



Participant Learning and Open Archives Through Artificial Intelligence and Citizen Humanities

DICK KASPEROWSKI 

KARL-MAGNUS JOHANSSON 

**Author affiliations can be found in the back matter of this article*

COLLECTION:
GLAM

RESEARCH ARTICLES

 ubiquity press

ABSTRACT

Online archives host millions of scanned historical documents. However, images of old handwriting remain unsearchable and largely inaccessible without transcription. These limitations can now be mitigated by advances in Handwritten Text Recognition (HTR), which uses machine learning to convert images of handwriting into text data. In the HTR and Citizen Humanities project The Detective Section, the Swedish National Archives invited volunteers to build and refine a large, high-quality training dataset for an HTR application.

This qualitative interview study explores how participants in The Detective Section perceived their involvement in the project in terms of learning and knowledge development. By applying a typology for understanding multifaceted learning and drawing on conceptual resources from science and technology studies (STS), the study highlights how participants perceived themselves as developing into epistemic subjects within an epistemic culture.

We found that participants learn both individually and collectively, often in unexpected ways, through and beyond predefined goals and protocols. Their strong emotional investment and interest in local history leads to additional activities, which in turn result in more accurate transcriptions. We conclude by discussing refinements to participatory practices within the GLAM sector that can support volunteer learning and help democratize access to archival collections.

CORRESPONDING AUTHOR:

Dick Kasperowski

University of Gothenburg,
Sweden

dick.kasperowski@gu.se

KEYWORDS:

Handwritten Text Recognition;
Machine Learning; Citizen
Humanities; Learning; Open
Archives

TO CITE THIS ARTICLE:

Kasperowski, D and Johansson, K-M. 2026. Participant Learning and Open Archives Through Artificial Intelligence and Citizen Humanities. *Citizen Science: Theory and Practice*, 11(1): 34, pp. 1–13. DOI: <https://doi.org/10.5334/cstp.911>

INTRODUCTION

Millions of scanned images of handwritten archival documents are available online. However, for most users, these materials remain largely inaccessible. The contents of such documents are not searchable, and they are frequently written in historical scripts that are difficult to decipher. Manual transcription, which could address these challenges, is a resource-intensive process and has not been a priority for archival institutions. Recent advances in a field of artificial intelligence (AI) known as handwritten text recognition (HTR) are poised to transform this situation. HTR relies on machine learning techniques, whereby an AI application is trained to automatically transcribe handwritten documents, significantly enhancing their accessibility and utility for research and public use.

In November 2019, the HTR and citizen humanities project The Detective Section (Detektiva avdelningen) was launched at the Swedish National Archives (SNA), in collaboration with the University of Gothenburg. The project focused on a series of approximately 25,000 pages from the archive of the Gothenburg Police Chamber, comprising handwritten police reports from 1868 to 1902, along with telegrams sent and received between 1865 and 1903. The content of this material holds considerable potential for research in numerous fields, including historical studies, cultural studies, and language technology, as well as for genealogical and local heritage research. However, the archival series had seen limited use, as the documents had not been digitized and were accessible only onsite in the reading room of the SNA in Gothenburg. Furthermore, the archival catalogue offered information about only the year each report book was produced, without any indexing of the individuals, locations, or events referenced therein. Making these report books available as transcribed text data would therefore significantly enhance their accessibility, searchability, and scholarly utility.

HTR forms part of an ongoing development within the archival sector's efforts in digital transformation, with the aim of enhancing access to archival information (Colavizza et al. 2021). A common feature of AI applications is the need for large volumes of high-quality training data—something that archival institutions rarely have the resources to produce by themselves (Rolan et al. 2018). This makes the field particularly conducive to collaboration with volunteers from the public, a form of engagement that is likely to expand and have considerable impact. In The Detective Section, the public was invited to participate specifically in the creation and improvement of data for training AI models.

In recent decades, a range of initiatives have emerged with the aim of making archives more accessible

by involving volunteers to collaborate with archival institutions and scholars. Today, there are numerous online platforms that support such collaborations (cf. Zooniverse, CrowdHeritage, FromThePage). Projects frequently involve discrete tasks in digital environments, such as the transcription of handwritten documents or the annotation of images (Ridge 2014; Ridge, Blickham, and Ferriter 2021; Heinisch et al. 2021).

In developing citizen science methodologies, there are often expectations of increased efficiency and the production of high-quality, usable data. As a result, many projects establish detailed protocols for task execution. A limited number of studies examine how volunteer participants are engaged in HTR projects to create groundtruth training data for machine learning models. These studies tend to focus on the instrumental benefits of involving participants in data production, aligning with a broader tendency in citizen science research to optimize outcomes (Wecker et al. 2020; Tarride et al. 2023). However, participation in the cultural heritage sector may generate effects beyond mere data production, not least for the participants themselves (Ridge 2013). Herein, we focus on such effects, particularly in the form of learning and knowledge development that participants associate with their involvement and co-creation.

THE DETECTIVE SECTION

The archival material for the HTR and citizen science project The Detective Section, was selected in dialogue with three volunteer participants from a previous participatory project at the SNA, together with researchers from the University of Gothenburg. The layout of the handwritten texts in the archival series—consisting mainly of running text, as in Figure 1—was also considered in the selection process, as it makes the HTR process more efficient.

The platform Transkribus was used to train the HTR model, that is, an algorithm for the automated transcription of handwritten text (Mühlberger et al. 2019; Terras 2022). Initially, volunteer participants were invited to transcribe training data via a web interface in Transkribus. No prior knowledge beyond basic computer skills was required, although the ability to read early 20th-century handwriting would be an asset. The invitation to participate clearly stated that the final transcriptions were to be published as text data on the website of the SNA, open and freely accessible to all.

During the first three months, more than 400 spreads—comprising 165,000 words—were manually transcribed by only five volunteer participants. These transcriptions were then used as training data for an HTR model, which achieved a character error rate of 2.7 percent. Since the transcriptions were to be published as text data with near

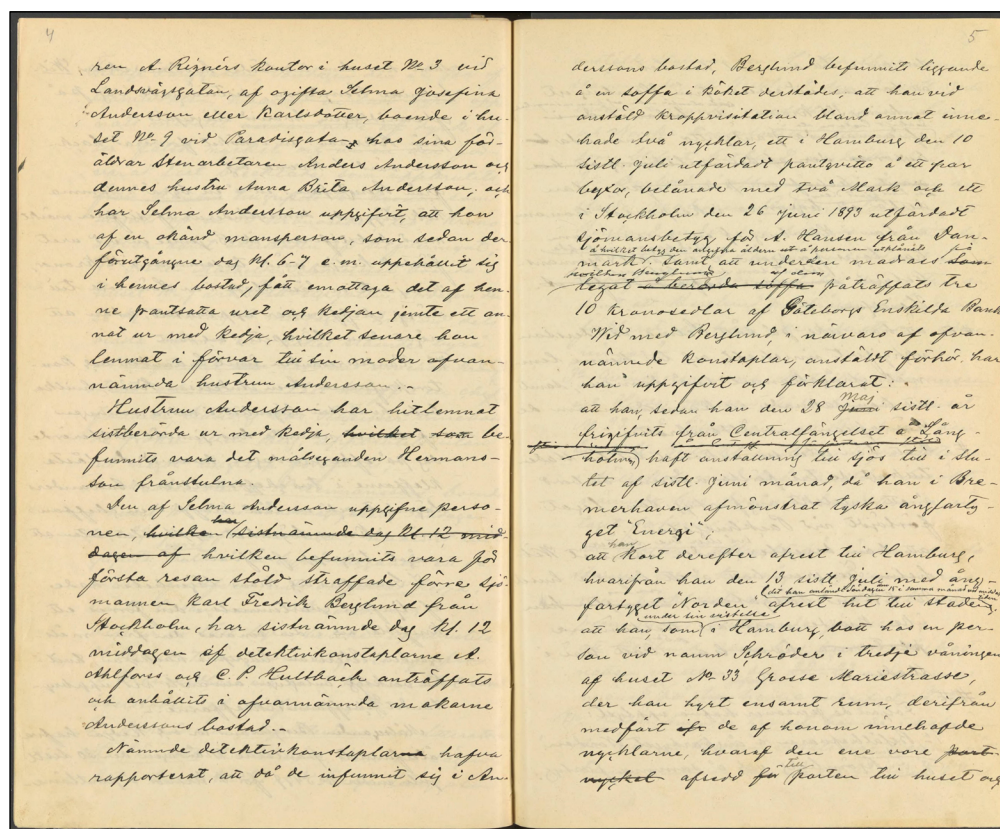


Figure 1 This example spread from the archival material shows a police report from 1894. Source: Göteborgs poliskammare före 1900, A II: 26 (The Swedish National Archives in Gothenburg).

100 percent accuracy and serve as training data for further HTR development, additional volunteer participants were invited to correct the remaining pages after they had been automatically transcribed by the HTR model. Eighteen new participants joined the project following the second invitation, and all previous participants chose to continue.

In designing the project, the project leader aimed to build on experiences from previous participatory projects at the SNA in Gothenburg, where participants had been invited to engage with analogue materials in workshop settings (Johansson 2021, 2019b, 2017). Evaluations of these earlier projects had highlighted the importance of collective learning and social interaction, emphasizing participants' notion of contributing to something of shared value. As a result, monthly discussion meetings were planned in the Detective Section, which, due to COVID-19 restrictions, were held online. The purpose of these meetings was to promote a sense of community, to reflect on the ongoing work within the project, to share experiences related to the challenges of evaluating transcriptions and corrections, and to discuss the information contained in the documents—how these relate to, or challenge, existing historical knowledge.

A researcher from the University of Gothenburg was always present at the meetings, and additional researchers who had studied the historical period of the

material attended some of the sessions. Their role was to contextualize the project's work and to engage in dialogue with the participants to gain specific insights into the material. Invited scholars were experts in gender studies, social history, language technology, criminology, and photography.

As part of the research project Transcription node Sweden (Transkriberingsnod Sverige), a study was conducted in 2023 to better understand the epistemic culture that emerged within The Detective Section (Kasperowski et al. 2024). The study was based on in depth semi-structured interviews with seven participants, focusing on four thematic areas: communication, working with AI, the participants' domain expertise, and the values they associated with their participation. Although the interview questions only briefly address the topic of learning, the conversations largely revolved around the development of knowledge that participants experienced and valued as part of the project.

Herein, we draw upon and return to the interview material from that study, now with a specific focus on learning. The purpose of this study is thus to examine how the participants in The Detective Section perceived their involvement in the project in terms of learning and knowledge development.

CONCEPTUAL CONSIDERATIONS

There is no shortage of studies on learning in citizen science. Recurring themes include how participants' learning and motivations relate to where and how effectively they can and wish to engage in the research process. At least three dimensions of learning have been central in these studies. (1) The learning and the development of skills specifically related to tasks participants are mobilized to perform, aimed at ensuring high-quality data. Such studies often (2) explore increased knowledge and self-efficacy regarding the specific scientific fields or disciplines in which projects are situated, such as ecology, biology, or astrophysics. Furthermore, studies on participant learning in citizen science have expanded to include (3) learning about scientific principles, methodologies, and the nature of scientific knowledge itself. This broader perspective encompasses how participation in citizen science can cultivate increased trust and interest in science by influencing stewardship and motivation. Scientific literacy frequently emerges as a primary goal when learning objectives are established and assessed. This substantial interest in participant learning has led researchers to seek effective methods for assessing learning outcomes, prompting calls for standardized quantitative approaches to measuring citizen science learning outcomes (Philips et al. 2018). Such approaches have facilitated comparative research across the three dimensions of learning outlined previously (Roche et al. 2020).

Few studies have investigated how participants collaboratively shape their own learning while actively training algorithms for HTR. To address this, we draw on conceptual resources from science and technology studies (STS) to highlight how participants in The Detective Section perceive themselves developing as epistemic subjects. The concept of epistemic culture was introduced in STS to analyze how the increasing delegation of tasks within research practices influences participants and their knowledge development (Knorr-Cetina 1999, 2007). Earlier studies of participants in citizen science have found that the delegation of tasks creates incentives for them as epistemic subjects and encourages them to develop resources that advance collective knowledge production. The interaction among epistemic subjects shape how epistemic cultures are constituted (see Kasperowski and Hillman 2018).

In this article, we focus specifically on how participants evolve as active epistemic subjects, and understand themselves, other participants, technologies, and resources. This approach also represents a shift away from a previously dominant perspective in citizen science studies, where distributing tasks to actors outside academia—or, as in our case, cultural heritage and memory institutions—

has primarily been framed in terms of institutional benefits and formalized learning.

We have adopted a perspective foregrounding the volunteers' own perception of their participation, exploring how knowledge development among participants can surpass the tasks initially assigned (Kasperowski et al. 2020). New technologies often raise concerns regarding deskilling and disempowerment (Rafner et al. 2021). As sophisticated technologies become integrated into citizen science, it is essential to address such concerns explicitly and explore how AI can augment, rather than diminish, volunteers' skills (Sharma et al. 2024).

We respond to calls to study the role of volunteer participants in AI design and development, particularly regarding how engagement in such processes empowers individuals to understand AI not merely as users but as co-creators (Duerinckx et al. 2024). In doing so, we seek to move beyond conventional perspectives that emphasize unidirectional flows of knowledge—where humans simply provide labeled data to train AI applications. Instead, we explore and illustrate a bidirectional relationship shaped actively by participants. Within this relationship, machines not only augment and extend human capabilities, but humans actively augment and extend machine capabilities as well (Østerlund et al. 2024).

We argue that highlighting these dynamics is important for cultural heritage institutions, especially as they increasingly seek to improve accessibility through advanced digital technologies like AI. Participants' own perspectives on their learning experiences and contributions to HTR applications are often rendered invisible by big data practices, despite widespread expectations of enhanced accessibility and openness in relation to archives and cultural heritage collections (see Shelton et al. 2024 for an exception).

METHOD AND EMPIRICAL MATERIAL

The empirical basis for this study consists of semi-structured interviews conducted between March and June 2022 with seven participants (hereafter R 1–7) in The Detective Section at the SNA in Gothenburg (Kasperowski et al. 2024). The interview guide is provided in Supplementary File 1: Appendix – Interview Guide. Respondents included as heterogeneous a sample as possible, facilitating the search for commonalities and differences within this diversity, thereby enabling a deeper understanding of the epistemic culture. Seven respondents were included based on methodological considerations, given that this study adopts a case-oriented, interpretive design. The purpose of the interviews was not to measure the prevalence or

distribution of opinions, but to generate empirically rich material on how participants are oriented to, engage with, and contribute to an epistemic culture.

In this sense, the interviews were designed to explore the situated practices, interpretive habits, and relations to epistemic objects (handwritten manuscripts) through which knowing is accomplished in this setting. Given this aim, the adequacy of the sample was assessed in relation to its specificity, the research questions, and the interpretive depth sought in the analysis. This position is in line with the argument that sample-size rationales should be grounded in the aim and purpose of the study. It is also supported by the notion of information power, which holds that smaller samples may be sufficient when respondents are closely aligned with the research focus and can provide material of high relevance to the study, as was the case in The Detective Section (Braun and Clarke 2006). Thus, generalizations were not based on representation but on comparing interview data to illustrate similarities and differences between volunteers' accounts and arguing for the most plausible interpretations of how learning was perceived. Interviews were coded independently by each author following a triangulation approach, with each interview analyzed in full to identify themes that transcended individual participants (Brinkmann and Kvale 2015; Favourate 2020). The interviews were conducted as dialogues and focused on the participants' experiences of the project.

All interviewees were retirees over the age of 65, confirming findings from previous studies indicating high levels of participation from this age group (Shelton et al. 2024). One participant had an education equivalent to compulsory secondary school, while the majority had attained education beyond the upper secondary level, in several cases at university level, including one doctoral degree in engineering. The gender distribution was three women and four men.

The participants' contributions to the project varied, ranging from having transcribed and/or corrected fewer than 100 documents to more than 3,000. Three of the interviewees joined the project during the initial phase when training data were being transcribed, while four joined during the second phase when automatic transcriptions required correction.

TYPOLOGIES FOR LEARNING IN CITIZEN SCIENCE AND CULTURAL HERITAGE

Review literature on crowdsourcing within the cultural heritage sector frequently identifies learning as a significant outcome of activities (Ridge et al. 2021). However, there

are few explorations of the cultures of such learning. Research in the field of citizen science has produced many studies focusing on learning within citizen science projects. (Kloetzer et al. 2021; Kasperowski et al. 2020). Most existing research has concentrated on participants' learning in areas such as environmental science, biodiversity, and sustainability. Learning within participatory projects in the humanities and cultural heritage remains relatively underexplored, particularly in relation to the development of advanced digital technologies like AI.

As a methodological tool for articulating, studying, and evaluating learning in citizen science, typologies have been developed that are based on intended learning outcomes. Two such typologies, which have inspired our work, are largely derived from citizen science projects within the natural sciences (Philips et al. 2018). They include categories that capture increased knowledge, awareness, and understanding of scientific concepts and processes; engagement and interest; attitudes toward science; competencies grounded in scientific concepts or methodologies; and behaviors related to science.

A common feature of both typologies is their emphasis on learning related to concepts, processes, and roles within the scientific domain of citizen science projects, not in the epistemic cultures of participants. A third alternative typology has been proposed for examining learning in online citizen science projects—collaborative initiatives in which participation occurs through digital applications, such as annotation or transcription platforms. What distinguishes this typology is its inclusion of learning about extra-scientific phenomena. It differentiates between learning that is subject-specific and learning beyond the immediate scientific domain, and it also considers personal development, extending beyond scientific competencies alone (Kloetzer et al. 2013).

Within the field of heritage education, the Heritage Learning Outcomes (HLO) framework has been developed as a tool for planning and evaluating various tasks. The framework is intended to support cultural heritage institutions in becoming more inclusive and relevant. HLO is based on the Generic Learning Outcomes framework, which was developed within the UK museum sector during the 1990s. The HLO framework categorizes learning into five areas, which partially align with the two previously mentioned typologies of learning in citizen science. These categories include: knowledge and understanding; skills; attitudes and values; enjoyment, inspiration, and creativity; and behavior and progression. HLO explicitly aims to make learning *through* heritage visible, rather than merely learning *about* heritage (Hansen 2014). Such an approach has also been recognized within the field of archival pedagogy (Grut 2016).

Based on the four typologies discussed above, we employed the following typology in our coding and analysis of the interviews to categorize the forms of learning that emerged in conversations with participants in The Detective Section. The term “tasks” refers to the specific activity that participants were invited to carry out: the transcription of handwritten texts from the late 19th century to generate training data for a HTR model, as well as the correction of the model’s transcriptions. Our aim is to understand both the learning about these specific tasks and the learning that emerged through task execution and through participation in the project more broadly.

Drawing on the conceptual resources offered by the notion of epistemic culture, and inspired by previous typologies of learning, we coded and analyzed the interview responses using the following categories:

- Knowledge, awareness, and understanding
 - (a) *about* the tasks and (b) *through* the tasks.
- Skills (a) required to carry out the tasks and (b) developed *through* the performance of the tasks.
- Engagement, interest, inspiration, and creativity.
- Attitudes and values.
- Behavior and personal development.

RESULTS: MULTIFACETED LEARNING IN THE DETECTIVE SECTION

All participants shared a common background as current or former genealogists. Individual skills brought into the project included the ability to read historical handwriting, local historical knowledge and interests, and, in one case, professional legal expertise. None of the participants had prior experience of training an HTR model or with any other form of AI, nor did the invitation to participate require such knowledge or experience. Many participants cited “curiosity” as a key motivation for joining—particularly an interest in the unknown archival material—and a desire to learn “how the digitization of archives works.” This motivation aligns with previous studies of participants’ reasons for engaging in transcription projects initiated by memory institutions (Shelton et al. 2024). Several participants also indicated that their limited understanding of AI contributed to their decision to participate.

KNOWLEDGE, AWARENESS, AND UNDERSTANDING ABOUT THE TASKS

None of the participants reported difficulties with the digital interface. However, one recurring challenge noted by several was the principle of source-faithful transcription. This principle requires that the transcribed and corrected

text replicate the original as closely as possible in terms of spelling, punctuation, capitalization, and other typographic features. Participants were instructed to transcribe exactly what was written, with no corrections or normalization allowed. Even obvious mistakes in the source documents had to be transcribed faithfully.

Several participants expressed initial discomfort with the requirement to “misspell” words deliberately, but those who raised this concern also acknowledged that such accuracy is necessary when training an HTR model. This illustrates how participants learned about the relationship between manual transcription work and AI, and about the importance of high-quality training data for the functionality and performance of HTR applications.

One participant (R2) articulated this point very clearly, emphasizing that the tasks are about being faithful only to the forms (of letters and signs), not attempting to compose or interpret content. HTR is understood by participants as based on pattern recognition: The AI must be trained to recognize images or visual patterns, not to understand meaning or content.

Through this process, participants developed a sense of themselves as epistemic subjects in relation to advanced technology. The insight of “gaining an understanding of how AI works” (R5) was regarded as particularly valuable by participants.

KNOWLEDGE, AWARENESS, AND UNDERSTANDING THROUGH THE TASKS

All participants agreed that their involvement in the project led to a deeper historical understanding of Gothenburg and Sweden in the late 19th century, particularly from a social-historical perspective. This was repeatedly described as the most valuable aspect of participating in the project. The documents from the police archives primarily depict phenomena and social groups that are invisible in traditional historical narratives, offering a unique insight into the social marginalization many individuals experienced during this period. The emotional challenge of confronting this knowledge was constitutive of the project’s epistemic culture. One respondent described gaining new insights into the social stratification of society while being emotionally affected by the accounts of poverty and hardship found in the archival material: “Some days you feel sick reading it” (R7).

Through the transcription process, participants not only learned about the technical development of the period, the emergence of modern infrastructure, and the evolution of the Swedish legal system; they also became increasingly inducted into a particular epistemic culture. In working with the archival material, they learned to engage with historical traces as objects of knowledge, relating fragmentary details in the records to broader sociohistorical developments. In

this sense, the project did not simply transmit information about the past, but enabled participants as epistemic subjects to acquire a mode of attentiveness characteristic of historical knowledge-making. When asked what had been especially valuable about the project, one respondent answered: “I’ve learned so much about the time period, about Sweden’s path toward democracy” (R4).

The project thus fostered not only a deeper understanding of the historical period, but also an emerging familiarity with the archives as a knowledge-setting with its own procedures, absences, and epistemic possibilities. Although some participants had prior archival experience, mainly through genealogy, none had previously worked with material from the police archives. This engagement provided what was described as “an entirely different perspective” (R1) and led to a “realization of how much was actually documented in the late 19th century” (R4). At the same time, participants’ encounters with the material made visible that archival knowledge is necessarily partial. Their reflections extended beyond what the records contained to what they excluded, suggesting an increasing sensitivity to the incompleteness of the archives and to the questions generated at the limits of documentation: “The material raises questions. What are we not seeing? How can we find more information?”

Archives held by archival institutions are typically organized according to the principle of provenance, which means that records are preserved in accordance with their original administrative context (Douglas 2010). As a result, information about a specific person, event, or location may be spread across various archival fonds and series, depending on who created the records and for what purpose. The principle of provenance therefore often requires knowledge of administrative history to effectively locate relevant information—a process that can be difficult and complex, even for experienced archive users (Burell and Sjögren 2018, p. 334). Through the knowledge gained in the project, several informants report having become more aware of how to search for information in archives and how different records and series relate to one another. One of the interviewees highlights as one of the most important lessons from the project the understanding of how the city’s archival sources—although separated by the principle of provenance and therefore organized into different archival fonds—are nevertheless “interconnected” (R5).

ABILITIES TO PERFORM THE TASKS

In the interviews, participants describe different ways of engaging in transcription and correction work. The material shows how participants are constituted as epistemic subjects within the project’s epistemic culture; that is, through their practical, affective, and interpretive relations

to the archival documents, the HTR model, and the auxiliary resources mobilized to secure source-faithful text data. Participants describe how they developed routines for correcting the HTR model’s automatic transcriptions, and in doing so, they reveal several shifts between text, context, and the visual forms of characters.

Correction is generally described as a movement back and forth between individual characters or letters as visual forms and the meaning or semantics of the text, as well as its broader context. This process depends on how difficult the handwriting in each document is to interpret, and participants move between different linguistic and contextual levels when correcting the machine-read text. Most do not begin at the semantic level because “the handwriting is so messy” (R7) and difficult to read. Instead, the point of entry is often the individual letters, that is, the forms the HTR model has failed to recognize. In this phase, visual attention tends to precede semantic interpretation.

Once these more fundamental errors have been corrected, the entire transcription is reread at semantic and contextual levels in order to identify additional errors introduced either by the HTR model or by the reviewer. One participant describes letting the text rest, sometimes for several weeks, before undertaking this final step (R4). During this proofreading phase, further resources may be needed to establish context. These include the SNA’s free online resources, as well as tools such as emigrant lists, historical dictionaries, genealogical handbooks, and address calendars. According to participants, the value of these resources lies in their capacity to validate uncertain words or spellings: “I often double-check geographical names, occupations, and archaic terms. It has to be correct” (R6).

The strong interest in tracing individuals in the archival material during the semantic reading of the texts points to a deepening engagement with the historical world to which the documents refer. This engagement also appears to enhance the conditions for source fidelity. Difficult-to-read handwriting that depends on contextual knowledge, such as street names, geographical locations, and legal terminology, can be more successfully deciphered when combined with participants’ growing familiarity with the historical setting and with the affective responses evoked by the lives documented in the archives. In this way, such engagements contribute to higher-quality training of the HTR model (Kasperowski et al. 2024). The opportunity to develop close bonds with individuals in historical documents, “to know a new person,” (R2) has also been identified as an important source of motivation in other transcription projects (Owens 2013; Shelton 2024 et al.).

From the perspective of the epistemic subject, these accounts suggest that accurate transcription is not only a matter of technical skill or individual competence. Rather,

it emerges through sustained engagement with epistemic objects that are at once incomplete and unfolding, and through the subject's attachment to the documentary traces, auxiliary resources, and historical persons that make interpretation possible. In this material, in-depth historical knowledge and emotional engagement with the actors in the archive support the deciphering of difficult-to-read handwriting and, in turn, the training of high-quality HTR models (cf. [Douglas et al. 2022](#)).

ABILITIES DEVELOPED THROUGH PERFORMING THE TASKS

Participants highlight the capacities they have developed through their involvement in the project, most notably an improved ability to read late nineteenth-century handwriting. Some also describe an expansion of their digital competence, for example by learning how to participate in online meetings or by offering specific suggestions for how the Transkribus interface could be made more user-friendly. Rather than presenting these developments simply as individual skills, participants largely frame them as achievements that emerged through their engagement with the project's tasks, tools, and epistemic objects.

As discussed previously, participants emphasize that understanding the content of the archival material is essential for producing high-quality transcriptions. When encountering certain types of information, such as names of persons, places, or companies, they make extensive use of additional sources and resources to support transcription and correction. Some of these resources, such as census databases, were already familiar to most participants before the project. Others were introduced through recommendations from the project leader, either in the project manual or during online sessions. Participants also independently identified additional resources valuable to their work, including reference books and information from other archives. In these accounts, the ability to search for information is closely linked to the ability to assess and validate it.

Several interviewees describe this "detective work" as intellectually stimulating: "I've gotten much better at searching for information—the challenges are the most fun part" (R4). This work largely involves navigating resources that contain large and heterogeneous bodies of information, which in turn places demands on participants' capacity for critical source evaluation: "I search a ton online and in libraries. You really must be source-critical, of course" (R4). From the perspective of epistemic culture, such practices are not merely supplementary to transcription work. Rather, they form part of the situated interpretive routines through which learning and new skills are developed. In this sense,

participants emerge as epistemic subjects through their ongoing relations to handwritten manuscripts, digital tools, and auxiliary resources.

ENGAGEMENT, INTEREST, INSPIRATION, AND CREATIVITY

Participants describe the digital meetings as a key reason for their engagement in the project and as something that cultivated a sense of community: "I feel like I'm not alone" (R1). However, many also found the individual work to be highly engaging. One participant shared that, despite many years of genealogical research, it was only through this project that they developed a broader interest in history: "I've actually never been interested in history, but now something just clicked" (R4). Another participant described the archival material "like a history book, but more alive" (R3).

Several participants recounted how the knowledge of archival material gained through the project helped them earn a greater understanding of Gothenburg's spatial development and continuity. This translated into new forms of engagement and inspired them to share the insights they had developed:

I'm learning so much. I had no idea about certain local historical details, but the project has given me a whole new perspective. When I'm out walking with friends, I often talk about cases that happened at the places we visit. It's almost like guided tours. (R1)

The content of the project's material could be directly linked to the spatiality of the city. Interest in and knowledge about the local context thus became part of the learning process. One participant reported, "I've learned so much about [the city district] *Majorna*," while another became curious about changes to the street layout: "Did *Södra Vägen* really go that far back then? It certainly doesn't today" (R2).

The tasks assigned to participants were clearly defined: to faithfully transcribe handwritten text and to correct automatically transcribed text. The limited digital interface in which participants worked allowed for little deviation in how the tasks were performed, and the manual distributed at the beginning of the project was oriented toward a single objective: how the tasks should be carried out as simply and accurately as possible. Yet despite this narrow framing, both collective and individual initiatives emerged in which participants, as epistemic subjects, exceeded the boundaries of the assigned tasks.

During several digital meetings, participants noted how frequently textiles were mentioned in the police reports—everything from lost scarves to stolen winter

coats and pawned wool sweaters. As a result, a collective decision was made to create a shared “textile glossary” within the project, which became a valuable resource for all participants. By the end of the project, this glossary contained 60 terms with descriptions and, in some cases, links to relevant entries in museum databases.

In another instance, one participant initiated a parallel project focused on the individuals mentioned in the police reports. These individuals were traced through other archival sources to identify details such as dates of birth and death, spouses, children, and, in some cases, photographs. The result was a database containing information on hundreds of individuals who lived in Gothenburg in the late nineteenth century, which the participant then shared with the project leader for distribution to students and others who “might benefit from it” (R3). These examples illustrate how participants, in creative ways, developed resources that extended well beyond the framework initially established by the project, a framework shaped in part by the constraints of the HTR technology. One participant reflected on the kinds of questions that emerged in the discussions between researchers and participants during the shared digital sessions: “The project has raised research questions that could be explored further, for instance, in new citizen science projects” (R4).

ATTITUDES AND VALUES

The design of the study makes it difficult to systematically trace how participants’ attitudes and values changed over the course of the project. However, the interviews do contain examples that suggest such shifts, particularly in relation to the archival material and to historical knowledge production. One participant noted: “I’ve reread all my old books about Gothenburg, which have now acquired a new meaning” (R7). Another described how they were initially frustrated by the lack of “blood” in the police reports, finding that they mostly concerned “boring” petty thefts (R6). Over time, however, this participant came to see that patterns emerge through the accumulation of what first appears trivial, and that such material is in fact essential for understanding the living conditions of the period. Details that might seem minor or insignificant from a present-day perspective had become meaningful: “You gain a new perspective when, for instance, you come to understand what a cap could signify” (R1).

Several respondents also reflect on broader social changes in Sweden. The participant who independently researched the personal histories of those accused noted that many of them died in their forties and concluded: “It really confirms that the social reforms that followed were truly necessary” (R3).

BEHAVIOR AND PERSONAL DEVELOPMENT

Some participants experienced the task of correcting automatically transcribed text as monotonous and somewhat dull. However, one of them, who joined the project during its first phase and chose to remain involved, emphasized that the correction process continually presents challenges that demand attention and effort, making the work engaging and worthwhile:

I enjoy the “micro-problems” that come up – it’s as fun as solving sudoku [...] Overall, it’s enjoyable and pleasant [to take part in the project], I work for an hour or so every morning. I don’t do more than I want to—I stop before it gets boring. (R4)

Another participant simply stated that “transcribing is insanely fun” (R2). Given that the project’s core tasks might easily be perceived as repetitive, it is noteworthy that, when asked, participants did not express a desire for progression toward greater responsibility or more demanding tasks. Most instead explained that the seemingly monotonous tasks are continually reanimated by the character of the archival material and by the transcription difficulties that arise in relation to it: “you always learn something new” (R2). In line with earlier studies, participants are not simply motivated to transcribe anything whatsoever; rather, they are mobilized by particular historical interests and by a project that invites them to engage with and learn through those topics (cf. [Shelton et al. 2024](#)).

One respondent, reflecting on possible future projects, expressed an interest in more distributed roles and shared decision-making (R4). The remaining interviewees, however, stated that they appreciated being invited into an already structured project and did not feel a need to influence its topic or scope. Instead, they expressed confidence that the archival institution could design engaging projects that, in turn, would enable participants to develop new interests and forms of knowledge—interests they may not previously have identified or been able to articulate:

Projects like this show just how incredibly much there is to discover [in the archives], so many different topics and so on. The project points you in a direction, and you get some free knowledge, and then you contribute, and you learn [...] What I like is that—you might already have an interest, and then you get nudged, and suddenly it becomes interesting, or maybe at first you think “oh well,” but the more you learn, the more interesting it gets. (R5)

Participants are not primarily seeking advancement in the form of hierarchy or expanded responsibility. Rather, they

are drawn into sustained engagement with epistemic objects whose incompleteness and difficulty continually generate new problems, interpretations, and attachments. In this sense, participants emerge as epistemic subjects not only because they move beyond routine work, but because routine work itself becomes epistemically productive through their ongoing relations to archival materials, transcription problems, and the interpretive practices through which these are handled.

CONCLUDING DISCUSSION

By adopting a cultural perspective on learning, this study has sought to understand how participants realize themselves as subjects within an epistemic culture. The delegation of tasks both individualizes and collectivizes learning within the HTR and citizen humanities project The Detective Section. The results illustrate how multifaceted learning emerged both through and beyond predefined goals and protocols.

Applying a typology for analyzing learning, we find that learning occurred in numerous ways, was experienced as informal and self-directed, and frequently arose due to situational demands—for instance, encountering difficulties and challenges in transcribing training data and correcting automatically transcribed text. These challenges were described as stimulating and brought together two complementary motivations: the desire to deepen historical understanding, and the ambition to produce highly accurate transcriptions.

Developing knowledge, awareness, and understanding of the tasks—and the capacity to perform it—is often individual. A clear individual component lies in how participants develop approaches to transcribing and correcting texts with the purpose of training the HTR application. The inspiration to independently develop methods aligned with prescribed protocols arises from the motivation to produce high-quality text data.

However, learning that emerges from the tasks often reveals collective dimensions. Knowledge, awareness, and understanding of historical contexts—spurred on through task completion—are discussed and disseminated within and beyond the group. Sometimes, collective initiatives create shared resources beneficial to all participants.

When participants reflect on the significance of their contribution, another, more expansive and temporal collective dimension emerges, as they perceive the project as meaningful and view their work as “a small part of something larger” (R6). Importantly, participants associate their work with future genealogists, who they believe will benefit greatly from the project’s outcomes,

enabling searches for names and other information within previously inaccessible archival material. Some take an even broader perspective, viewing their contributions as part of a democratization process extending accessibility to those unfamiliar with archival search methods or historical handwriting interpretation:

The public will really benefit from this in the future. Those who are not as experienced will find it so much easier to access the material. (R2)

Thus, individual learning—developing methods to ensure accurate transcription—resonates at a broader collective level, driven by the goal of creating accessible resources for present and future archival users.

This illustrates that within distributed collaborative cultural heritage projects defined and guided by the needs of AI technology, diverse forms of learning and knowledge development emerge among participants. The results indicate a cultural potential for heritage education, with the learning generated through participation being highly valued by participants and serving as a significant source of motivation. Through archive users’ participation, new and significantly improved opportunities to access and utilize archival materials are co-created, cultivating heritage education and contributing to making cultural heritage more accessible and meaningful to the broader public.

This qualitative study contributes insights into learning to earlier survey-based studies examining motivations for participation in virtual volunteer transcription projects ([Shelton et al. 2024](#)) and extends the discussion to participatory cultural heritage initiatives that include AI applications.

The study builds on existing understandings of learning in citizen science and the GLAM sector but foregrounds the fact that learning cannot be fully captured as a set of predefined or expected individual outcomes associated with knowledge, skills, attitudes, or behavior. Our analysis extends these understandings by locating learning within an epistemic culture in which it emerges through participants’ relations with tasks, archival materials, digital technologies, other participants, and the historical worlds encountered through the records. In this respect, the distinction between learning about the task and learning through the task is particularly important. For citizen science research, this advances an understanding of participants not simply as recipients of scientific knowledge or providers of data, but as epistemic subjects who actively shape the conditions of knowledge production, including the capacities of AI systems. For GLAM institutions, the findings extend approaches to heritage learning by demonstrating that participation can generate learning through heritage

while simultaneously contributing to the interpretation, enrichment, and increased accessibility of collections.

Memory institutions could consider these results by facilitating the epistemic culture that complements individual and collective learning and inspires participants to explore new historical areas, share knowledge, develop a sense of closeness to historical individuals, and enhance the accessibility of collections. The act of transcribing handwritten materials requires significant focus and evokes learning, immersing individuals in a personal and intimate encounter with the past.

DATA ACCESSIBILITY STATEMENT

Interview data from this study will be made available for research purposes upon request from the authors. The interview guide is provided in Supplementary File 1: Appendix – Interview Guide.

SUPPLEMENTARY FILE

The Supplementary file for this article can be found as follows:

- **Supplementary File 1.** Appendix – Interview Guide.
DOI: <https://doi.org/10.5334/cstp.911.s1>

ETHICS AND CONSENT

This study is conducted as academic research in accordance with the Swedish Ethical Review Act (*Lag (2003:460) Lag om etikprövning av forskning som avser människor*). In Sweden, an ethical review by the National Ethical Review Authority (*Etikprövningsmyndigheten*) is mandatory only if the research meets specific “statutory thresholds.” This study does not meet these thresholds since interviews focus strictly on the participants’ practical experiences, motivations, and reflections regarding their contributions to the citizen science project. No sensitive personal data, as defined by Article 9(1) of the GDPR and Swedish law, such as health, political opinions, religious beliefs, racial or ethnic origin, or sexual orientation were collected or processed. The interview questions (see Supplementary File 1: Appendix – Interview Guide) pose no physical or psychological risk, nor do they involve any risk of legal, financial, or social harm to the participants. All interviewed citizen scientists were consenting adults aged 18 or over, meaning that no minors or other vulnerable groups were targeted. The level of risk is therefore classified as negligible or minimal. However, researchers must still comply with

good research practice and data protection legislation. The study strictly adhered to the Declaration of Helsinki and general European research ethics. All participants provided informed consent, participation was entirely voluntary, and all data have been pseudonymized or anonymized to protect participants’ privacy in accordance with the GDPR.

FUNDING INFORMATION

This research was supported by a grant from The Swedish National Heritage Board, RAÄ-2021-2704, *Transcription node Sweden: machine learning and citizen science combined*.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Dick Kasperowski  orcid.org/0000-0003-4539-2055
University of Gothenburg, Sweden

Karl-Magnus Johansson  orcid.org/0000-0002-4734-5060
The Swedish National Archives, Stockholm, Sweden

REFERENCES

- Braun, V.** and **Clarke, V.** (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), pp. 77–101. DOI: <https://doi.org/10.1191/1478088706qp0630a>
- Brinkmann, S.** and **Kvale, S.** (2015) *InterViews: Learning the Craft of Qualitative Research Interviewing*. Los Angeles: SAGE.
- Burell, M.** and **Sjögren, C.** (2018) *Information i verksamhet och arkiv: Regler och standarder med digitalt perspektiv*. Stockholm: Föreningen för Arkiv och Informationsförvaltning.
- Colavizza, G., Blanke, T., Jeurgens, C., and Noordegraaf, J.** (2021) Archives and AI: an overview of current debates and future perspectives, *ACM Journal for Computing and Cultural Heritage* 15:1, pp. 1–15. DOI: <https://doi.org/10.1145/3479010>
- Douglas, J.** (2010) Origins: evolving ideas about the principle of provenance. In: Eastwood, T., and MacNeil, H. *Currents of Archival Thinking*, Santa Barbara: Libraries Unlimited, pp. 23–43. DOI: <https://doi.org/10.5040/9798216963882.0005>
- Douglas, J., Alisauskas, A., Bassett, E., Duranseaude, N., Lee, T., and Mantey, C.** (2022) “These are not just pieces of paper”: acknowledging grief and other emotions in pursuit of person-centered archives, *Archives & Manuscripts* 50:1, pp. 5–29. DOI: <https://doi.org/10.37683/asa.v50.10211>

- Duerinckx, A., Veeckman, C., Verstraelen, K., Singh, N., Van Laer, J., Vaes, M., Vandooren, C., and Duysburgh, P.** (2024) Co-creating artificial intelligence: designing and enhancing democratic AI solutions through citizen science, *Citizen Science: Theory and Practice*, 9(1), pp. 1–15. DOI: <https://doi.org/10.5334/cstp.732>
- Favourate, Y.** (2020) Saturation controversy in qualitative research: complexities and underlying assumptions: a literature review, *Cogent Social Sciences* 6:1, pp. 1–17. DOI: <https://doi.org/10.1080/23311886.2020.1838706>
- Grut, S.** (2016) Lärande genom arkiv: tankar om arkivens outnyttjade potential. In: Hoswar, M., Røsjø, E., Jensen, B., Jensen, C.S.H., Ketola, A., og Östergren, M.L. #arkivdag: relevans, medvirkning, dialog, Oslo: Oslo Met Skriftserien, pp. 75–83.
- Hansen, A.** (2014) The heritage learning framework and the heritage learning outcomes. In: Christidou, D. *Implementing Heritage Learning Outcomes*, Fornvårdaren 37. Östersund: Jamtli Förlag, pp. 7–24.
- Heinisch, B., Oswald, K., Weißpflug, K., Shuttleworth, S., and Holton, H.** (2021) Citizen Humanities. In: Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R., et al. *The Science of Citizen Science*. Cham: Springer, 97–118. DOI: https://doi.org/10.1007/978-3-030-58278-4_6
- Johansson, K.-M.** (2017) Hungerkravallerna 1917: deltagande i fokus i arbetet med ny utställning i Göteborg, *Nordisk Arkivnyt* 2, pp. 80–81.
- Johansson, K.-M.** (2019b) Medskapande, design och film: ny arkivutställning i Göteborg, *Nordisk Arkivnyt* 1, pp. 9–10.
- Johansson, K.-M.** (2021) Film som arkivhandling och som utgångspunkt för arkivutställning. In: Johansson, K.-M. *Filmens Göteborg: spår, erfarenheter och resultat från ett samverkande forskningsprojekt*, *Arkiv i Väst*, nr 28, pp. 59–83.
- Kasperowski, D., and Hillman, T.** (2018) The epistemic culture in an online citizen science project: programs, antiprograms and epistemic subjects, *Social Studies of Science* 48:4, pp. 567–568. DOI: <https://doi.org/10.1177/0306312718778806>
- Kasperowski, D., Johansson, K.-M., and Karsvall, O.** (2024) Temporalities and values in an epistemic culture: citizen humanities, local knowledge, and AI-supported transcription of archives, *Archives & Manuscripts* 51:2, pp. 3–22. DOI: <https://doi.org/10.37683/asa.v51.10937>
- Kasperowski, D., Kullenberg, C., and Rohden, F.** (2020) The Participatory Epistemic Cultures of Citizen Humanities: Bildung and epistemic subjects. In: Pierroux, P., Hetland, P., and Esborg, L. *A History of Participation in Museums and Archives: Traversing Citizen Science and Citizen Humanities*. London & New York: Routledge, pp. 238–242. DOI: <https://doi.org/10.4324/9780429197536-17>
- Kloetzer, L., Lorke, J., Roche, J., Golumbic, Y., Winter, S., Jogeve, A.** (2021) Learning in Citizen Science. In: Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R., et al., *The Science of Citizen Science*. Cham: Springer, pp. 283–307. DOI: https://doi.org/10.1007/978-3-030-58278-4_15
- Kloetzer, L., Scheider, D., Jennett, C., Iacovides, I., Eveleigh, A., Cox, A., L., and Gold, M.** (2013) Learning by volunteer computing, thinking and gaming: what and how are volunteers learning by participating in virtual citizen science. In: Käßlinger, B., and Robak, S. *Changing configurations of adult education in transitional times: conference proceedings*. Berlin: Peter Lang, pp. 73–92.
- Knorr-Cetina, K.** (1999) *Epistemic cultures: how the sciences make knowledge*. Cambridge, Mass: Harvard University Press. DOI: <https://doi.org/10.4159/9780674039681>
- Knorr-Cetina, K.** (2007) Culture in global knowledge societies: knowledge cultures and epistemic Cultures, *Interdisciplinary Science Reviews* 32:4, pp. 361–375. DOI: <https://doi.org/10.1179/030801807X163571>
- Mühlberger, G., Seaward, L., Terras, M., Oliviera, S.A., Bosch, V., Bryan, M., Colutto, S., et al.** (2019) Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study, *Journal of Documentation* 75:5, pp. 954–976.
- Østerlund, C., Crowston, K., Jackson, C.B., Wu, Y., Smith, A.O., and Katsagellos, A.K.** (2024) Supporting human and machine co-learning in citizen science: lessons from gravity spy, *Citizen Science: Theory and Practice*, 9(1): 42, pp. 1–14. DOI: <https://doi.org/10.5334/cstp.738>
- Owens, T.** (2013) Digital cultural heritage and the crowd, *Curator: The Museum Journal*. 56(1), pp 121–130. DOI: <https://doi.org/10.1111/cura.12012>
- Philips, T., Porticella, N., Constan, M., and Bonney, R.** (2018) A Framework for articulating and measuring individual learning outcomes from participation in citizen science, *Citizen Science: Theory and Practice* 3:2, pp. 1–19. DOI: <https://doi.org/10.5334/cstp.126>
- Rafner, J., Dellerman, D., Hjorth, A., and Veraszto, D.** (2021) Deskillling, upskilling, and reskillling: a case for hybrid intelligence, *Morals & Machines*. 1(2), pp. 24–39. DOI: <https://doi.org/10.5771/2747-5174-2021-2-24>
- Ridge, M.** (2013) From tagging to theorizing: deepening engagement with cultural heritage through crowdsourcing, *Curator: The Museum Journal* Vol 56(4), pp 435–450. DOI: <https://doi.org/10.1111/cura.12046>
- Ridge, M.** (2014) *Crowdsourcing our Cultural Heritage*. London: Routledge.
- Ridge, M., Blickham, S., and Ferriter, M.** (2021). *The Collective Wisdom Handbook: Perspectives on Crowdsourcing in Cultural Heritage*. The British Library.
- Roche, J., Bell, L., Galvão, C., Golumbic, Y.N., Kloetzer, L., Knoben, N., Laakso, M., et al.** (2020) Citizen science, education, and learning: challenges and opportunities, *Frontiers in Sociology*, 5, pp 1–10. DOI: <https://doi.org/10.3389/fsoc.2020.613814>

- Rolan, G., Humphries, G., Jeffrey, L., Samaras, E., Antsoupova, T., and Stuart, K.** (2018) Artificial intelligence in the archive, *Archives & Manuscripts* 47:2 pp. 179–203. DOI: <https://doi.org/10.1080/01576895.2018.1502088>
- Sharma, N., Colucci-Gray, L., Lakeman-Fraser, P., Robinson, A., Newman, J., Van der Wal, R., Rueger, S., et al.** (2024) Image recognition as a “dialogicAI partner” within biodiversity citizen science—anempirical investigation, *Citizen Science: Theory and Practice*, 9(1): 35, pp. 1–16. DOI: <https://doi.org/10.5334/cstp.735>
- Shelton, A., Algee, L., Osborn, C., and Owens, T.** (2024) Giving back, learning, relaxing, and having fun: personal motivations and impacts of a virtual volunteer transcription program, *International Journal of Digital Humanities* 6, pp. 143–168. DOI: <https://doi.org/10.1007/s42803-024-00087-0>
- Tarride, S., Faine, T., Boillet, M., Mouchère, H., and Kermorvant, C.** (2023) Handwritten text recognition from crowdsourced annotations, pp. 1–7. <https://arxiv.org/abs/2306.10878>.
- Terras, M.** (2022) Inviting AI into the archives: The reception of handwritten recognition technology into historical manuscript transcription. In: Jaillant, L., *Archives, Access and Artificial Intelligence: Working with Born-Digital and Digitized Archival Collections*. Bielefeld: transcript Verlag, pp. 179–204. DOI: <https://doi.org/10.14361/9783839455845-008>
- Wecker, A.J., Schor, U., Raziel-Kretzmer, V., Elovits, D., Lavee, M., Kuflik, T., and Stökl Ben Ezra, D.** (2020) Opportunities for personalization for crowdsourcing in handwritten text recognition. In: *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization*. UMuAI, Genova, Italy, pp. 373–375. DOI: <https://doi.org/10.1145/3386392.3402436>

TO CITE THIS ARTICLE:

Kasperowski, D and Johansson, K-M. 2026. Participant Learning and Open Archives Through Artificial Intelligence and Citizen Humanities. *Citizen Science: Theory and Practice*, 11(1): 34, pp. 1–13. DOI: <https://doi.org/10.5334/cstp.911>

Submitted: 24 September 2025 **Accepted:** 28 September 2026 **Published:** 27 October 2026

COPYRIGHT:

© 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <https://creativecommons.org/licenses/by/4.0/>.

Citizen Science: Theory and Practice is a peer-reviewed open access journal published by Ubiquity Press.