integration.php`

For removing a tool, apply the same steps in reverse. For deactivating all tool integration, simply deactivate the plugin. If a tool is moved, i.e., its domain changed, the corresponding entry in `cisi-message-iframe.html` has to be edited as well.

A.4. Used Wordpress Plugins and Their Functionality

Name	Type	Functionality
Akismet	Security, Administration	Automatic spam protection
Buddypress	Collaboration, Community	Provides most of the community functionalities of the platform, i.e., user profiles, projects structure and displaying last user activities
Buddypress Docs	Collaboration Contribution	Provides most of the contribution functionality
BuddyPress Group Email Subscription	Collaboration	Allows users to receive e-mail notification for their projects
BuddyPress Auto Group Join	Administration	Enforces mandatory membership in certain projects
BuddyPress Restrict Group Creation	Administration	Allows platform operators to specify certain conditions a user must meet, e.g., its global user type, in order to create a new project
BP Group Announcements	Collaboration	Allows project moderators and administrators to post announcements on the project's page
Captcha on Login	Security	Adds a mandatory captcha field to the login page to prevent brute force attacks
CISI Custom Plugin Settings	Core functionality	Modifications to other plugins, e.g., front-end editor-field size, layout changes and minor tweaks

A APPENDIX

CISI Data Analysis Tool Communication	Core functionality	Adds the communication framework for tools to send data to the editor
CISI Group Description	Core functionality	Adds the editable project description with revision control to a project
CISI Group Likes Integration	Core functionality	Implements the sorting of projects by their received likes
CISI Group Overview	Core functionality	Adds the overview page (new contributions, last activities, statistics) to the projects
CISI Help Wanted	Core functionality	Allow users to set a notice on contributions, explicitly requesting other members to participate
CISI Redirect on logout	Core functionality	Redirects users back to the home page after they log out instead of leaving them on the login screen
CISI Restrict Admin-Dashboard Access	Core functionality	Blocks unneeded access to the admin area for basic registered users; Wordpress allows them by default to enter the admin area, which would confuse most users as it has an unfamiliar layout compared to the main site
CISI Toolbar	Layout	Customizations to the default Wordpress toolbar displayed on top of the page. Mostly removes references to Wordpress, its logo, and the admin area.
CISI User Stats	Core functionality	Tracking of user actions on the platform and displaying user and project statistics
CISI WP ULike	Core functionality	Fork of the WP ULike plugin; allows users to rate contributions and projects.
Comments wpDiscuz	Core functionality	Adds a more advanced comment system, mainly allowing users to edit their comments after submission.
Cryout Serious Theme Settings	Administration	Provides settings for the Parabola theme used for the platform's layout
Email Obfuscator	Security	Obfuscates e-mail addresses mentioned on the platform in order to prevent spam
Google Analytics Dashboard for WP	Administration	Integrates Google Analytics for statistics and tracking purposes into the platform
Mail On Update	Security	Sends an e-mail notification to platform operators when new versions of plugins are available.
Revisr	Development	Git version control integration into the administration area

Plugin Vulnerabilities	Security	Displays alerts for known vulnerabilities of plugins
TinyMCE Advanced	Layout, Functionality	Advanced options for the text editor for contributions, for example displaying the html source
Update Control	Security	Toggles automatic updates for Wordpress
UpdraftPlus Backup/Restore	Administration	Allows automatic backups of Wordpress System, uploads and database.
White Label CMS	Layout	Removes references to Wordpress from front-end, allows setting custom logos
Wordfence Security	Security	Provides firewall and IP blocking of attackers, traffic analysis and other security features
WP Sync DB	Administration, Development	Allows easy mirroring of the database for backup or local development.
WP DB Manager	Administration, Maintenance	Adds access to the database in the administration area and automatic backup scheduling.

A.5. Amount of Source Code Written

A.5.1. Citizen Science Platform

Counting the lines of code (LOC) written for this thesis (excluding blank lines) for all additions to the basic Wordpress installation, done in the *CISI plugins* and the *parabola-child* theme, it totals in approximately **3430 LOC**.

In comparison, the full current installation of the citizen science platform, including the Wordpress core and third party plugins, contains roughly 800.000 LOC.

These numbers were computed with the open-source tool *sloc*,¹⁷ excluding readme and translation files, and additionally excluding third-party JavaScript libraries when necessary:

```
sloc -e "*.po|.*mo|.*txt|.*md" <selected-cisi-plugin-dir>
```

A.5.2. ARTigo Search

Using the *git* version control system's statistic functionality reveals that 593 LOC have been added or edited in the course of implementing the communication with the citizen science platform as well as the needed changes to provide bookmarkable URLs for search queries and static links to images. Subtracting the 141 lines of *.xhtml* template code, simply duplicated to create a second version of the search specifically for the citizen science project (without

¹⁷<https://github.com/flosse/sloc>

A APPENDIX

changing the existing one), leaves **452 LOC** added or changed. The exact command used to count the changed LOC can be seen below.

```
git log -w --numstat --pretty="%H" <start-of-work-commit>..<final-commit> |  
    awk 'NF==3 {plus+=$1; minus+=$2} END {printf("+%d, -%d\n", plus, minus)}  
    }
```

A.5.3. Analytics Center

Using the same method for counting the changed LOC in the Analytics center, reveals that **1732 LOC** were added or changed. It should be noted that the actual amount of changes necessary for the features discussed in Chapters 3.7, 4.3.1 and 4.3.2 were well below that number. Additionally to the work on preparing the Analytics Center for integration into the citizen science platform, some existing classes and methods were refactored along the way in order to provide a better structured and maintainable code base.

References

- [AH10] Abdulmonem Alabri and Jane Hunter. Enhancing the quality and trust of citizen science data. In *Sixth International Conference on eScience*. IEEE, 2010.
- [ALH⁺00] Paul V. Adams, Erick D. Langer, Lily Hwa, Peter N. Stearns, and Merry E. Wiesner-Hanks. *Experiencing world history*. NYU Press, 2000.
- [AS13] Georgios Askalidis and Greg Stoddard. A theoretical analysis of crowdsourced content curation. In *The 3rd Workshop on Social Computing and User Generated Content*, 2013.
- [BBJ⁺09] Rick Bonney, Heidi Ballard, Rebecca Jordan, Ellen McCallie, Tina Phillips, Jennifer Shirk, and Candie C. Wilderman. 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*, 2009.
- [Ben06] Yochai Benkler. *The wealth of networks: How social production transforms markets and freedom*. Yale University Press, 2006.
- [Ber] Berkeley SETI Research Center. *SETI@HOME*. http://setiathome.berkeley.edu/sah_about.php (accessed June 09, 2016).
- [BHH⁺13] Anne Bowser, Derek Hansen, Yurong He, Carol Boston, Matthew Reid, Logan Gunnell, and Jennifer Preece. Using gamification to inspire new citizen science volunteers. In *Proceedings of the first international conference on gameful design, research, and applications*, pages 18–25. ACM, 2013.
- [BHP13] Anne Bowser, Derek Hansen, and Jennifer Preece. Gamifying citizen science: Lessons and future directions. In *Workshop on Designing Gamification: Creating Gameful and Playful Experiences*, 2013.
- [BML05] Steve Benford, Carsten Magerkurth, and Peter Ljungstrand. Bridging the physical and digital in pervasive gaming. *Communications of the ACM*, 48(3):54–57, 2005.
- [BN06] Yochai Benkler and Helen Nissenbaum. Commons-based peer production and virtue. *Journal of Political Philosophy*, 14(4):394–419, 2006.
- [Bol00] Paul Boltwood. An amateur astronomer’s experiences with amateur-professional relations. In *Amateur-Professional Partnerships in Astronomy*, volume 220, pages 188–195, 2000.
- [Bra13] Daren C. Brabham. *Crowdsourcing*. The MIT Press Essential Knowledge series. MIT Press, 2013.
- [Bus] BusinessDictionary.com. *upselling*. <http://www.businessdictionary.com/definition/upselling.html>(accessed August 31, 2016).
- [BW12] François Bry and Christoph Wieser. Squaring and scripting the esp game: Trimming a gwap to deep semantics. In *International Conference on Serious Games Development and Applications*, pages 183–192. Springer, 2012.

References

- [CKT⁺10] Seth Cooper, Firas Khatib, Adrien Treuille, Janos Barbero, Jeehyung Lee, Michael Beenen, Andrew Leaver-Fay, David Baker, Zoran Popović, et al. Predicting protein structures with a multiplayer online game. *Nature*, 466(7307):756–760, 2010.
- [Daw12] Diane DeDe Dawson. Open science and crowd science: Selected sites and resources. *Issues in Science and Technology Librarianship*, 2012.
- [DDKN11] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, pages 9–15. ACM, 2011.
- [Dou08] Michael Douma, editor. *Pigments through the ages*. <http://www.webexhibits.org/pigments>, 2008. (Accessed August 25, 2016).
- [EAGLdG12] Enrique Estellés-Arolas and Fernando González-Ladrón-de Guevara. Towards an integrated crowdsourcing definition. *Journal of Information science*, 38(2):189–200, 2012.
- [EJLC13] Alexandra Eveleigh, Charlene Jennett, Stuart Lynn, and Anna L. Cox. “I want to be a Captain! I want to be a Captain!”: Gamification in the Old Weather citizen science project. In *Proceedings of the first international conference on gameful design, research, and applications*, pages 79–82. ACM, 2013.
- [FS14] Chiara Franzoni and Henry Sauermann. Crowd science: The organization of scientific research in open collaborative projects. *Research Policy*, 43(1):1–20, 2014.
- [FSE03] Judy Foster-Smith and Stewart M. Evans. The value of marine ecological data collected by volunteers. *Biological Conservation*, 113(2):199–213, 2003.
- [GHL⁺14] Anita Greenhill, Kate Holmes, Chris Lintott, Brooke Simmons, Karen Masters, Joe Cox, and Gary Graham. Playing with science: gamified aspects of gamification found on the online citizen science project-zooniverse. In *15th International Conference on Intelligent Games and Simulation*, pages 15–24, 2014.
- [Gol97] Michael H. Goldhaber. The attention economy and the net. *First Monday*, 2(4), 1997.
- [Gre09] François Grey. Viewpoint: The age of citizen cyberscience. *Cern Courier*, 29, 2009.
- [Hak13] Muki Haklay. Citizen science and volunteered geographic information: Overview and typology of participation. In *Crowdsourcing geographic knowledge*, pages 105–122. Springer, 2013.
- [HGMR12] Aaron Halfaker, R. Stuart Geiger, Jonathan T. Morgan, and John Riedl. The rise and decline of an open collaboration system: How wikipedias reaction to popularity is causing its decline. *American Behavioral Scientist*, page 0002764212469365, 2012.

- [Hob10] Eric Hobsbawm. *Age of Empire: 1875-1914*. Hachette UK, 2010.
- [Hoi14] Florian Hoidn. The Analytics Center: Devising a citizen science data mining tool for the ARTigo image tagging project. Bachelor’s thesis, Institute of Informatics, Ludwig Maximilians University of Munich, 2014.
- [How06] Jeff Howe. The rise of crowdsourcing. *Wired magazine*, 14(6):1–4, 2006. <http://www.wired.com/2006/06/crowds/> (accessed September 5, 2016).
- [IJCTC13] Ioanna Iacovides, Charlene Jennett, Cassandra Cornish-Trestrail, and Anna L. Cox. Do games attract or sustain engagement in citizen science?: a study of volunteer motivations. In *CHI’13 Extended Abstracts on Human Factors in Computing Systems*, pages 1101–1106. ACM, 2013.
- [KR96] Tim Kasser and Richard M. Ryan. Further examining the american dream: Differential correlates of intrinsic and extrinsic goals. *Personality and social psychology bulletin*, 22(3):280–287, 1996.
- [LSS05] Sonja Lyubomirsky, Kennon M. Sheldon, and David Schkade. Pursuing happiness: The architecture of sustainable change. *Review of general psychology*, 9(2):111, 2005.
- [LW79] Bruno Latour and Steve Woolgar. Laboratory life: The social construction of scientific facts. *Beverly Hills: Sage*, 1979.
- [McG08] Jane McGonigal. Engagement economy: The future of massively scaled collaboration and participation. *Institute for the Future*, 2008.
- [McG11] Jane McGonigal. *Reality is broken: Why games make us better and how they can change the world*. Penguin, 2011.
- [NAA11a] Oded Nov, Ofer Arazy, and David Anderson. Dusting for science: motivation and participation of digital citizen science volunteers. In *Proceedings of the 2011 iConference*, pages 68–74. ACM, 2011.
- [NAA11b] Oded Nov, Ofer Arazy, and David Anderson. Technology-mediated citizen science participation: A motivational model. In *ICWSM*, 2011.
- [NAA14] Oded Nov, Ofer Arazy, and David Anderson. Scientists@Home: What drives the quantity and quality of online citizen science participation? *PloS one*, 9(4), 2014.
- [Nat] National Audubon Society. *History of the Christmas Bird Count*. <http://www.audubon.org/content/history-christmas-bird-count> (accessed August 15, 2016).
- [Nic12] Scott Nicholson. A user-centered theoretical framework for meaningful gamification. *Games+ Learning+ Society*, 8(1):223–230, 2012.
- [Nic15] Scott Nicholson. A recipe for meaningful gamification. In *Gamification in education and business*, pages 1–20. Springer, 2015.
- [NWC⁺12] Greg Newman, Andrea Wiggins, Alycia Crall, Eric Graham, Sarah Newman, and Kevin Crowston. The future of citizen science: emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment*, 10(6):298–304,

References

- 2012.
- [Pal07] Sylvia Paletschek. Zurück in die Zukunft?: Universitätsreformen im 19. Jahrhundert. In Wolfgang Jäger, editor, *Das Humboldt-Labor: Experimentieren mit den Grenzen der klassischen Universität*, pages 11–15. Albert-Ludwigs-University Freiburg, 2007. (In German).
- [PC12] Nathan R. Prestopnik and Kevin Crowston. Citizen science system assemblages: understanding the technologies that support crowdsourced science. In *Proceedings of the 2012 iConference*, pages 168–176. ACM, 2012.
- [Rae06] Joost Raessens. Playful identities, or the ludification of culture. *Games and Culture*, 1(1):52–57, 2006.
- [RBG⁺10] M. Jordan Raddick, Georgia Bracey, Pamela L. Gay, Chris J. Lintott, Phil Murray, Kevin Schawinski, Alexander S Szalay, and Jan Vandenberg. Galaxy zoo: Exploring the motivations of citizen science volunteers. *Astronomy Education Review*, 9(1), 2010.
- [RBG⁺13] M. Jordan Raddick, Georgia Bracey, Pamela L. Gay, Chris J. Lintott, Carie Cardamone, Phil Murray, Kevin Schawinski, Alexander S. Szalay, and Jan Vandenberg. Galaxy Zoo: Motivations of citizen scientists. *arXiv preprint arXiv:1303.6886*, 2013.
- [RD00] Richard M. Ryan and Edward L. Deci. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, 25(1):54–67, 2000.
- [RHP⁺14] Dana Rotman, Jen Hammock, Jenny Preece, Derek Hansen, Carol Boston, Anne Bowser, and Yurong He. Motivations affecting initial and long-term participation in citizen science projects in three countries. *iConference 2014 Proceedings*, 2014.
- [RP13] Hauke Riesch and Clive Potter. Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. *Public Understanding of Science*, 2013.
- [RPH⁺12] Dana Rotman, Jenny Preece, Jen Hammock, Kezee Procita, Derek Hansen, Cynthia Parr, Darcy Lewis, and David Jacobs. Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 217–226. ACM, 2012.
- [SBW⁺12] Jennifer L. Shirk, Heidi L. Ballard, Candie C. Wilderman, Tina Phillips, Andrea Wiggins, Rebecca Jordan, Ellen McCallie, Matthew Minarchek, Bruce V. Lewenstein, Marianne E. Krasny, et al. Public participation in scientific research: a framework for deliberate design. *Ecology and Society*, 17(2):29, 2012.
- [Sch14] Corina Schemainda. *Qualitative Analysis of the ARTigo Gaming Ecosystem*. Bachelor’s thesis, Institute of Informatics, Ludwig Maximilians University of Munich, 2014.

- [Sch15] Maximilian Schich. Figuring out art history. *arXiv preprint arXiv:1512.03301*, 2015.
- [SDW06] Matthew J. Salganik, Peter Sheridan Dodds, and Duncan J. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *science*, 311(5762):854–856, 2006.
- [SF14] Henry Sauermann and Chiara Franzoni. Participation dynamics in crowd-based knowledge production. In *Academy of Management Proceedings*. Academy of Management, 2014.
- [SFZ14] Hendrik Send, Sascha Friesike, and Ayca Nina Zuch. Participation in on-line co-creation: Assessment and review of motivations. *HIIG Discussion Paper Series Discussion Paper 2014-01*, 1, 2014.
- [Sig15] Franz Siglmüller. Advanced user interface for artwork search result presentation. Bachelor’s thesis, Institute of Informatics, Ludwig Maximilians University of Munich, 2015.
- [Sil09] Jonathan Silvertown. A new dawn for citizen science. *Trends in ecology & evolution*, 24(9):467–471, 2009.
- [Sim13] Tom Simonite. The decline of Wikipedia. *Technology Review*, 116(6):50–56, 2013.
- [Sno59] Charles Percy Snow. *The two cultures and the scientific revolution: The Rede Lecture*. Cambridge University Press, 1959.
- [Ste10] Bartholomäus Steinmayr. *Designing Image Labeling Games For More Informative Tags*. Diploma thesis, Institute of Informatics, Ludwig Maximilians University of Munich, 2010.
- [Stö12] Florian Störkle. *Combino – a GWAP for generating combined tags*. Bachelor’s thesis, Institute of Informatics, Ludwig Maximilians University of Munich, 2012.
- [Sto15] Greg Stoddard. Popularity dynamics and intrinsic quality in reddit and Hacker News. In *Ninth International AAAI Conference on Web and Social Media*, 2015.
- [VA06] Luis Von Ahn. Games with a purpose. *Computer*, 39(6):92–94, 2006.
- [VAD04] Luis Von Ahn and Laura Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 319–326. ACM, 2004.
- [VAD08] Luis Von Ahn and Laura Dabbish. Designing games with a purpose. *Communications of the ACM*, 51(8):58–67, 2008.
- [VD09] José Van Dijck. Users like you? theorizing agency in user-generated content. *Media, culture, and society*, 31(1):41–58, 2009.
- [WBBL13] Christoph Wieser, François Bry, Alexandre Bérard, and Richard Lagrange. Artigo: building an artwork search engine with games and higher-order latent semantic analysis. In *First AAAI Conference on Human Computation and*

References

- Crowdsourcing*, 2013.
- [WBI05] Candie C. Wilderman, Alissa Barron, and Lauren Imgrund. Top down or bottom up? allarms experience with two operational models for community science. In *Proceedings of the 4th National Monitoring Conference*, 2005.
- [WC11] Andrea Wiggins and Kevin Crowston. From conservation to crowdsourcing: A typology of citizen science. In *System Sciences (HICSS), 2011 44th Hawaii international conference on*, pages 1–10. IEEE, 2011.
- [Wie14] Christoph Wieser. *Building a semantic search engine with games and crowdsourcing*. Doctoral thesis, Institute of Informatics, Ludwig Maximilians University of Munich, 2014.
- [Win94] Wilhelm Windelband. *Geschichte und Naturwissenschaft. Rede zum Antritt des Rectorats der Kaiser-Wilhelms-Universität Strassburg, geh. am 1. Mai 1894*. Heitz, 1894. (In German).
- [Wro08] Luke Wroblewski. Sign up forms must die. *A List Apart*, 255, March 25, 2008. <http://alistapart.com/article/signupforms> (accessed August 15, 2016).
- [You10] Jeffrey R. Young. Crowd science reaches new heights. *The Chronicle of Higher Education*, May 28, 2010. <http://chronicle.com/article/The-Rise-of-Crowd-Science/65707/> (accessed September 5, 2016).
- [ZGB⁺07] C. S. Zerefos, V. T. Gerogiannis, D. Balis, S. C. Zerefos, and A. Kazantzidis. Atmospheric effects of volcanic eruptions as seen by famous artists and depicted in their paintings. *Atmospheric Chemistry and Physics*, 7(15):4027–4042, 2007.