Secretaries at Work: Accessing Astrid Lindgren's Stenographed Manuscripts through Expert Crowdsourcing

Karolina Andersdotter [1,2] and Malin Nauwerck [3,4]

[1] Åbo Akademi University, Tuomiokirkontori 3, FI-20500 Turku, Finland
[2] Uppsala University Library, Box 510, 751 20 Uppsala, Sweden
[3] The Swedish Institute for Children's Books, Odensgatan 61, 113 22 Stockholm, Sweden
[4] Uppsala University, Box 256, 751 05 Uppsala, Sweden

Abstract

The digitisation of cultural heritage collections has made large parts of our digital heritage available online. However, the collections have often been difficult to access in a meaningful way, still requiring item by item handling of digital images of text to decipher manuscripts and printed materials due to e.g. limited OCR (optical character recognition)/HTR (handwritten text recognition) capabilities or insufficient metadata. While these technologies are under rapid development, in some cases they require training data to learn both machine and handwritten texts, and in some cases it just makes more sense to transcribe texts manually.

Enter crowdsourcing: a method where a crowd of people is involved to transcribe, describe or otherwise enrich the digital heritage collections with data [1]. However, the labour and cost efficiency of crowdsourcing in a cultural heritage context has been questioned [2] – is the quality of the crowdsourced results worth the investment in launching and running a crowdsourcing project?

The Astrid Lindgren Code project [3] explores Swedish author Astrid Lindgren's original manuscripts in Melin shorthand (stenography). Lindgren's stenography has for long been considered "undecipherable" [4, 5] and has therefore never been subjected to study, making manual interpretation the only existing possibility of accessing the material as well as providing training data for future research [6].

Nevertheless, crowdsourcing has proven to be unexpectedly successful in producing transliterations of Lindgren's stenographed notepads. With 170 volunteers signing up for decoding, prolific attempts during the Spring of 2021 have resulted in a full transliteration of the drafts to novel *The Brothers Lionheart* (1973) in approximately five weeks.

This paper presents the method development securing this successful crowdsourcing process, focusing on the importance of joint ownership, planned communication efforts, and community building through online hackathons. The paper also considers how the particularly challenging circumstances of a pandemic year might have contributed to the avid response from a crowd that normally might have lacked the confidence and time to participate. Transliterating stenography is a particular skill, situated in time and associated with the profession of the former secretary. While substantially limiting the recruitable crowd of volunteers, the paper argues that the requiring of expert skill has been central in the success and the methodological development of the project.

---

1. Introduction

Astrid Lindgren (1907–2002) is a world-famous Swedish author whose writing method was shorthand. More than 670 shorthand notepads containing all of Lindgren's literary production are preserved in the Astrid Lindgren Archives and at The Swedish Institute for Children's Books. Up until now, Lindgren's original manuscripts have never been the subject of research, mainly because "no one has managed to decrypt Lindgren's stenographic code", as literary scholar Vivi Edström once put it [2], and this has become an acknowledged truth in Lindgren reception.[1]

Although true for most people, the indecipherability of Lindgren's shorthand does not apply to everyone. The premise for the ongoing digital humanities project The Astrid Lindgren Code (2020–2022) [3] is that Lindgren's manuscripts can indeed be read by those with knowledge of the Melin system of shorthand which Lindgren practised.

The mixed methods approach for accessing Lindgren's shorthand includes HTR experiments, applied and developed simultaneously with collective transliteration[2] through volunteer expert crowdsourcing [6].[3] At this stage, the primary material mainly consists of 52 digitised notepads containing drafts and manuscripts to the fantasy novel *The Brothers Lionheart* (1973).

This paper addresses the process of securing high quality transliteration of Lindgren's shorthand notepads through citizen science. Our focus is the recruiting, retaining, and utilising of volunteers with expert skills and whose demographics (on the group level) is atypical for crowdsourcing in general.

Mirroring our iterative work process, the paper is structured the following way: First, we provide an overview of the material we are studying and its defining characteristics, as well as the underlying theories and methods we use to approach the subject matter. Second, we describe the four different phases of the project, which each contain methodological considerations, results, and conclusions that led us to the following phase. Lastly, we present our conclusions, reflecting on the initial research questions and main outcomes of the expert crowdsourcing project.

2. Crowdsourcing as a method in digital humanities

Through the development of the internet, crowdsourcing in a scientific context has become an increasingly available and popular method for researchers and citizen scientists alike.[4] The method can "increase the accuracy of computer automated tasks, lower costs, increase the scale of research, transcend boundaries and borders, produce novel discoveries and increase the speed of research progression, among other benefits" [8].

The method is not perfect – common challenges include recruiting and retaining a crowd, finding a diverse enough group to avoid skewed end results, maintaining engagement and interest in the crowdsourcing task over a long time, and completion of a crowdsourced task.[5]

As with many cases in the information society, technological possibilities can sometimes overshadow ethical implications. While a research project may be non-commercial, the allocated

---

[1]For more about why the manuscripts have not been explored, and Lindgren's own role in cloaking her shorthand in mystery see M. Nauwerck, "Storyteller, stenographer, and self-published superstar: how Astrid Lindgren's multiple roles in book production created the Lindgren myth", forthcoming 2022.

[2]Transliteration is the correct term for the conversion of a text from one script to another which also involves the swapping of letters. Therefore we use the term transliteration for "transcription" of shorthand.

[3]Notably, this mixed methods approach is now a built-in feature in the Transkribus platform, and will for example be used in the research project Gustav's Hand [7]. Similar methods have been used in the medical sciences, where expert crowdsourcing is combined with image recognition software [13].

[4]The term crowdsourcing was coined in the Wired magazine in 2006 and is a portmanteau of "outsourcing" and "crowd". It was initially presented to utilise web 2.0 (i.e. the participatory or social web) to lower the cost of labour by assigning simple work tasks to a group of volunteers [8] but in an academic context it has come to be used interchangeably with citizen science and is described as a way to "broaden the scope and appreciation of humanistic enquiry" [1]. Citizen science can be defined as "non-professionally trained individuals conducting science-related activities" [9].

[5]Although in some cases completion may be neither desirable nor attainable, e.g. hedgehog observations [39]. However, crowdsourcing of cultural heritage collections usually does have an end point since collections contain a specified number of items, and many well-known heritage crowdsourcing projects are far from a 100% completion rate (e.g. the Transcribe Bentham project [36], What's On the Menu? [37], Occasional Poetry Catalogue [38]) even though there are examples of finished or almost finished projects as well (e.g. Anti-Slavery Manuscripts [34] and Georgian Papers (91% done as of 2022-02-10) [35]).

research funds are used to pay researchers and possibly other involved actors, which means remuneration for work is given to some and not others.

The challenges of the hybrid format of paid and non-paid work are thoroughly discussed by Lund [9] regarding the Wikipedia project.[6] He describes Wikipedia as a good example of how crowdsourcing in combination with non-profit foundations and open licences creates new ways of producing use values in society. Osman [10] highlights collaboration as an important aspect of engaging with Wikipedia, creating a mutually beneficial relationship between involved entities which goes beyond co-labouring. In conclusion, what crowdsourcing volunteers are not paid in financial remuneration can be replaced with other intangible rewards, such as the contribution to something perceived as important or being part of a collaborative community.

Another aspect of the labour in crowdsourcing is whether it yields enough results to be an affordable method in humanities research. While the workforce is participating on a volunteer basis, the infrastructure of crowdsourcing projects (e.g. hosting and supporting a crowdsourcing platform, adding and editing material to the platform, technical and topical support to volunteers) costs money, time, and engagement. Experiences from the Transcribe Bentham project (a project to transcribe the manuscripts of the British moral philosopher Jeremy Bentham) suggest that cultural heritage crowdsourcing projects need an "ambitious and well thought-through project plan at the very beginning, and ongoing institutional support, commitment, and resources to successfully meet the crowdsourcing programme's goals, or it is unlikely that the cost-avoidance or, indeed, any other aims will be obtained" [2].[7]

The cost of crowdsourcing infrastructure can be measured against the results, such as the amount of transcribed pages or the accuracy rate of transcribed pages or be compared with alternate solutions to the crowdsourcing task. For instance, the Transcribe Bentham project staff evaluated the economic investment by comparing crowdsourcing costs with the potential cost of hiring transcribing staff [2]. They also evaluated the outcome of a three-year crowdsourcing effort and noted that while 6,000 transcriptions finished and 70,000 transcriptions to go "does not sound all that impressive", the expected trajectory is that all Bentham's manuscripts could be transcribed within two decades. While this may sound like a long time to complete the task, they note that it is still "faster than had Transcribe Bentham never existed" [11].

2.1.   Expert crowdsourcing

A feature of crowdsourcing is that it invites a crowd to participate in a work process, with or without previous knowledge. Crowdsourcing projects may direct their recruitment efforts so that their volunteers have a relevant connection to the project (using the Transcribe Bentham project as an example, the staff reached out to the academic community and to schools [12]) and potentially also specific skills to make the crowdsourcing task easier (e.g. knowledge of the subject matter or experience in reading/deciphering handwriting).

Expert crowdsourcing (or expert sourcing)[8] is, just like its hypernym crowdsourcing, used in many different ways, spanning from analysing documents written by experts [13], to employing experts (professionals) instead of a crowd (non-professionals) [14],[9] to compiling individual work efforts in combination with machine learning [15].

For the Astrid Lindgren Code project, a tailored crowd was a necessity due to the specific and nowadays rare knowledge of Melin shorthand. We refer to this method as expert crowdsourcing, meaning that only people with a specific expertise were invited to be part of the crowd in this crowdsourcing effort.

2.2.   Hackathons

The original meaning of hackathon, a portmanteau of 'hack' and 'marathon', is a collaborative event for exploratory programming [16]. In this project, hackathon is used to describe a collaborative event for solving digital crowdsourcing tasks.

The use of the term was derived from previous experimental crowdsourcing events at Uppsala University Library and was chosen for its connotation to technology to de-dramatise the concept of data and the digital elements of the event, in the hopes of increasing the digital skills and confidence among the participants. As this project progressed we noted that another benefit of using the term hackathon (in lieu of e.g. 'transcribathon') is that it is favoured in media storytelling through its contradictory connotations with the project matter (shorthand vs. writing software code) and demography (older generation vs. younger generation).

We chose hackathons as a method because findings from the hackathons at Uppsala University Library suggested that they would increase the sense of ownership among volunteers (which in turn creates more engagement and motivation to continue to contribute) and make the process itself better than if it was pre-defined and static (in terms of workflow efficiency, pedagogical instruction of the workflow, end result of crowdsourcing, or all of these). However, experimental hackathons had not been used to fully transcribe a whole corpus of text. Combining a defined research goal with a crowdsourcing process based on social events and active participation with the process and the task was a method that had not previously been tested.

3. The task: transliterating Astrid Lindgren's shorthand notepads

The Melin system of shorthand was widely used in Sweden during the 20th century. Lindgren herself learned it as part of her professional secretary training at the Bar-Lock Institute in Stockholm in 1926–27. Although the skill has become obsolete in Swedish professional life, it is still practised nationally, for example through *Melinska stenografförbundet* [17], a social society with several local branches in Sweden. This meant that a potential expert crowd of volunteers existed. However, the primary material of Lindgren's shorthand notepads still posed several challenges to general models for transcription-based crowdsourcing projects. The lack of previous attempts to use citizen science or crowdsourcing for shorthand transliteration[10] in combination with the project's reliance on the participation of a specific crowd consequently required material-based method development as well as continuous sensitivity to how the user experience could best be adapted to the demographics and prerequisites of the volunteers.

The volunteers were subsequently invited to take part in developing this method from the start, resulting in a more iterative and experimental rather than curated and generalised form of transliteration based on 1) the specificity of Lindgren's shorthand notepads; and 2) the user experience of the volunteers. In addition to what turned out to be a quick process and a high completion rate, advantages with this approach also include the increased involvement of volunteers in building the project (favouring a sense of joint-ownership) as well as the opportunity for us to explore the potential of letting volunteers play a more dynamic part in the process, as co-developers of methods and tasks.

3.1. The unlikely volunteer: secretaries as code breakers, grunt workers, and technical pioneers

To a high degree, the volunteers of The Astrid Lindgren Code reflect the medial and historical context of shorthand in Sweden during the twentieth century. Based on both registration letters and evaluations, a majority of the volunteers are retired former professionals, born in the 1930s and 40s, who learned shorthand as part of their professional education, and used it in their careers as

[10]Since then, the Dickens Code project (2021–2022) [18] has worked with public calls and crowdsourcing to decipher Charles Dickens' preserved documents written in his own system of "brachiography". Dickens' material shares some traits with Lindgren's but is in comparison very sparse, putting the individual note, page, or letter at the centre. Based on cryptographic code breaking rather than shared professional experience/skill set, the crowdsourcing activities of the Dickens Code project also require different kinds of expert volunteers. However, the parallels in storytelling around both projects are notable, and in both cases seem to have served their purpose in attracting volunteers.

secretaries, administrators, office workers, or stenography teachers. Whereas some of the younger volunteers are still working in shorthand related professions, for example as parliamentary secretaries or journalists, today most of them practise shorthand only as a specialised hobby. The geographical distribution of the volunteers is evenly spread across Sweden, with an equal representation of rural and urban areas. Roughly estimated, 90% of the volunteers are women.[11]

On a general level, crowdsourcing activities tend to mainly attract male volunteers, most famously perhaps in the case of Wikipedia [19]. A comprehensive study of crowdsourcing projects on the Zooniverse citizen science platform suggests that scientific culture and favourable socio-economic conditions also generally benefit participation in citizen science activities, yet more extensive participation in one country does not mean that volunteers to a higher degree reflect its demographics [20]. Whereas there are indications that the gender gap might be closing among younger volunteers, the age gap remains [21].

As previously mentioned, this is reflected in the discourse around The Astrid Lindgren Code where Swedish media coverage has acknowledged that terms such as "hacking" and "code breaking" are generally associate with young men, and that an expert crowd of mostly middle aged and older women in this sense is unexpected [22].

The undercurrent of women's coding and grunt work is however an integral part of 20th century history of computing,[12] and today it also serves as a pop culture trope reflected in television shows and films.[13] Notably, Lindgren herself worked at the Swedish secret service's department for letter censorship from 1940 to 1945, an experience that came to influence her literary work and to some extent also her deliberate use of shorthand as a 'secret language'.

Although the crowdsourcing tasks of The Astrid Lindgren Code are more about interpretation and puzzle solving than decrypting or code breaking in a literal sense, the close relationship between the secretarial profession, problem solving, and technical development has likely been integral to securing participation from an older generation. During the twentieth century, stenography, a technical aid, was eventually replaced by tape recorders, computers and smart devices in the workplace. Volunteers who repeatedly have had to adapt to new technology as part of their profession are arguably more likely to have acquired the necessary digital literacy required for participation.[14]

3.2. Senior citizen science: digital literacy in the time of a pandemic

On a general level, access to the technologies and skills required to participate in online activities for data creation and sharing has increased during the Covid-19 pandemic. "[T]he crisis has urged older adults to adopt new technologies to facilitate their tasks, as well as to provide them with an effective means against loneliness and social isolation caused by the confinement", as noted by Martínez-Alcalá [23]. Although the country never practised hard lockdown during the pandemic, Sweden's applied recommendations for social distancing and isolation were especially directed toward the age group of "70+" who during the first year of the pandemic were encouraged to keep distance and stay at home. Even if the digital literacy among an older generation is proportionally high in Sweden and has increased during the pandemic, the prior knowledge among senior citizens varies to a great extent [24]. In this case, the digital tools and platforms used in the project were new to almost all volunteers and required introductory tutorials and technical support.

4. The iterative phases of expert crowdsourcing

---

[11]An evaluation was sent out on the project's mailing list in February 2022 and received 35 responses. These responses primarily reflect experiences from highly motivated individuals who have stayed on as volunteers throughout the project's progress, making them a relevant yet inconclusive material for evaluating all participation in the project.

[12]The first computers were indeed women, and only later was the name designated to machines. Arguably, the pioneer project of digital humanities was also carried out by female computers. For further reading on the women of Father Roberto Busa's punch card project, see Terras [40], Eveleth [41], and Nyhan [42].

[13]Such as *The Bletchley Circle* (2012–2014), *The Imitation Game* (2014), and *Hidden Figures* (2016).

[14]Still, the project has along the way lost volunteers who have not been comfortable with the digital platforms and programs used in the project, resulting in a younger average age within the volunteer group. There is also the group of potential volunteers who we had to decline at the project's initial stage as they reached out through relatives, phone, or letter and for whom digital participation was never an option.

In this section we define and describe the phases of our expert crowdsourcing process. Our aim was to create a social environment around the crowdsourcing task where volunteers could actively contribute to forming the process. We defined four phases of the expert crowdsourcing process: 1) finding a crowd, a platform, and a workflow; 2) introducing and engaging the crowd at hackathons; 3) post-hackathon transliteration and wrap-up; and 4) future development of the project and volunteer initiatives.

4.1. Phase 1: Finding a crowd, a platform, and a workflow

For the crowdsourcing component of the project, the initial plan was to organise physical hackathons, but since we began in early 2021, we had to adapt to the conditions stipulated by the Covid-19 pandemic. The first phase consisted of finding a crowd, deciding on a crowdsourcing platform and developing workflows for the crowdsourcing process.

4.1.1. Recruiting the crowd

Recruitment of volunteers started as the project was presented in an in-depth interview in the Swedish national radio show *Vetenskapsradion forskarliv* (P1) [25]. During this presentation we sent out a call for stenographers who wanted to participate in deciphering Astrid Lindgren's manuscripts. The call received an overwhelming response, with more than a hundred stenographers signing up within three weeks of the radio interview. As The Astrid Lindgren Code has continued to attract media attention, volunteers have continued to join. Today there are approximately 170 assigned volunteers in the project, with a more continuously active core of approximately 40 individuals.[15]

Recruiting and retaining a crowd is generally considered a challenge for crowdsourcing projects (for instance, two-thirds of volunteers on the Zooniverse platform make only one classification and do not return [20]), and since we also requested a very specific skill set we were fortunate to receive such a large response. The volunteers offered several reasons for signing up, from the intellectual, to the professional, to the emotional. Recurrent themes also mentioned were the high profile of the subject matter, devotion to the craft of stenography and its survival, ambition to aid research, interest in a challenge aimed at them particularly, and a curiosity to test if their shorthand knowledge was sufficient for the task.

Notably, the volunteers often mentioned an identification with or admiration for Astrid Lindgren. It is noteworthy that this type of emotional reward which indicates an importance of personal connection to the material is not explicitly mentioned in Estellés-Arolas' suggestion that recompense from crowdsourcing "would always look to satisfy one or more of the individual needs mentioned in Maslow's pyramid: economic reward, social recognition, self-esteem or to develop individual skills" [26].

4.1.2. Finding a crowdsourcing platform

Our technical specifications for the platform were: preferably open source, easy to manage (i.e. not requiring coding expertise) and easy for volunteers to use. Because of the crowd demographic, we knew that we needed to put special emphasis on the user experience of the platform.

We decided to try out two crowdsourcing platforms which were both feasible options for us as facilitators: Zooniverse, which is used for crowdsourcing a wide range of subjects and is hosted online [27], and Omeka-S, a collection management software installed on a separate server [28], with the Scripto plug-in, a tool for transcribing documents [29] (henceforth referred to as Omeka).

As facilitators, Zooniverse would have been advantageous because of the online hosting, the online tutorials for users, and the simple file upload system. Advantages of using Omeka would be the structure for uploading data (the image/item/item set structure is simpler than the manifests used in Zooniverse) and the transparency and availability of data storage and export.

---

[15]Although the formation of a core group of more active users is a general tendency in crowdsourcing projects, there are indications that many of the volunteers who originally signed up either found the material too challenging to work with or lost interest for other reasons.

We convened with the expert group of twelve stenographers on two occasions (on 29 January 2021 to test Zooniverse, and on 19 February 2021 to test Omeka). After the first meeting we sent out a survey to gather feedback on the platform, on the project's structure on the platform, on transliteration instructions, and on the source material itself (the perceived difficulty of reading it, etc.). After the second meeting we sent out another survey with similar questions, as well as questions about how the two platforms compared.

Omeka's weaknesses were the lack of a Swedish interface and that users had to click a save button to save the transliteration lest all text be lost (this was especially frustrating when one had accidentally navigated away from the Omeka page, e.g. by clicking the back/forward arrows).

In the Zooniverse platform it was difficult to organise the material in a meaningful way as the interface seems to favour a smaller number of files in one item (i.e. a letter with two to four pages is more suitable than a notepad with 60 to 100 pages). The stenographers expressed frustration with lack of context (which is particularly important for shorthand, where interpretation is often based on context) when we uploaded the manuscripts in sets of three pages that appeared to the volunteers in a random order. While we think this issue could have been solved in Zooniverse with a bit more tinkering, it was easier to use Omeka as it already had the hierarchic item organisation that the volunteers requested.

To summarise, both platforms had pros and cons but the deciding factor was the volunteers' preference for Omeka's presentation of the manuscripts as well as the comprehensive editing interface, so we used Omeka.

4.1.3. Creating a workflow

Having decided on a platform, the next step was to ensure that the transliterations were produced with quality and consistency. Both the stipulations of shorthand transliteration and the future purpose of the transliterations required particular consideration. Typing up shorthand in a professional context generally implies both interpretation of intent and the conforming of colloquial and oral elements into the appropriate written style. Although such a standard would be familiar to the volunteers, it would eliminate much of what is intriguing about the manuscripts from a literary point of view, such as the oral aspects of Lindgren's creative process [30]. A model for transliteration designed to reflect the Melin system, for example by mirroring phonetic signs and abbreviations, would however not only be very time-consuming and complicated for the volunteers to perform, but also generate text too unwieldy to work with for most purposes. Ultimately, we decided on line-to-line transliteration, following regular spelling conventions but without any adding of what in shorthand is generally only implied (such as punctuation). Our main considerations were facilitating transliteration while still preserving relevant features of shorthand as well as creating feasible transliterations for HTR development.

We decided not to use the inbuilt editing features of Omeka, as these would have required the volunteers to learn and use a dual set of systems for editing. For HTR purposes, we also needed the transliterations to be very clear on, for example, the exact placement of additions and deletions, and the inbuilt editing features of Omeka were not sufficient. Furthermore, using the Omeka features caused confusion when editing the text, as pressing [B] generated the tags "" and "" in the transliteration, a format that many of the volunteers were not familiar with. For similar reasons we decided not to use the TEI (the Text Encoding Initiative)[16] or similar transcription guidelines as the learning curve would have been too steep. Instead, we created our own convention for transliteration and presented it in a manual [31] which was developed and refined in close dialogue with the test group of volunteers.

4.2. Phase 2: Introducing and engaging the crowd at hackathons

---

[16]Text Encoding Initiative, https://tei-c.org/.

With the manuscripts prepared in Omeka, we launched the crowdsourcing platform for the whole group of stenographers. The transliteration activities were centred around a series of hackathons where stenographers could ask questions, collaborate on tricky details in the manuscripts, and continue to develop the workflow. In between the hackathons we communicated through newsletters to encourage continued transliteration and to share news on the project and new guidelines for the workflow.

The inaugural hackathon took place on 7 April 2021, and was followed by five more hackathons on 23 April, 8 May, 20 May, 3 June, and 19 June. The hackathons were two hours long; the length was evaluated at the end of the first hackathon where participants agreed that two hours was an appropriate time and that a longer duration would have been too long to spend in front of a screen. To include as many stenographers as possible, the hackathons were scheduled on different weekdays and hours during the day. Participation was around 20–25 people for each hackathon (except the first one which was attended by around 40 people) and were from a core group of approximately 40 stenographers.

4.2.1. Digital tools and skills

The hackathons were held through the Zoom video online conference tool which was a new kind of software for many participants. Thus, the hackathon participants were confronted with two new pieces of software: Zoom and Omeka. The digital competency skills varied within the group and so a part of each hackathon was dedicated to helping people get online in Zoom, find their bearings with its controls,[17] as well as instructing them in the use of Omeka. As the hackathons progressed one-sided support from us was complemented by peer-to-peer support, especially when we started using breakout rooms in Zoom.

Teaching the volunteers how to use the new digital tools was a crucial step in creating an inclusive crowdsourcing environment. On the one hand, they were sought after experts because of their shorthand skills, on the other hand, they were beginners in the digital tools we used. In this environment, self-confidence was mixed (both regarding shorthand skills and digital skills), and it was important to us to emphasise the volunteers' expertise to motivate the digital tools learning curve. Some volunteers decided to drop out of hackathon participation or out of the project as a whole due to technical difficulties or hardship in understanding the functionalities of Zoom and/or Omeka.

4.2.2. Co-creative workflows

The transliteration workflows during and in between hackathons were continuously developed together with the hackathon participants; the hackathons became a forum for interaction and renegotiation of currently agreed practices. If volunteers had provided input on the process via e-mail, this input was discussed during hackathons to get feedback on the issues. Any new guidelines or recommendations for the project were then communicated through newsletters.

We tried working in different ways during the hackathons to find a suitable form for transliteration workflows, for example solitary work while logged on, working in breakout rooms depending on interest or need (e.g. technical support, coffee break, solving encountered difficulties in the texts), and working in breakout rooms with specific tasks (group work in solving difficult paragraphs and words in an assigned set of notepads).

Some examples of practices that were implemented after the hackathons include: a notepad for difficult passages where volunteers could ask for help and discuss solutions together (this was designed to be a forum for discussion for volunteers in between hackathons, as we decided we could not share personal information such as e-mail addresses with everyone); the manual division of pages to transliterate so that everyone had a more specific task; a review process where a participant could ask for a second reading on specific page(s) they had transliterated; and how to sign and save the transliteration so that it was clear who had transliterated which page.

---

[17]E.g. mic on/off, camera on/off, screen sharing, chat function, joining/participating in breakout rooms.

4.2.3. Communication between hackathons

With the hackathons as a backbone for the crowdsourcing effort, we filled the gaps in between with a newsletter to make sure everyone involved got information about changes in the workflow that had been agreed upon on the hackathons. As mentioned in the previous section, each hackathon had 20–25 participants and the full contact list consisted of 170 people.

Drawing on previous experiences from crowdsourcing projects (e.g. [11, 26]) we knew that the sense of community and involvement was a strong motivational factor for volunteers. Therefore, it was crucial to communicate regularly with this community so that everyone would be in the loop on developments and news on the project. By communicating everything from transliteration riddles to media coverage, and by offering the volunteers space in the newsletter and the social media channels of the project, we wanted to create a sense of co-ownership. On several occasions we included calls for volunteers to participate in local and national media outlets that had been in contact with the project.

4.2.4. When is "done" done? Review and export of transliterations

When working with difficult texts, it may be hard to know when a transliteration (or transcription) is done. Words or sentences may be marked as uncertain and require further review by other stenographers. Deciding on when a transliteration is complete has been an iterative and inconclusive process. The inconclusiveness is partly due to the different uses/meanings of the "final transliteration": 1) export the text to train an HTR algorithm, 2) create a text suitable for a forthcoming critical edition, and 3) knowing as a volunteer when a page is closed for editing.

Because of these conflicting values we did not end up with a clear definition of when a text is done. We tried a peer-review system similar to the Library of Congress' *By the People* project [32], but quickly realised that the different levels of expertise within the crowd led to a very uneven quality in the end result. Next, we considered peer-review by a selected group of reviewers, drawing from the pool of superusers within the project. We wrote a draft for a manual for reviewing (in addition to the manual for transliteration) but ended up not using it. The reviewing process is still not finalised but will be a part of the upcoming phase of the project when we will start to export transliterations from Omeka for new purposes.

4.3. Phase 3: Toward an independent crowd

After the last hackathon in June, we created a mailing list using Google Groups to allow for direct connection between volunteers without disclosing their personal data (they had to accept the invite to the group and were made aware of the visibility of their name and email address if they accepted). The mailing list was also created with the hope of ensuring that the crowdsourcing could continue without extensive support from us and instead relying on peer support from the volunteers.

We invited the volunteers to make use of the digital meeting space (accessible through the same link as had been used for all hackathons) to arrange their own hackathons. While no such initiatives were made for the whole group, we know that some of the volunteers met up in smaller groups. These constellations had formed using breakout rooms in the Spring 2021 hackathons.

Most volunteers fell back to solitary work. Some expressed (through emails and the mailing list) that they wanted hackathons to be arranged. We could not meet that need, but the request confirmed conclusions from other crowdsourcing projects: a crowd requires maintenance to be kept together. This became even more important when we were running out of material for the task (something we had not anticipated and therefore not planned for) as the material was an important motivational factor; we know that the lack of material to transliterate caused volunteers to lose interest in the project.

4.3.1. Christmas code cracking: a hybrid hackathon finale

Because of changes in the Covid-19 pandemic restrictions, we were able to organise a physical hackathon in Stockholm as a wrap-up for the crowdsourcing part of the project. The hackathon was arranged as a hybrid event on 8 December 2021, allowing for remote and in-person participation (the number of physical attendees was limited).

About 70 people participated at the event. In addition to the shorthand expert crowd, we also invited experts from the Astrid Lindgren Society [33], thus merging two expert crowds to complete one task. At this event, the task was to decipher unidentified manuscripts, and the combined expertise of two expert crowds rendered a positive result.[18]

The shorthand volunteers expressed joy in participating in the hybrid hackathon, some because of a hackathon taking place after a long hiatus and some emphasised the pleasure of finally meeting us and the other volunteers in person.

For us, it was interesting to arrange a physical event according to what we originally planned in the research application; the Covid-19 pandemic forced us to explore different methods (which in the end may have been more successful for the completion rate) and had the benefits of being able to include volunteers who otherwise would have been restricted from attending due to geographical distances. By using the hybrid format, we could combine the best aspects of in-person and remote participation.

4.4.   Phase 4: Future development of the project and volunteer initiatives

The fourth phase is a look into future potential development of this project and the outcome of what we have done so far. How can the data best be utilised for HTR development, and how will we proceed with a genetic/critical edition of *The Brothers Lionheart* based on the transliterations? Today more than 600 notepads remain in the Astrid Lindgren Archives and there is a keen interest among volunteers to continue transliteration. However, the time frame is an issue: how long will it take to secure funds for continued digitisation and how long will the digitisation process take? Will the interest in the project decrease if the current momentum is lost?

Several volunteers report that primarily, new challenges and tasks are needed to motivate them to continue working with the manuscripts. One solution to all the issues above is to utilise the initiatives created by the volunteers themselves. So far, these have included: lexicons with Lindgren's shorthand images and typed counterparts in standard Melin from the 1960s; independent research in order to identify unknown texts in the notepads; strategies for recognising when Lindgren has written shorthand in English or German; and developing pedagogical material for teaching shorthand to beginners in collaboration with the Melin shorthand society.

On the one hand, further involvement of the 'super-users'[19] among expert volunteers closes the gap between 'citizens' and 'science'. On the other there are several ethical aspects to consider when lines between volunteer and paid researcher/expert start to blur. If certain individuals are lifted from the crowd of stenographers, what happens to the crowd? Could it potentially have a negative impact on their motivation? We wonder if this is particularly difficult when using expert crowdsourcing, as the crowd of experts is already singled out from the crowd. When recognition and praise is our main compensation to the volunteers, it is a tricky balance to uphold.

5. Results and conclusions

5.1.   A sustainable crowdsourcing lifecycle

An important result of this crowdsourcing project is that it was completed. While there are good examples of finished crowdsourcing projects (e.g. Anti-Slavery Manuscripts [34], Georgian Papers (91% done as of 2022-02-10) [35]), there are also plenty of projects which are in progress and expected to be so in the foreseeable future (e.g. Transcribe Bentham [36], What's On the Menu? [37], Occasional Poetry Catalogue [38]). Because of our unique scope and the special skill set required to

---

[18]This event was covered extensively by national media, for example by news agency TT who reported on the findings of the three unidentified shorthand notepads [33].
[19]The most active users in a crowdsourcing project, cf. Terras 2015.

participate in the expert crowdsourcing, we could not predict where our project would fall on this scale, so we were pleasantly surprised by the level of completion.

The Transcribe Bentham project has made an estimation that the four-year period average of 2,704 transcripts a year would result in a completion of the Bentham texts in 2036 [2]. If we make a similar estimation based on our results, all shorthand material in the Astrid Lindgren archives could be transliterated in less than a year.[20] Even if it is difficult to predict if the transliteration speed would be consistent, even a slightly less optimistic outlook would give a good prognosis timewise. And this is without considering the potential contributions from HTR.

Two practical outcomes of the completed transliterations so far are that we have a full text corpus that can be used by the HTR part of the project as well as a full transliteration of shorthand material related to *The Brothers Lionheart* which can be used for literary analysis. On a broader scale, the transliterated data and its contribution to HTR will contribute to the digital humanities field as a whole, with developed algorithms and training data available as part of larger infrastructures and new projects.

A recommendation we would pass on to other cultural heritage crowdsourcing projects is to centre the crowdsourcing around a limited and well-curated content. This is a bit contradictory to the workflows of mass digitisation of cultural heritage collections where it is sometimes a better strategy to digitise larger quantities at a time,[21] but it could be beneficial (for crowdsourcing as well as other public engagement) to break out certain parts of a collection even if the initial result is an "incomplete" digitised collection. Limiting the content and having a clearly defined and attainable end point also limits the social and emotional work that goes down into the upkeep of a crowdsourcing project.

5.2. Shortcut to the super-user: benefits of expert crowdsourcing

Dependence on expert knowledge can actually be an advantage when recruiting, retaining, and motivating a crowd. The targeting of people for their expertise adds value to the user experience and enhances the sense of a specialised and exclusive community, both important factors in motivating the volunteers. From a researcher's point of view, expert crowdsourcing can be a shortcut to the much sought-after super-users. Within The Astrid Lindgren Code the initiatives of such invested volunteers have provided several opportunities for mixed methods development and sub-projects.

Arguably, the benefits of using experts in crowdsourcing must be balanced against the higher "costs", i.e. a higher degree of investment and interaction from the project facilitators. In this sense expert labour might 'cost' more even when it is free.

5.3. Covid-19 aftermath: a ray of light in a dreary time

Except for the hybrid finale, the crowdsourcing process, like many other activities during the Covid-19 pandemic, has been entirely based on remote participation through online software.

On the direct question of how the pandemic has influenced their participation in the project, approximately 50% of the volunteers who responded to our evaluation claim that it did not affect their participation. The other half primarily report that the pandemic has resulted in more time to do the work, because of short term permutations or recommendations to stay at home. One respondent mentioned a general fatigue that drained them of the energy to participate more, whereas others stated that the crowdsourcing activities have felt like something meaningful to do together with others despite isolation, or that participating in the project has been a ray of light in a dreary time.

Although evaluations indicate that the crowdsourcing process would have been successful regardless of the Covid-19 pandemic, it is likely that the pandemic situation favoured broad participation in terms of age and geography, both by contributing to improved digital literacy and by

---

[20]Based on the rough estimation that 52 notepads (8% of the shorthand material in the Astrid Lindgren archives) were transliterated in five weeks worth of time.

[21]To illustrate, in our project the matter of how much and how to digitise the Astrid Lindgren archives was a negotiation with the cultural heritage institution where the material was deposited.

creating a need for intellectually challenging and/or socially meaningful activities compatible with remote participation.

5.4. Utilising the expertise: co-creation and community building

We can conclude that the hackathons were a successful model to create and maintain a sense of community within the crowd. The hackathons received positive feedback from volunteers both during the spring when they took place and six months after when they were invited to give some general feedback on the project.

The volunteers did not seem to reflect on being co-creators of changes in the workflow, but many of their preferred ways of working are direct results of discussions during and between the hackathons. For instance, many volunteers noted that they preferred the hackathons when they were working in smaller groups to solve problems, which was how the last two hackathons were organised.

While our work focused heavily on including the volunteers in the co-creative process around the crowdsourcing workflow, an unexpected but thrilling outcome was the amount of co-creative projects they themselves initiated (see examples under Phase 4). Regardless of whether these projects are a direct result of our empowerment and co-ownership efforts or of the volunteers' overall engagement, they have been supported by an accommodating and inclusive atmosphere that encourages initiative and values competence.

5.5. Reaching the unlikely volunteer: communication and personal relationships

We conclude that the citizen science activities have benefitted largely from the media attention given to the Astrid Lindgren Code project. Research communication through press releases and social media as well as continuous interviews have generated public visibility, essential in recruiting volunteers as well as keeping them. Inviting volunteers to participate in interviews and acknowledging their work when results from the project were communicated in social and regular media, have contributed to the overall user-experience and sense of joint ownership.

This is also the case with the continuous communication between volunteers and researchers through email, hackathons, social media, and newsletters. A general focus on reliability, accessibility, inclusion, and appreciation through personal communication has been integral in motivating the volunteers.

The rare demographics of the volunteer group, consisting primarily of Swedish senior women, can possibly shed some new light on how to motivate this target group specifically and why they are often proportionally absent within general crowdsourcing activities although present in other participatory cultures.

A major finding of this study is the importance of a subject or research question which engages the volunteer on a personal level. Here, the professional background and skill set shared with Lindgren as well as the author's impact on the volunteers through her literary fiction have generated an emotional affinity with the primary material which turned out to be a guarantee for prolific results.

For many volunteers of crowdsourcing the contribution to science is both a community building factor and considered a reward in itself. Therefore, it might be less important whether you apply your knowledge to the classification of ladybirds or counting of hedgehogs. For volunteers of expert crowdsourcing other motivational factors seem to be at play: Does it matter if you transliterate Pippi Longstocking or parliamentary protocols? Does it matter if the task is open for general participation or if participation means being part of an exclusive community? Does it matter if anyone could do the job, or if it is your specific skills that can solve a problem? This study suggests that it does.

6. References

[1] M. Terras, Crowdsourcing in the Digital Humanities, in: S. Schreibman, R. Siemens, J. Unsworth (Eds.), A New Companion to Digital Humanities, Wiley Blackwell, Chichester, 2015, pp. 420–438. doi: 10.1002/9781118680605.ch29.

[2] T. Causer, K. Grint, A.-M. Sichani, M. Terras, 'Making such bargain': Transcribe Bentham and the quality and cost-effectiveness of crowdsourced transcription, Digit. Scholarsh. Humanit., Jan. 2018, doi: 10.1093/llc/fqx064.

[3] M. Nauwerck, Riksbankens jubileumsfond, Astrid Lindgren-koden: Astrid Lindgrens stenograferade originalmanuskript genom digital bildanalys, genetisk kritik, bok- och mediehistoriska perspektiv (dnr: P19-0103:1), 2020. URL: https://www.rj.se/anslag/2019/astrid-lindgren-koden-astrid-lindgrens-stenograferade-originalmanuskript-genom-digital-bildanalys-genetisk-kritik-bok—och-mediehistoriska-perspektiv/.

[4] V. Edström, Astrid Lindgren och sagans makt, Rabén & Sjögren, Stockholm, 1997.

[5] L. Törnqvist, Rapport : Astrid Lindgrens arkiv – nya forskningsmöjligheter, Barnboken Tidskr. för barnlitteraturforskning 34 (2011) 59–67.

[6] R. Heil, M. Nauwerck, A. Hast, Shorthand Secrets: Deciphering Astrid Lindgren's Stenographed Drafts with HTR Methods, in: D. Dosso, S. Ferilli, P. Manghi, A. Poggi, G. Serra, G. Silvello (Eds.), Proceedings of the 17th Italian Research Conference on Digital Libraries, Padua, Italy (virtual event due to the Covid-19 pandemic), February 18-19, 2021, pp. 169-177. URL: http://ceur-ws.org/Vol-2816/short5.pdf.

[7] M. Alm, Riksbankens jubileumsfond, Gustav's Hand: Digitisation, Digital Enhancement, and Dissemination of the Gustavian Collection, 2021. URL: https://www.rj.se/en/grants/2021/gustavs-hand-digitisation-digital-enhancement-and-dissemination-of-the-gustavian-collection/.

[8] K. Wazny, Crowdsourcing's ten years in: A review, J. Glob. Health 7 (2017). doi: 10.7189/jogh.07.020601.

[9] A. Lund, Frihetens rike : Wikipedianer om sin praktik, sitt produktionssätt och kapitalismen, Tankekraft förlag, Hägersten, 2015.

[10] K. Osman, The Free Encyclopaedia that Anyone can Edit: The Shifting Values of Wikipedia Editors, Cult. Unbound 6 (2014). doi: 10.3384/cu.2000.1525.146593.

[11] T. Causer, M. Terras, 'Many hands make light work. Many hands together make merry work': Transcribe Bentham and crowdsourcing manuscript collections, in: M. Ridge (Ed.), Crowdsourcing our Cultural Heritage, Ashgate, Farnham, 2014, pp. 57–88.

[12] T. Causer, V. Wallace, Building A Volunteer Community: Results and Findings from Transcribe Bentham, Digit. Humanit. Q., 6 (2012). URL: http://www.digitalhumanities.org/dhq/vol/6/2/000125/000125.html.

[13] A. Androutsopoulou, F. Mureddu, E. Loukis, Y. Charalabidis, Passive Expert-Sourcing for Policy Making in the European Union, in: E. Tambouris, P. Panagiotopoulos, Ø. Sæbø, M. A. Wimmer, T. A. Pardo, Y. Charalabidis, D. Sá Soares, T. Janowski (Eds.), Electronic Participation, vol. 9821, Springer International Publishing, Cham, 2016, pp. 162–175. doi: 10.1007/978-3-319-45074-2_13.

[14] I. Bekker, Y. Felus, Quality Control for Crowdsourcing Large Scale Topographic Maps, Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci., vol. XLII-2-W13 (2019) 1201–1205. doi: 10.5194/isprs-archives-XLII-2-W13-1201-2019.

[15] S. Auer, A. Oelen, M. Haris, M. Stocker, J. D'Souza, K. E. Farfar, L. Vogt, Lars, M. Prinz, V. Wiens, M. Y. Jaradeh, Improving Access to Scientific Literature with Knowledge Graphs, Bibl. Forsch. Prax. 44 (2020) 516–529. doi: 10.1515/bfp-2020-2042.

[16] Wikipedia, "Hackathon", 2021. URL: https://en.wikipedia.org/w/index.php?title=Hackathon&oldid=1052850889.

[17] Melinska Stenografförbundet. URL: http://stenografi.nu/.

[18] University of Leicester, The Dickens Code, 2022. URL: https://le.ac.uk/dickens-code.

[19] E. Hargittai, A. Shaw, Mind the skills gap: the role of Internet know-how and gender in differentiated contributions to Wikipedia, Inf. Commun. Soc. 18 (2015) 424–442. doi: 10.1080/1369118X.2014.957711.

[20] K. Ibrahim, S. Khodursky, T. Yasseri, Gender Imbalance and Spatiotemporal Patterns of Contributions to Citizen Science Projects: The Case of Zooniverse, Front. Phys. 9 (2021). doi: 10.3389/fphy.2021.650720.

[21] C. L. Thigpen, C. Funk, Younger, more educated U.S. adults are more likely to take part in citizen science research, Pew Research Center, 2020. URL: https://www.pewresearch.org/fact-tank/2020/06/25/younger-more-educated-u-s-adults-are-more-likely-to-take-part-in-citizen-science-research/.

[22] Å. Stibner, Här knäcker de koden till Astrid Lindgrens anteckningar, December 8, 2021, TV4 Nyheter. URL: https://www.tv4.se/artikel/71VsDfwZhTk9rFRDBDGVjJ/haer-knaecker-de-koden-till-astrid-lindgrens-privata-anteckningar.

[23] C. I. Martínez-Alcalá, A. Rosales-Lagarde, Y. M. Pérez-Pérez, J. S. Lopez-Noguerola, M. L. Bautista-Díaz, R. A. Agis-Juarez, The Effects of Covid-19 on the Digital Literacy of the Elderly: Norms for Digital Inclusion, Front. Educ. 6 (2021). doi: 10.3389/feduc.2021.716025.

[24] Svenskarna och internet 2021, Internetstiftelsen, 2021. URL: https://svenskarnaochinternet.se/rapporter/svenskarna-och-internet-2021/.

[25] Y. C. Warnborg, Malin ska knäcka 'Astrid Lindgren-koden', October 14, 2020, Vetenskapsradion Forskarliv, Sveriges Radio. URL: https://sverigesradio.se/avsnitt/1582494.

[26] E. Estellés-Arolas, F. González-Ladrón-de-Guevara, Towards an integrated crowdsourcing definition, J. Inf. Sci. 38 (2012) 189–200. doi: 10.1177/0165551512437638.

[27] Zooniverse, 2022. URL: https://www.zooniverse.org/.

[28] Omeka, 2022. URL: https://omeka.org/.

[29] Scripto, 2022. URL: https://scripto.org/.

[30] M. Nauwerck, Sagoberättaren, sekreteraren och den spelande linden, in: J. Pennlert and L. Ilshammar (Eds.), Från Strindberg till Storytel: korskopplingar mellan ljud och litteratur, Daidalos, Göteborg, 2021, pp. 197–219.

[31] M. Nauwerck, K. Andersdotter, Manual för transkribering i Omeka. Version 1 (webbversion)., Svenska barnboksinstitutet, 2021. URL: https://www.barnboksinstitutet.se/wp-content/uploads/2021/03/Manual-fo%CC%88r-transkribering-i-Omeka-webbversion.pdf.

[32] Library of Congress, By the People – How to Review, 2022. URL: https://crowd.loc.gov/help-center/how-to-review/.

[33] Astrid Lindgrensällskapet, 2022. URL: https://www.astridlindgrensallskapet.se/.

[34] Boston Public Library, Anti-Slavery Manuscripts, 2020 URL: https://www.antislaverymanuscripts.org/.

[35] Georgian Papers Programme, Transcribe Georgian Papers, 2022. URL: https://transcribegeorgianpapers.wm.edu/.

[36] UCL, Transcribe Bentham, 2022. URL: https://blogs.ucl.ac.uk/transcribe-bentham/.

[37] NYPL Labs, What's on the menu?, 2022. URL: http://menus.nypl.org/.

[38] Uppsala universitetsbibliotek, Personverser, 2022. URL: https://ub.uu.se/bibliotekskataloger-a-till-o/personverser/.

[39] Artportalen, 2022. URL: https://www.artportalen.se/.

[40] M. Terras, For Ada Lovelace Day, Father Busa's Female Punch Card Operators, 2013. URL: http://melissaterras.blogspot.co.uk/2013/10/for-ada-lovelace-day-father-busas.html.

[41] R. Eveleth, Computer Programming Used to Be Women's Work, Smithsonian Magazine, 2013. URL: http://www.smithsonianmag.com/ist/?next=/smartnews/2013/10/computer-programming-used-to-be-womens-work/.

[42] J. Nyhan, Gender, knowledge, and hierarchy: on Busa's female punch card operators, 2014. URL: http://archelogos.hypotheses.org/135.