ARTICLE

Temporalities and Values in an Epistemic Culture: Citizen Humanities, Local Knowledge, and AI-supported Transcription of Archives

Dick Kasperowski[1]* , Karl-Magnus Johansson[2] and Olof Karsvall[2]

[1]Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg, Sweden; [2]The Swedish National Archives, Stockholm, Sweden

Abstract

An enormous amount of handwritten documents in archives can only be accessed by experts trained in reading older handwriting. Through artificial intelligence (AI)-supported technology, they can now be transcribed and made available for wider audiences. To produce transcriptions an AI needs training and a feasible way is to invite citizens to fulfil such tasks. To understand how an epistemic culture develops in such work, this study conducted interviews with participants on how they associate value, meaning and recognise themselves as active epistemic subjects in relation to the project. Despite that the formation of an epistemic culture are beyond the influence and control of project owners, findings show a strong relation between participants' knowledge of local history, and personal and emotional ties to archival content, for achieving high quality in AI-transcriptions.

Keywords: *Citizen humanities*; *archives*; *artificial intelligence*; *handwritten text recognition*; *epistemic culture*

The concept of citizen humanities (CH) has in recent years been used to denote the involvement of the public in different aspects of digital humanities and archival research.[1] This aligns with developments within archival studies, where 'participatory archives' has been described as a reconceptualisation of archival practices that questions the inherent power dynamic between archivists and users,[2] changing the archivist's role towards community-based and participatory archiving, as a consequence of the digitalisation and democratising of archives.[3] Initiatives of public involvement include projects on large scale platforms accessible for global audiences, as well as more local arrangements catering for participants with more specific domain expertise.[4] This type of distributed work, beyond the boundaries of professional expertise, can be situated in a general and global context of an 'openness paradigm' encompassing the distribution of tasks in advanced knowledge production known as citizen science (CS) or, as in this paper, CH.

*Corresponding author: Dick Kasperowski, Email: dick.kasperowski@gu.se

Making archives available in digital form is a main task for most archive institutions. For decades, the archives have invested in digitisation, that is, by scanning and photographing physical documents. The Swedish National Archives, for example, holds more than 200 million raster image files of archival documents. In this way, the archives become available instantly, at any time, reducing the need to visit the reading rooms of the institutions. However, documents as raster images, especially handwritten documents, need to be read and interpreted manually.

Developments in artificial intelligence (AI) has changed the conditions for research based on handwritten sources. It also affects the roles of volunteer participants in CH. The enormous number of handwritten documents in the archives, which have long been reserved for experts trained in reading older handwriting, can now be transcribed by machine learning technology, known as handwritten text recognition (HTR). Access to large quantities of machine transcriptions would broaden and deepen research, and notably benefit local heritage and genealogical research. Including HTR and other AI technologies would therefore be an important step to take for the archive institutions. In particular, it would facilitate in-depth full-text searches of archives that are difficult to access and use today.

However, the potential of HTR technology involves a greater challenge: access to high quality training data. The machine needs to learn the language, semantics and handwriting that appear in the various archives. Machine interpretation is thus dependent on transcriptions created by humans manually, as 'ground truth' for training. There are digital platforms such as *Transkribus* that facilitate the work with HTR for non-programmers.[5] But transcribing by hand is time-consuming and requires experience in or knowledge of reading and interpreting older handwriting. Given the enormous amount of preserved historical documents in archives, the need for such transcription processes is far exceeding the available resources at most archival institutions.

A feasible way is to invite citizens to participate with their interest and knowledge of handwritten sources. Galleries, libraries, archives and museums (GLAMs) have a long legacy of promoting participation with the public, and have institutional aims to promote their collections and archives, attracting as wide an audience as possible.[6] This now includes initiatives to invite volunteer participants to either perform tasks that are usually carried out by professional scholars, or work that has never been done by paid employees.[7] Collaborative transcription now appears prominently within the field of participatory archives, as a way to share control of archival content curation with users who often identify themselves as stakeholders in relation to the archives' content.[8] Such work now also encompasses the use of AI, including the training and correcting of handwritten manuscripts for the use of HTR.[9] However, the cultural aspects of such work, particularly the epistemic cultures developing among volunteer contributors are largely unknown. To study the cultural aspects of CH implies to consider the values of participation developing among volunteer participants, values that often are beyond the influence of project owners. Studying epistemic cultures in relation to AI applications, such as HTR, will nuance the focus on optimisation of time and resources often associated with this technology.[10] As such cultural studies offer a new and much needed perspective on the often-occurring presumption that CH participants should be aligned with technology and its protocols for optimisation.

Purpose and research questions
The overarching purpose of this paper is to understand how an epistemic culture in CH develops. More specifically, this leads to the questions of this paper, namely: How do volunteer participants recognise themselves as active epistemic subjects? How do they associate meaning

and value to their engagement in transcribing training data and correcting the machine transcription of historical handwritten documents? To answer this, we interviewed volunteer participants engaged in a collaborative AI-supported transcription project initiated by the Swedish National Archives. The project, *The Detective Section*, invited volunteer participants to train an HTR to process 25,000 pages of handwritten text from the 19th century. The findings reported in this paper build upon accounts of volunteer participants' practices and experiences of the project.

Following recent studies of CH and the expected changes in AI-implementations,[11] we apply conceptual resources from studies of epistemic cultures, to understand how volunteer participants associate values and meaning when training an AI. The answers, we believe, will produce a nuanced understanding of why and how tasks are performed by volunteer participants in distributed heritage work involving AI. Eventually such understandings can point to how volunteers value participation. To consider how such practices tie in with the formation of values benefit reflections on the design and scale of future HTR projects. In other words, this study presents a meta-perspective on CH projects combined with AI, which we believe needs to be explored in more detail.

The present paper starts with a description of the studied case, *The Detective Section*, and its relation to CH. Next, the theoretical and methodological framework is outlined, situating our study in motivational studies in CH and CS. We then proceed with our theoretical resources before presenting our empirical results. This is followed by a discussion on the implications of our findings, how meaning and domain expertise in relation to the historical material is important for attaining quality in an HTR process. The paper concludes with how our results transcend standard configurations of volunteer participants in practices of archives and CH.

The Detective Section as a case of handwritten text recognition and citizen humanities

In November 2019, the HTR and CH transcription project *The Detective Section* (Detektiva avdelningen) was initiated at the Swedish National Archives in collaboration with GPS400: Centre for Collaborative Visual Research at the University of Gothenburg.

The archival material in *The Detective Section* included 25,000 pages from the Gothenburg Police department consisting of a series of handwritten police reports 1868–1902, and handwritten copies of received and sent telegraph messages 1865–1903.[12] The information in the material could potentially be of value to many research fields such as historical, cultural and linguistic studies, as well as for amateur research, for example family historians. The material was previously rarely used, however. Since it was not digitised, it had to be read in its original physical form in the reading room; and the catalogue only gave information about what year the records were from, not which persons, places or events that were mentioned. Therefore, to make this series available as fully transcribed text data would radically improve its accessibility and ability to be searched and used. The material was selected both in dialogue with three volunteer participants from a previous participatory project at the National Archives in Gothenburg, as well as with researchers at GPS400. Also, the layout of the handwritten text in the series, with spreads of plain running text, was taken into consideration in the selection since it tends to make the HTR process more efficient.

In the project, the HTR platform *Transkribus* was used to train an HTR model, that is, an algorithm that later on was used for automated transcription of handwritten text.[13] Initially, volunteer participants were invited to join the project and transcribe the training data in an online user interface in *Transkribus*. For this purpose, a mass email was sent on 2 February 2020 to the approximately 500 email addresses used during the last 8 years to inform the public about upcoming lectures at the Gothenburg branch of the Swedish National Archives. Volunteer participants were approached regarding an opportunity to

create new ways of conducting research on the local history of Gothenburg. The invitation stated that no previous knowledge was required besides basic computer skills; however, the ability to read handwriting from the years around 1900 would be an asset. The invitation described how the project would radically improve the material's accessibility and ability to be searched and used. It also clearly stated that the final transcriptions would eventually be published as text data on the website of the National Archives, open and free to use for all. Social media platforms were also used to invite volunteer participants with a similar but condensed message. Some of the social media posts went viral, for example, the Twitter invitation reached 7,800 users on the platform.

During an initial 3-month phase, more than 400 spreads with 165,000 words were manually transcribed by five volunteer participants. The transcribed text was then used as training data for an HTR model that was trained to a character error rate of 2.7%. Since the transcribed series in the project was going to be published as text data with as close to 100% correct transcription – as suggested by both the participants and the involved researchers – volunteer participants were again invited to the project, now to proofread and correct the rest of the pages after they were automatically transcribed by the HTR model.

A second invitation mass email was sent on 23 September 2020, specifically asking for participation with proofreading and correcting the automated transcribed documents. The invitation described the success of the initial phase of the project, and also gave examples of the content of the archival material. At the same time, the printed newsletter *Västanbladet* for the local genealogical association, *GöteborgsRegionens släktforskare* [The GothenburgRegion Genealogists] carried a presentation of the project, and the webpage for the Centre for Collaborative Visual Research at the University of Gothenburg described the project and invited volunteer participants to join. Eighteen new volunteer participants joined the project after the second campaign, and all of the previous participants chose to continue in the project.

At the beginning of the project, the project leader wrote a manual in consultation with two of the volunteer participants. The manual was sent to all new participants and included basic instructions on how to perform the tasks as well as recommendations on resources that might be helpful when working in the project, such as databases and encyclopaedias.

When designing the participatory aspects of the project, it was important for the project managers to leverage the experiences gained from previous onsite participatory projects at the National Archives in Gothenburg where participants were invited to interact with analogue material in workshops. Those experiences highlighted the importance of learning and social aspects of such projects, as well as the participants' account of a rewarding feeling of contributing to a greater good.[14] Therefore, starting from 2 months after the first invitation, monthly meetings with volunteer participants were arranged. Because of the restrictions following the coronavirus disease 2019 (COVID-19) pandemic, the meetings had to be digital (via Zoom). The agenda of the meetings was to discuss the work that was done in the project, to share experiences about the difficulties in the assessment of transcriptions and corrections, and to share stories contained in the documents and how they relate to or challenge historical knowledge. The volunteer participants were invited to the meetings by the project leader, and the lead researcher at GPS400 was always present. Other researchers that had studied the historical period of the material were invited to some of the meetings to contextualise the work in the project as well as to gain specific knowledge from the volunteer participants. Since the project's archival material reflects the inequalities of Swedish society at the time, experts on gendered crime and social injustice were among the invited scholars. They took part in the discussions adding contextual knowledge about the social structures in which the plaintiffs, defendants and witnesses of the police reports were part. Other invited researchers were scholars of language technology, criminology, and photography.

Theoretical and methodological considerations

Motives for participation in citizen humanities and citizen science

The theoretical framework for this study relies on the concept of epistemic cultures. It was introduced to nuance the distributed knowledge practices in institutional settings of science and research, and to take closer account of how members of such epistemic cultures realise themselves as active epistemic subjects.[15] In this study, these concepts are employed to create a more finely granulated understanding of the motives, values and notions of time among volunteer participants than what is usually offered by studies of participation in CH and CS.

Studies of CH have been concerned with classifying the activities and tasks performed by volunteers, commonly finding that participants are included in assignments of tagging, transcribing, categorising, mapping, georeferencing, contextualising, and translating empirical material.[16] The 'main tasks' in CH projects have been identified as involving refining and collecting data, and in some cases also contributing domain expertise in more collaborative roles of co-creating projects.[17]

The preoccupation with classifying volunteer participants' tasks into different forms of CH has been paired with an interest in what motivates them to engage in projects according to project aims, and the value to institutions of their contributions.[18] In pointing out new directions for 'crowdsourcing in the cultural heritage sector', the increased value for institutions,[19] the possibilities of technological development and understanding volunteer participants' motivation are often suggested as important to explore in the future development of best practices in participatory projects.

Recurrent findings in motivation studies include commitment, learning experiences, personal rewards, interest, purpose, addiction, and good cause.[20] These largely institutional approaches to understanding volunteer participation in the humanities and sciences often find themselves at home in managing and sustaining coordination of distributed work to facilitate productivity, efficiency and timesaving for scientific and cultural heritage institutions. Accordingly, volunteer participants are often invited with tasks that are open to anyone, regardless of training and knowledge,[21] with fine-tuned strategies to uphold motivation,[22] including retention, targeted invitations,[23] and using technology to survey, standardise and speed up processes of data collection and project performance.[24] Recent ethical discussions concerning motivational and retention strategies and the extensive instruction and tutoring of volunteers before being able to contribute to CS and CH projects, have found that they create unethically excessive demands on the time and effort of contributors. Concerns about the increased demands on volunteers to be more engaged, for some bordering on exploitation, will eventually be addressed by participants themselves, who will refrain from involvement and abandon projects.[25]

Institutional approaches often neglect or downplay how volunteer participants more dynamically engage with tasks and the material on hand. Universal frameworks and categories for assessing motivation among volunteer participants to achieve retention have also been suggested. Applying such categories, volunteer participants are configured as inclined to contribute to research, including benevolence (helping people within one's own circle), and self-direction (creating, exploring) as the most important motivators. Categories with a low ranking are self-enhancement motivations of power (gaining recognition and status), achievement (personal success), as well as conformity (adhering to social expectations), personal image or reputation.[26]

We suggest that such approaches and findings largely capture the temporal adaptation, or subordination of volunteer participants to contribute, imposed in speeding up the scientific process, justifying institutional motives, however, largely outside of participants' influence. The cultural perspective employed in this study holds that there are reasons to reconsider such

conceptual frameworks concerning participant's motives when understanding participatory cultural heritage work.[27]

A cultural perspective on distribution, tasks and time in citizen humanities

With an increased recognition of the capacity (and necessity) among citizens to be actively involved in research, largely facilitated by the digital development, some studies explore the epistemological ideals and values developing among volunteer participants, as different research fields configure to accommodate 'outsiders' as contributors.[28] The analytical resources offered by the perspective of epistemic cultures lead us to ask different questions regarding the volunteer participants' reasons for participating in CH projects.

The interest in volunteer motivation as a condition for successfully conducting such projects has overshadowed enquiries into what values and knowledge volunteer participants themselves develop as epistemic cultures are formed in such projects. Our conjecture, building on the limited number of studies of epistemic cultures in CS and CH,[29] is that meaning created by volunteer contributors transcends categories of motivation, retention and institutional benefits, as participatory initiatives are considered by museums and archives.[30]

Thus, what do volunteer participants do and how do they find value and meaning in what they do? It has been shown that participants' engagement results in practices and values to create meaning beyond the goals of a project as formulated by owners and initiators. Namely, content to be tagged, transcribed, classified among others, with the help of volunteer participants contain anomalies, surprises, uncertainties and previously unseen or experienced phenomena that spur different types of engagement and practices among volunteer participants.[31] These types of activities have been described in terms of 'epistemic stratification' occurring over time in projects, as actors are endowed with or find temporal epistemic meanings in training an AI.[32]

In recent research on elderly persons engaging in voluntary archival work, results point to the 'good relations' that volunteer participants develop to documents, technologies and the individuals present in historical documents.[33] These findings resonate with what anthropologists and sociologists have called *social time,* offering a conceptual framework of understanding time as 'multiple, heterogeneous and arising from unequal entanglements between various social formations'.[34] This perspective, how time is *made* meaningful and valued as it operates as mediator or intermediary for social collaboration – and coordination, we tentatively suggest, is a resource for understanding the epistemic cultures of CH and distributed cultural heritage work.

The active forming of relationships with time, including to the historical archival material, but also to the future users of the digitised archive, as well as to the task of training the AI for HTR, we postulate, are intimately connected to meaning and values developed in the epistemic cultures of volunteer participants. To be able to transform the historical handwritten documents of *The Detective Section* into text data, volunteer participants trained an AI that was fed with these documents. This transformation relies on the volunteer participants' ability to read sometimes difficult handwritten text. This capacity cannot be separated from their knowledge and interest in the historical time period, as well as the participants' familiarity, knowledge and use of archival and other resources assisting them in their work. The empirical question is what this task will produce in terms of time(s), not only to the material as such, but also in relation to the work requested of, and the time devoted to the project by volunteer participants. How time is made meaningful might also be different between individuals and tasks.[35]

If the assignment to train an HTR model and correct its automated transcriptions is experienced by volunteer participants as a number of well aligned intermediaries making the task

smooth and easy, for instance an easy to follow handwriting, an easy to use interface for training the HTR model among others, work might flow, but also be less challenging, and therefore uninteresting. However, if the same tasks are experienced as mediators to overcome, causing the volunteer difficulties, time will be experienced differently, both in relation to the historical material as well as time spent for assignments.

Interviews with volunteer participants

To understand how an epistemic culture develops in distributed AI-supported cultural heritage work, semi-structured interviews with seven volunteer participants in *The Detective Section* were conducted during March–June of 2022. Respondents were approached with the methodological intention of having as heterogeneous a sample as possible of respondents in order to search for commonalities and differences across this diversity and understand the epistemic culture. For this purpose, it was more interesting to examine a few cases in depth and the inclusion of seven respondents was based on methodological considerations that an interval of 6–8 respondents would result in empirical material that would be relatively independent of individual respondents personal or subjective opinions. Generalisations were thus not based on representation, but on comparisons of data from interviews with the aim of identifying similarities and differences between volunteer participants' accounts and arguing for the most likely interpretations of how time was made meaningful and valued. Interviews were coded and analysed by each of the authors independently following a triangulation approach, and each interview was analysed in full.

The aim was to identify themes that arose across interviews, striving for theoretical saturation working back and forth between theory and empirical data to identify shared values and meanings in relation to time.[36] This yielded an identification and preliminary coding of themes that was then compared and grouped together by the authors. This work resulted in four themes, as reported below, of how time was socially made in forming values and meaning by volunteer participants in the project.

Each interview session was conducted in the presence of the interviewing researcher, the project leader, and a volunteer participant. The sessions strived to utilise the interviews in a dialogic way between the individuals present. Conversations revolved around the participants' accounts of their communicative practices as volunteer participants in the project, how they developed and relied on earlier knowledge and experiences as they completed tasks, and how they found meaning and relevance in their involvement.

All interviews were performed with the informed consent of the respondents. All direct quotes from participants used in this paper have been provided following informed consent from respondents and according to the ethical guidelines established by the Swedish Research Council.[37]

All respondents were retired (over 65 years) and the majority had an education well beyond secondary school, in several cases at the university level, including one doctoral exam. The gender distribution was three women and four men. The respondent's contributions to the project ranges from transcribing and/or correcting less than 100 to more than 3,000 documents. Three of the respondents joined the project in the initial phase when the training data was transcribed, and four joined in the second phase. One of the respondents chose to end his engagement because of illness before the project was completed. All of the interviewed volunteer participants had previous knowledge and domain expertise in the use of archives as active genealogists or local historians. Thus, there existed a culture of shared interest and trust among the volunteer participants and the archive. This is exemplified by participation in public lectures and earlier collaborative projects initiated by the National

Archives in Gothenburg, and even in one case by being engaged by them as a lecturer for a public lecture. Some participants had also developed their interest and knowledge of history in university courses, and in authorship of books on local history. This resonates with earlier research on the demographics and education of participants in CS, which found that many participants are highly educated, upper-middle class, middle-aged or older, of higher income, and white.[38] This has also been the case with participants in CHs.[39] The contexts from which many CH projects emanate (genealogy, local history) – having been institutionalised at archives for decades – may prevent more inclusive and equitable participation, however also providing the necessary domain expertise.

Results

In this section, we present our results from interviews with the volunteer participants. This is followed by the four themes that arose across interviews on participants as epistemic subjects; communication and community, working with AI, multiple kinds of relational knowledge, and values in and beyond *The Detective Section*.

Communication and community

Volunteer participants' accounts of being invited into the project point to little or no experience of CS and CH or with training an AI in HTR. The values and meanings associated with participating and devoting time to the project are therefore of a different kind. Instead, it is the unknown, 'exciting' archive and its content that provides the main value. To learn about Gothenburg in the late 1800s and to develop generic knowledge to be applied elsewhere, 'to rummage around in the archive and see what you can find' as one respondent puts it. Or, that it is 'fascinating to follow the fate of humans in a material that brings individuals to life in ways that is not usually found in archives used by genealogists'.

This resonates with the accounts given by respondents on the value of taking part in the project. The particularities of the archive, where people from the past come alive in ways not encountered before, is a shared value emphasised by the volunteer participants, and, but to a lesser extent, the relevance for future research in improving the archive's accessibility and ability to be used and searched. The ground-breaking project of transcribing old handwritten documents with applied AI is not valued to the same extent.

It is apparent that established relations to the archival institution is an important aspect of joining the project. Many participants express having trust in the local branch of the National Archives in Gothenburg. Trust is not about the quality of data, so commonly addressed in CS and CH, but a trust in that the archival institution will facilitate the project in ways that will make taking part interesting and worthwhile. As a respondent with experience from earlier participatory projects at the archives, formulates it:

> There is so much to be found in the Archives, the project helps me to focus on something specific, that I did not know I was interested in. You get that for free, you can contribute, and you learn beyond that.

This form of trust, making things interesting and meaningful, is often grounded in former experiences of taking part in events and activities offered by the National Archives in Gothenburg. These include the established traditions of lectures and workshops offered since 2013. All respondents point to the newsletters as having been the place they were invited into the project rather than the invitations posted on social media, which virally reached many more potential volunteer participants.

The project leader of *The Detective Section* held regular Zoom-meetups with the participants, and invited lecturers to some of the meetings. For ethical and integrity reasons, email addresses were not shared between participants, and no other means of communication between participants, such as social media, was offered. Where volunteer participants required tutoring or encountered problems, the project leader was approached through email. Volunteer participants did not regard the on-line Zoom meetings as specifically discussing the transcription process, and the training of the HTR model. However, they referred to these meetings as opportunities to further develop their historical knowledge and interests. through lectures by invited researchers.

The Zoom meet-ups have been very nice. I felt really privileged to meet the invited researchers. That added value to the project.

Zoom meetings were described by respondents as 'always interesting', but not there for a dialogue on the specifics of AI-supported transcription. However, they created a feeling of belonging to a collective with a shared interest among volunteer participants, and of being needed and appreciated by the project leader and invited lecturers. Some respondents missed the opportunities of in-person meetings to communicate with other volunteer participants on the content of the archives and the local history of Gothenburg in the late 1800s. Physical meetings could have provided better conditions for such discussions, some respondents tentatively suggest, however the 'pandemic put a stop' to such initiatives.

Although the volunteer participants had earlier experience and domain expertise in genealogy and local history, taking part in the project did not inspire or facilitate communication in such networks on their behalf. One exception is a participant who presented the project in a bulletin for a genealogical society. However, several of the interviewed participants shared details of the content of the historical police reports with close friends and family via email and social media but mostly in personal encounters, which evoked 'interest and fascination'.

I've been talking to all of my friends about this project. I'm saying 'I'm doing something really fun', making everyone envious now during Covid.

Me and my friends often take walks in the city. For many places that we pass I can relate to cases and events in the police reports, and I'm always talking about that. It's almost like I'm doing guided city walks for my friends.

Despite this, no one has been able to recruit volunteer participants into the project from these personal networks. The main communication on the tasks distributed within the project, that is, correcting automated transcriptions, has taken place on an individual level between the project leader and volunteer participants.

I have sent texts that I do not understand to [the project leader] and then we have read [them] together.

One respondent wished for possibilities to communicate directly with developers at *Transkribus*. However, this is a rare exception, as Zoom meetings, but interactions with friends and family about the close encounters with individuals in the archives of *The Detective Section*, have been the main content shared with actors outside the project. Thus, the social making of time in the community of volunteer participants is largely defined by the fascination with the past, in a highly local or personal context.

Working with HTR

Although initially invited to transcribe handwritten text as training data for an HTR model, the overwhelmingly majority of time that the volunteer participants spent in the project was working with correcting text that the model had automatically transcribed.

In the interviews, participants closely describe their approach to accomplishing the task of correcting transcriptions as one of switching between levels. This is explained as a move back and forth between individual signs or letters, and the meaning of the text and the context. How this switching between levels occurs is dependent on the complexity of what the HTR model has transcribed:

> First you read to simply correct the straightforward mistakes of form and pattern recognition. The AI might have mistaken a letter because of an unusual form that can be assigned to a particular police officer's handwriting. But when the obvious mistakes are corrected, you have to go to the level of meaning and context, unless you are presented with something very complicated from the start, then you must go for the level of meaning right from the start, to grasp the correct reading of individual letters that the AI has made mistakes about. Then you need the meaning of the text, and maybe the context also.

The usual approach for correcting transcriptions is, however, not to start at the level of meaning as 'the handwriting is so messy'. Participants express the need to start with the single letters that the HTR model has failed to recognise, as it is not yet sufficiently sensitive to the different styles of handwriting in the archive. Moving up to the level of meaning and context is useful at the second stage of correcting. This is when you benefit from understanding the 'flow of the text'.

The repetitive nature of correcting transcriptions is brought up by several respondents. It is the difficult cases, the correct interpretation of vague styles of handwriting to train the HTR model, understanding meaning and text in relation to image and pattern, that challenges participants. However, the difficulties, for some impossibilities, of producing a flawless, perfect, correction is also recognised. Participants have developed slightly different methods of achieving accuracy, repeatedly switching between levels, sometimes letting the first corrected version of a text rest for weeks before returning to it:

> It is very hard to be completely perfect. At the first glance, I see 6–8 errors per page. Then I switch to the level of meaning and it is at this instance that I understand the text for the first time. Earlier, it was just word for word to get the correction of the transcription right, not on the level of meaning. There is no flow when you concentrate on the first correction. [...] When I switch to the level of meaning I find even more mistakes. Then I let the text rest for about two weeks before returning to it, controlling it letter by letter, finding at least one to two additional mistakes per page.

> It is about being true to forms of letters, not trying to write up some content. The AI recognises images and patterns, it is not about content and meaning. That is something for us humans.

This volunteer participant clearly states that reading on the level of meaning and the use of local historical knowledge is a means to attain the non-contextual pattern recognition ability of a well-trained HTR model. To be able to produce an HTR model that can be applied to different text material from the time period and travel extensively between archives, meaning and local context is necessary in producing a high quality HTR model.

The more complicated the handwriting, and the correction needed for the mistakes made by the HTR model, the more you benefit from knowledge about the local historical context.

Usually, the content of reports filed in the archive are not regarded as particularly complicated to read, but the level of difficulty varies as it is connected to the complexity of the handwriting. This is directly related to which scribe the writing can be assigned to, as his style of handwriting would entail recurrent 'unusual forms' of letters, misspellings, word sequences and general style and use of language. Some of the handwriting was regarded by participants and the project leader as the most difficult in the material. In such cases, the automated transcriptions could actually 'train' the participant, hence working with AI is a process of training and being trained:

> Some pages were very difficult to transcribe. To guide me, I checked the automated transcription done by the AI [...] so, you could say that I was also trained by the AI.

Furthermore, the names of streets and geographical locations, the stolen goods, the modus operandi of the crime etc., would provide context from which corrections to the transcripts could be derived. Even though participants had no experience in the type of distributed work that involves training an HTR model on complicated handwritten text, they all had expertise, and in some cases advanced expertise in reading historical handwriting. The added task of working with HTR did not therefore create any great difficulties, except for some very difficult personal styles of handwriting and linguistic conventions, but then the AI could actually be a resource, as its original translations could be checked for clues in how to correct the translations. However, the mistakes by the HTR model could also be a source of irritation:

> Punctuation makes me insane! Particularly when the AI has added it when it is not there in the original text.

The introductory manual was regarded as informative on HTR, but, relying on their domain expertise in deciphering handwritten text, a more common approach among participants was to directly try out the *Transkribus* interface in correcting the automated transcriptions. The value of the manual was instead associated with it being a guide to external resources for completing the task:

> I used several of the digital resources mentioned in the manual, such as historical census data, The Swedish Academy Dictionary and The Swedish State Calendar.

Individual participants also extensively used already cultivated external resources, ranging from Google, to databases at the National Library of Sweden, historical dictionaries, shipping lists, digitised newspapers, place name registers, Wikipedia, catechetical registers in church records etc. The familiarity with archival research and the domain expertise in genealogy and local history was hereby evident among participants.

Additional resources were also developed by individual participants and shared through the project leader and in Zoom meetings as the project progressed. This included a glossary of different textiles, as the most reported crime was the theft of clothes and other textiles, and a list of mortgage offices in Gothenburg. The result of this participants' initiative was an extensive description of over 100 historical terms that was then included in the manual as a resource for all.

Multiple kinds of relational knowledge

Despite their familiarity with archival research, and domain expertise in areas such as the local history of Gothenburg or in reading handwritten historical documents, participants are often

eager to point out a cognitive or social distance to 'proper' research, and identify as amateurs who appreciate the opportunity to be of use and provide assistance. In this context, many respondents also point out that their contribution is modest, and that the time spent on tasks in the project is not extensive, as 'I stop before I get bored'.

Increased and detailed knowledge about the local history, however, is a recurrent theme in the respondents' accounts of the value of participating in the project. One participant has, after enrolling in the project, started to re-read her vast collection of historical Gothenburgiana, 'with fresh eyes'. Participants encounter and learn about the fate of individuals, and the places and circumstances of petty crimes in the archives of *The Detective Section*, which they describe as 'exciting' and not encountered before. They also point to the wealth of documentation in the archive that extends beyond the reports on criminal activities and concerns the wider social and technological developments during the time period. Respondents frequently testify to cultivating a 'closeness' to the individuals present in the historical material.

> When you get close to the people you read about in the material, you sometimes feel that you *know* specific bicycle thieves.

Individual scribes of the Police administration are recognised through their use of language and handwriting. In some cases, the personality of the scribe is constructed from the handwriting:

> Someone, in the beginning I called him 'The Klutz', starts to write and then revise over and over again. Writing in between the lines with smaller and smaller letters. Impossible to decipher. It was like as if he was very uncertain about how to communicate and express himself, he seemed to lack self-confidence. His first draft usually was good, but then he always succumbed to extensive revisions. Such a conflict of ambition and ability.

Multiple kinds of relations to the historic material are made by volunteer participants. For yet another participant, with a professional background as a lawyer and judge in the 1970s, closeness is manifested in the familiarity with the language used in police reports: 'I have no difficulties in understanding the content of the material'. Instead, it is the difficult handwriting and era-specific grammatical conventions that constitute an historical challenge and motivate this participant: 'Legal text from the 1890s is not difficult to understand, but what is really meaningful and motivating is to have a difficult handwriting to decipher'.

The challenges of learning to interpret difficult handwriting, and the generic proficiency in this required for genealogical studies in general, is shared by participants: 'You become better in interpreting and reading old handwritten text'. The development of this skill is also connected to having insights into how AI is used to digitise historical archives. It is 'interesting to see how the AI is developing as I help to train it'. The accounts of how knowledge develops during the task of transcribing text and correcting the transcriptions of the HTR model often revolves around how participants develop hermeneutical skills to interpret and understand old handwritten text – and in some instances also the author behind the text.

The language used in the police reports is of central concern to the participants, who refer to the reports as narratives about individuals that you come close to in ways that are rare in historical documents: 'You get to know the individuals'. Geographical references, street names and buildings are spaces adding a closeness in time to the historical period, 'street names are the same as today, it feels very modern'. To have a connection to 'real'

people in a transformative period in Swedish history, 'the social and technological develop-ment of the period' is experienced as very rare and of a high value. The historical material, through its focus on the fate of individual lives, brings the history alive and closer to partic-ipants. In particular, the social inequalities displayed in the material seem to create empa-thy: 'The insights I got from the material was so interesting, yet so very tragic and sad'.

Some respondents describe an emotional connection to the subjects in the police reports, or, as described in another section above, to scribes identified from their specific handwriting. Recent discussions in archival studies highlight and theorise such connections. People working with archives form emotional relationships with individuals 'from' the archives, sometimes including very sensitive and personal details. Most archival records, and not least the police reports, have an 'intrinsic humanity'.[40]

It is through 'facts' (street names, geographical locations, legal jargon etc.), intertwined with the interpretation and emotions associated with the fate of individuals –'why are they committing these crimes'– that participants return to in explaining how they develop knowledge and meaning in the project. The joining of facts and emotions, text and con-text in the making of time, actually makes interpretation of difficult handwriting, and the training of the HTR model, possible. In fact, the more you develop your knowledge about the local history of both police officers and their individual use of language and handwriting as well as the delinquents' deeds and fates, the better you can fulfil the task of training the HTR model.

Values in and beyond The Detective Section

For volunteer participants, taking part in *The Detective Section* has enabled them to get access to an historical archive they have not before considered or even known about. As a result, they have been introduced to and developed new knowledge of the local history of Gothenburg:

> I have learned so much about the time period from this project.

> Learning was an important part of expectation when I joined this project … now I have learned so much and found new interest.

Some participants also point to the understanding they have gained in the role of AI and HTR, but foremost, the high value associated with the opportunity to follow individual cases, 'people' that come alive in an 'exciting archive', and thereby to deepen their understanding of the local history of Gothenburg. Being given the opportunity to understand the workings of an historical public authority and its contemporary context is highly valued by participants. This active forming of 'good relations' to the material and individuals in the historical archive is closely connected to meaning and value among volunteer participants.[41] Some considered this the sole value of taking part:

> The project has created value for me in creating access to the archive and through the tasks of developing the AI. I don't think so much about the wider significance of the work that we have put in.

Others regard their participation as 'making a difference' for the future use of the archive:

> Many will, once the archive is digitised, be able to search through the archive, using names, places, time. There is a need, future genealogists will be able to find their relatives and what they were up to. You are contributing to a common good.

Everyone who has any experience of archive use has also experienced the difficulty of access. What has been done in this project might be important for future generations that would be interested in increased access and searchability of archives. It will be so much easier to have a digitised archive to search and it will be easier for those who do not have the knowledge background, but would like to develop it.

For some, their engagement has spurred further, more specific historical interest in the role of law enforcement. The archive of *The Detective Section* depicts petty crime, mostly thefts, and volunteer participants now ask where reports on other, more serious aspects of law enforcement can be found and investigated. This provides an opportunity to initiate and develop new projects, building on and developing the knowledge on local history that has been gained from taking part in *The Detective Section*. In one case, a volunteer participant initiated their own parallel project out of curiosity about the subjects described in the police reports and started researching them in other archives to find their date of birth, date of death, spouses, children, and even photos of them. The result was a database with information about hundreds of historical individuals, which the volunteer participant then sent to the project leader for distribution to students and others that might find it useful. This parallel project, as well as the glossary list of textiles described earlier, are examples of how the epistemic culture of the project supported participants to initiate and do things beside the main protocol of training an AI.

One respondent is also asking about the possibility of assuming roles in projects beyond correcting automated transcriptions, and whether volunteer participants could be given the opportunity to choose which archives should undergo digitalisation. This also includes the question of how distributed cultural heritage work will be initiated and governed in the future. What roles would archives play in this? What would be the ethical concerns? However, the majority of respondents point to the value of participation *per se*, trusting the archives to facilitate and develop their domain expertise and interests. Here, the participants express an attitude of not questioning their position in the participatory structure, as long as other values in the epistemic culture can be nurtured.

In addition to developing the participants' historical knowledge, and improving their ability to read difficult handwriting while training and developing an HTR model, the value of taking part in the project thus also extends to expectations on the part of the participants of more active participation and access to archives in the future. This connects to the perception among participants of the value of developing AI in making archives accessible to a larger general public and to improve the conditions for future research. Free accessibility to digital archives is emphasised: 'It is important that money is not made from our work. It is a matter of trust'.

Sweden has a long tradition of open access to information in archives, with the world's oldest regulation regarding public access to official records dating back to 1766. Personal and other kinds of sensitive information can be subjected to secrecy, but, not for longer than 70 years. Also, the Swedish implementation of General Data Protection Regulation (GDPR) can limit the distribution of information, but only regarding persons still alive. The material of the project dates from 1865 to 1903, and thus are juridically granted open access.

However, in addition to the juridical aspects, the notion of access is far from neutral.[42] As most of the volunteer participants are part of the family history-community in Sweden, they can be seen as stakeholders in the transcription process that leads to open access. One of the interviewed participants exemplifies this clearly when telling that her husband's family was living in the poorest area of Gothenburg at the time. Another interviewee revealed that his

great-great-grandfather was a police officer at the time, but lost his position because of alcohol consumption at work. For both the public access to such information, was a most important reason to even join the project and spend time in it.

Discussion

The respondents stated that local knowledge is a key aspect of understanding this remediation of data. This aligns with studies that challenges whether being digital is being independent from local constraints. Such a perspective suggest that data are entangled within a knowledge system and inscribed in a place, and that all knowledge systems are rooted in practices and politics related to their time and space – that all data are local.[43] This study also acknowledges that these are important aspects to consider in regards to how data quality is attained in distributed cultural heritage work. A large number of studies have been devoted to the question of how to engage volunteer participants in scientific work without compromising the collection and classification of data,[44] and how to facilitate the development of skills that ensure the quality of data. Usually these discussions lead to recommendations for having a low cognitive threshold for volunteer participants, thereby minimising the need for instruction and learning. However, projects such as our *The Detective Section* also rely on the existing domain expertise of the volunteers.[45]

Danielsen et al. (2005) argue that 'locally-based methods are generally more vulnerable than professional techniques to various sources of bias', suggesting 'thorough training' as a solution.[46] However, extensive training is expensive, time consuming and demands infrastructural solutions, therefore the answer often found, according to Danielsen, is to create stable protocols that put volunteer contributors on par with professionals with regard to their tasks. If such stabilisation cannot be attained, professionals, researchers or other actors, including machines (like *Transkribus*) relying on advanced knowledge, will remain sceptical about the results.[47] Volunteer participants before autumn 2020 had created training data for an HTR-model that reached 97.3% certainty in transcribing the handwritten text. As the project was striving for even higher certainty in the HTR transcription ability, it needed more help. The interviews show that the volunteer participants are aware that higher certainty is attained through more persistent work rooted in local knowledge of history and interests in the specific source material. This is despite the fact that AI (HTR) often makes invisible what knowledge and relation to the material is needed to produce an even higher data quality.[48]

Such nuances of delegated work are largely missing in discussions of the significance of domain expertise in distributed cultural heritage work for creating high quality in HTR. It is not only the level of standardisation in the interface or participatory protocol that determines the data quality achieved by volunteer contributors.[49] The focus of our study – how volunteer participants realise themselves as active epistemic subjects – yields that the development of relations to the historical archive illustrates how quality in HTR is improved in ways not usually part of discussions in CH.

Taking account of our empirical results, we have to consider both the spatial aspects of the term 'local' but also its temporal aspects. These temporal features include the relations formed not only with a technological interface or protocol but also with the conditions in which those data have been manifested, that is, close and personal relationships with a local archive. Here respondent accounts of meaning are associated with the stories of individual fates in the historical material. This is important for arriving at a more finely granulated understanding of the training of an AI. It is also clear that the local knowledge systems of connected resources such as databases, other archive series, and encyclopaedias are important when performing

such tasks. These circumstances are partly supported by findings in both large globally distributed as well as in smaller scale local participatory projects. Volunteer participants will engage dynamically with the material at hand, often beyond the tasks they have been invited to perform, and will create new resources to share among members.[50] However, this study clearly connects such dynamic relations (of time and local circumstances) to the quality of the work asked for – training an HTR model. Furthermore, emotional relations seem to be important for both creating meaning as well as for the quality of training of the HTR. Emotions are not usually considered important or relevant to include in research, in fact the opposite: they should be avoided or controlled so as not to create bias, thereby contrasting the usual standards for the rigours of knowledge creation.[51] In fact, Danielsen et al. associate 'locally-based' methods with bias.[52]

On the other hand, the accounts from the volunteer contributors in this study show that relations, emotions and empathy together with local historical knowledge are at the core of creating an HTR with high accuracy. Inviting participants with domain expertise in the local history to work together with local historical archives is important for the quality of HTR. The participants' interview responses point to locally 'rooted' projects, in the sense of archival material and voluntary participants with domain expertise, as well as trusted relationships to archival institutions, being favourable. For distributed projects in science and the humanities, open archives and data repositories tend to invite individuals to facilitate efficiency and speed in science and research. In this way, they tend to follow dominant narratives of acceleration so often reiterated in the justification of open science and open data, namely increased production of knowledge and research. However, as the data from the (albeit a limited amount of) in-depth interviews in this study show, the design of projects in archival HTR benefit from a perspective sensitive of time as not exclusively influenced or defined by dominant narratives that describe time as uniform, external to participants and in a state of continuous acceleration.[53] The results align with recent reconceptualisations within archival studies, where indigenous scholars question the pace of archival work and suggest that slowing down creates possibilities to emphasise how such epistemological processes are entangled with a series of relationships.[54] Such 'slow archives' work makes time for people-centred and reflective approaches, and, as shown in this study, this is to the benefit of data quality. These are aspects of an epistemic culture that should be considered in relation to participants' local domain expertise and interests in creating improved digital access to archives.

Conclusion

The purpose of this paper has been to understand how an epistemic culture develops in a distributed CH-project, namely training an AI in HTR. To this end, interviews were conducted with volunteer participants to gain insights into how participants associate value and meaning in relation to the historical material as they transcribe training data and correct automated transcriptions of handwritten documents.

The central finding in this study is the relationship between the volunteer participants' knowledge of local history and their achievement of high quality when correcting the automated transcriptions of the archived material prepared by the HTR model. A recurrent narrative is the participants' accounts of specific local historical knowledge as an important asset for the quality of the correction of transcriptions. The more you develop your knowledge about the local history, in fact establish a personal and emotional relationship to the police officers and scribes, as well as delinquents, the better you can accomplish the task of training the HTR model. The participants' accounts of their engagement

with the project show that values and meaning are formed in developing a relationship to the historical material.

Rather than considering optimisation from a technical perspective, we have investigated it in relation to the epistemic culture and the way volunteer participants realise themselves as active epistemic subjects, foremost how they associate meaning and value to their engagement. This perspective has also been recognised in recent studies on participatory AI, acknowledging that communities and citizens beyond technical designers have knowledge and interests that would benefit such projects.[55] The specific tasks at hand, of transcribing training data and correcting automated transcriptions, are not referred to by respondents as the main reason or the meaning of engaging in the project. Volunteer participants are curious about AI and HTR, but it is the relationships formed, bridging time, and the possibilities of discoveries in the archival material that are of central value for the respondents. However, the participants emphasise the importance of generally increased digital access to archives, and in that sense, HTR is a key factor.

These results transcend the categorisations often associated with volunteer participants' motives for taking part in CH. In this case, *The Detective Section,* volunteers have been invited with the task of working with AI in transcribing historical material. Thus, the project is contributive in this respect. It is also collaborative in refining data and creating resources for volunteer participants and furthermore develops domain expertise among them. To be a volunteer participant in *The Detective Section* is to create meaning as you engage and situate yourself in, but also beyond, such categorisations.

Author biographical notes
Dick Kasperowski
Professor of Theory of Science at the University of Gothenburg. His interests include citizen science, governance of science, participatory and activist practices in science and the humanities and open collaborative projects in scientific work. The analytical focus of his research concerns how new technologies configure relations and the development of epistemic cultures between actors claiming different experiences and knowledge.

Karl-Magnus Johansson
Senior archivist at the Swedish National Archives. In practice and theory, his main interests are understanding engagement in and use of archives, as well as the intersection between media theory, contemporary art and archives.

Olof Karsvall
Research manager at the Swedish National Archives and PhD in Agrarian History. He has been working in several research projects concerning digitalisation, digital methods and research data at the National Archives.

Notes

1. Joni Adamson, 'Gathering the Desert in an Urban Lab: Designing the Citizen Humanities', in Joni Adamson and Michael Davis (eds.), Humanities for the Environment: Integrating Knowledge, Forging New Constellations of Practice, Routledge, London, 2018, pp. 106–19.

2. Edward Benoit III and Alexandra Eveleigh, 'Defining and Framing Participatory Archives in Archival Science', in Edward Benoit III and Alexandra Eveleigh (eds.), Participatory Archives: Theory and Practice, Facet, London, 2019, pp. 1–7.

3. Terry Cook, 'Evidence, Memory, Identity, and Community: Four Shifting Archival Paradigms', Archival Science, vol. 13, 2013, pp. 113–16; Craig Gauld, 'Democratising or Privileging: The Democratisation of Knowledge and the Role of the Archivist', Archival Science, vol. 17, 2017, p. 227.

4. Dick Kasperowski and Thomas Hillman, 'The Epistemic Culture in an Online Citizen Science Project: Programs, Antiprograms and Epistemic Subjects', Social Studies of Science, vol. 48, no. 4, 2018, pp. 567–68.

5. https://readcoop.eu/transkribus/

6. Compare with Melissa Terras, 'Crowdsourcing in the Digital Humanities', in Susan Schreibman, Ray Siemens and John Unsworth (eds.), *A New Companion to Digital Humanities*, John Wiley & Sons, Chichester, 2016, p. 6.

7. Amy Clothworthy, 'The Experience of the Citizen Scientist', Edited interview with Amy Clotworthy, Danish National Archives, 2019b, available at https://training.parthenos-project.eu/wp-content/uploads/2019/05/Amy-Clotworthy-Interview-on-the-Citizen-Scientist-experience-April-2019.pdf; Tim Causer, Kris Grint, Anna-Maria Sichani and Melissa Terras, '"Making Such Bargain": Transcribe Bentham and the Quality and Cost-Effectiveness of Crowdsourced Transcription', Digital Scholarship in the Humanities, vol. 33, no. 3, September 2018, pp. 467–487. see Danish National Archives' crowdsourcing portal https://cs.rigsarkivet.dk.

8. Sumayya Ahmed, 'Engaging Curation: A Look at the Literature on Participatory Archival Transcription', in Edward Benoit III and Alexandra Eveleigh (eds.), Participatory Archives: Theory and Practice, Facet, London, 2019.

9. Compare with Melissa Terras, 'Inviting AI into the Archives: The Reception of Handwritten Recognition Technology into Historical Manuscript Transcription', in Lise Jaillant (ed.), Archives, Access and Artificial Intelligence: Working with Born-Digital and Digitized Archival Collections, Bielefeld University Press, Bielefeld, 2022, p. 188.

10. Compare with Gregory Rolan, et al., 'More Human than Human? Artificial Intelligence in the Archive', Archives and Manuscripts, vol. 47, no. 2, 2019, p. 186.

11. Compare with Jonathon Hutchinson, 'Digital Intermediation: Unseen Infrastructures for Cultural Production', New Media & Society, August 2021.

12. https://riksarkivet.se/psidata/goteborgs-poliskammare

13. For a thorough presentation of Transkribus and its impact on research, see Guenter Muehlberger et al., 'Transforming Scholarship in the Archives through Handwritten Text Recognition: Transkribus as a Case Study', Journal of Documentation, vol. 75, no. 5, September 2019, pp. 954–976.

14. Karl-Magnus Johansson, 'Hungerkravallerna 1917: deltagande i fokus i arbetet med ny utställning i Göteborg', Nordisk Arkivnyt, no. 2, 2017; Karl-Magnus Johansson, 'Medskapande, design och film: ny arkivutställning i Göteborg', Nordisk Arkivnyt, no. 1, 2019.

15. Karin Knorr-Cetina, 'Culture in Global Knowledge Societies: Knowledge Cultures and Epistemic Cultures', Interdisciplinary Science Reviews, vol. 32, no. 4, 2007, pp. 361–375; Karin Knorr-Cetina, Epistemic Cultures: How the Sciences Make Knowledge, Harvard University Press, Cambridge, MA, 1999.

16. Stuart Edale Dunn and Mark Charles Hedges, 'From the Wisdom of Crowds to Going Viral: The Creation and Transmission of Knowledge in the Citizen Humanities', in Christothea Herodotou, Mike Sharples and Eileen Scanlon (eds.), *Citizen Inquiry: Synthesising Science and Inquiry Learning*, Routledge, London, 2017; Dick Kasperowski, Christopher Kullenberg and Frauke Rohden, 'The Participatory Epistemic Cultures of Citizen Humanities: Bildung and Epistemic Subjects', in Palmyre Pierroux, Per Hetland and Line Esborg (eds.), A History of Participation in Museums and Archives: Traversing Citizen Science and Citizen Humanities, Routledge, London, 2020.

17. Kasperowski et al., p. 238; Nina Simon, The Participatory Museum, Museum 2.0, Santa Cruz, 2010.

18. Terras, p. 8; Lesley Parilla and Meghan Ferriter, 'Social Media and Crowdsourced Transciption of Historical Materials at the Smithsonians Institution: Methods for Strenghtening Community Engagement and Its Tie to Transcription Output', The American Archivist, vol. 79, no 2, 2016, pp. 438–460.

19. Ibid, p. 13.

20. Ibid, pp. 13–14; Rose Holley, 'Crowdsourcing: How and Why Should Libraries Do It?', *D-Lib Magazine*, vol. 16, no. 3–4, March 2010, p 1–21.

21. Kasperowski et al., p. 237.

22. Anne Land-Zandstra, Gaia Agnello and Yaşar Selman Gültekin, 'Participants in Citizen Science', in Katrin Vohland, et al. (eds.), *The* Science of Citizen Science, Springer, Cham, 2021, p. 247 ff.

23. Barbara Heinisch, et al., "Citizen Humanities", in Katrin Vohland et al. (eds.), The Science of Citizen Science, Springer, Cham, 2021, pp. 97–118.

24. Bálint Balázs et al., 'Data Quality in Citizen Science', in Katrin Vohland et al. (eds.), The Science of Citizen Science, Springer, Cham, 2021, p. 152.

25. Loreta Tauginienė et al., 'Ethical Challenges and Dynamic Informed Consent', in Katrin Vohland et al. (eds.), The Science of Citizen Science, Springer, Cham, 2021, p. 408; Dick Kasperowski, Niclas Hagen and Frauke Rohden, 'Ethical Boundary Work in Citizen Science: Themes of Insufficiency', *Nordic Journal of Science and Technology Studies*, vol. 10, no. 1, 2021, p. 18.

26. Land-Zandstra et al., p. 249.

27. Compare with Amy Clotworthy, 'Engaging the Human in Digital-Humanities Projects: How Participating in Crowdsourcing Projects Impacts Quality of Life among Volunteers at the Danish National Archives', Paper presented at 4th Digital Humanities in the Nordic Countries, Copenhagen, Denmark, 2019a; Kasperowski and Hillman, 2018; Kasperowski et al., 2020.

28. Knorr-Cetina, p. 367; Kasperowski et al., 2020.

29. Ponti, M., Kasperowski, D. & Gander, A.J. Narratives of epistemic agency in citizen science classification projects: ideals of science and roles of citizens. AI & Soc (2022). https://doi.org/10.1007/s00146-022-01428-9

30. See Mia Ridge et al., *The Collective Wisdom Handbook: Perspectives on Crowdsourcing in Cultural Heritage*, community review version (1st ed.), British Library Publications, London, UK, 2021.

31. Kasperowski and Hillman, 2018, p 582. Full reference for Latour is missing. Should be Bruno Latour Reassembling the social: an introduction to actor-network-theory. Oxford New York: Oxford, University Press, p. 5.

32. Compare with Ponti et al., 2022, p. 13.

33. Clotworthy 2019b, pp. 1–4.

34. Larissa Pschetz and Michelle Bastian, 'Temporal Design: Rethinking time in design', *Design Studies*, vol 56, 2018, pp. 169–184, 172.

35. Latour Reassembling the social: an introduction to actor-network-theory. Oxford New York: Oxford, University Press, p. 5.

36. See Svend Brinkmann and Steinar Kvale, *InterViews: Learning the Craft of Qualitative Research Interviewing*, 3rd ed., Sage Publications, Los Angeles, CA, 2015; Favourate Y. Sebele-Mpofu, 'Saturation Controversy in Qualitative Research: Complexities and Underlying Assumptions: A Literature Review', *Cogent Social Sciences*, vol. 6, no. 1, 2020, 1838706.

37. Swedish Research Council Expert Group on Ethics, *Good Research Practice*, Swedish Research Council, Stockholm, Sweden, 2017.

38. Karen Purcell, Cecilia Garibay and Janis L. Dickinson, 'A Gateway to Science for All: Celebrate Urban Birds', in Janis L. Dickinson and Rick Bonney (eds.), *Citizen Science: Public Participation in Environmental Research*, Cornell University Press, Ithaca, NY, 2012; Rajul Pandya and Kenne Ann Dibner (eds.), *Learning through Citizen Science: Enhancing Opportunities by Design*, The National Academies Press, Washington, DC, 2018; B. Troy Frensley et al., 'Bridging the Benefits of Online and Community Supported Citizen Science: A Case Study on Motivation and Retention with Conservation-Oriented Volunteers', *Citizen Science: Theory and Practice*, vol. 2, no. 1, 2017, 1–14; Elli J. Theobald et al., 'Global Change and Local Solutions: Tapping the Unrealized Potential of Citizen Science for Biodiversity Research', *Biological Conservation*, vol. 181, 2015, pp. 236–44; Hillary K. Burgess et al., 'The Science of Citizen Science: Exploring Barriers to Use as a Primary Research Tool', *Biological Conservation*, vol. 208, 2017, pp. 113–120; Dale R. Wright et al., "Understanding the Motivations and Satisfactions of Volunteers to Improve the Effectiveness of Citizen Science Programs", *Society & Natural Resources*, vol. 28, no. 9, 2015, pp. 1013–29; Mari Jönsson et al., 'Long-Term Trends in Age and Gender Participation in an Online Biodiversity Citizen Science Project', accepted for Ambio, 2023.

39. Chiara Bonacchi et al., "Participation in Heritage Crowdsourcing", Museum Management and Curatorship, vol. 34, no. 2, 2019, pp. 166–82.
40. Jennifer Douglas, et al., '"These Are Not Just Pieces of Paper": Acknowledging Grief and Other Emotions in Pursuit of Person-Centered Archives', Archives & Manuscripts vol. 50, no. 1, 2022, pp. 5–29.
41. Compare with Clotworthy, 2019a and Clothworthy, 2019b.
42. Agostinho, Daniela, 'Archival Encounters: Rethinking Access and Care in Digital Colonial Archives', Archival Science, vol. 19, 2019, pp. 141–165.
43. Yanni Alexander Loukissas, All Data Are Local: Thinking Critically in a Data-Driven Society, MIT Press, Cambridge, MA, 2019.
44. Myriah L. Cornwell and Lisa M. Campbell, 'Co-Producing Conservation and Knowledge: Citizen-Based Sea Turtle Monitoring in North Carolina, USA', Social Studies of Science, vol. 44, no. 1, 2012, p. 105.
45. Hauke Riesch and Clive Potter, 'Citizen Science as Seen by Scientists: Methodological, Epistemological and Ethical Dimensions', Public Understanding of Science, vol. 23, no. 1, 2014, pp. 107–120.
46. Finn Danielsen, Neil D. Burgess and Andrew Balmford, 'Monitoring Matters: Examining the Potential of Locally-Based Approaches', Biodiversity and Conservation, vol. 14, no. 11, 2005, pp. 2524–2542.
47. Ibid, p. 2527.
48. Compare with Susan Leigh Star, 'This Is Not a Boundary Object: Reflections on the Origin of a Concept", Science, Technology, & Human Values, vol. 35, no. 5, 2010, p. 607.
49. Jeffrey P. Cohn, 'Citizen Science: Can Volunteers do Real Research?', BioScience, vol. 58, no. 3, 2008, p. 194.
50. Compare with Kasperowski, Kullenberg and Rohden, 2020; Clotworthy 2019a; Clotworthy 2019b.
51. Compare with Minna Santaoja, 'Insect Affects: A Study on the Motivations of Amateur Entomologists and Implications for Citizen Science', Science & Technology Studies, vol. 35, no. 1, 2022, pp. 58–79.
52. Danielsen, Burgess and Balmford, 2005, p. 2524.
53. Compare with Pschetz and Bastian, 2018, p. 174.
54. Kimberly Christen and Jane Anderson, 'Toward Slow Archives', Archival Science, vol. 19, 2019, pp. 87–116.
55. Abeba Birhane et al., 'Power to the People? Opportunities and Challenges for Participatory AI', Paper presented at the Equity and Access in Algorithms, Mechanisms, and Optimization Conference, October 6–9, 2022, George Mason University, Arlington VA, USA