

Past voices, present tools. Citizen Science and the Hesse-Nassau Dictionary

*Nathalie Mederake, Robert Engsterhold
Marburg University, Research Center Deutscher Sprachatlas
E-mail: mederake@uni-marburg.de, engsterhold@uni-marburg.de*

Abstract

This paper examines the role of Citizen Science in dialect lexicography through the case of the Hesse-Nassau Dictionary (HNWb). It argues that participatory practices have long been integral to dialect documentation, particularly through large-scale questionnaire surveys and the involvement of local speakers as linguistic informants. These historically established forms of collaboration constitute both the material and methodological foundation of the HNWb archive, which comprises thousands of questionnaires and hundreds of thousands of handwritten paper slips.

In light of this context, the paper conceptualises Citizen Science in dialect lexicography not as a methodological innovation introduced by digital media, but as an epistemic reconfiguration of historically embedded participatory practices. Building on this legacy, it analyses how contemporary digital tools allow for a systematic continuation and formalisation of participation under present-day conditions. Two dedicated applications support citizen contributors in the transliteration and annotation of historical dialect materials that cannot be efficiently processed using automated handwritten text recognition.

Through these workflows, contributors move from the role of informants to that of collaborators who actively participate in the archive's transformation, structuring, and enrichment. The paper argues that incorporating Citizen Science into lexicographic workflows enhances the sustainability, accessibility, and societal relevance of long-term dialect dictionary projects. Digital infrastructures thus allow past voices to be reactivated and reinterpreted, strengthening the preservation and contemporary usability of regional linguistic heritage.

Keywords: Citizen Science; Dialect Lexicography; Digital Archives

1 Citizen Involvement in Dialect Documentation

Citizen Science provides a useful conceptual framework for understanding the long-standing involvement of non-academic contributors in dialect lexicography. Rather than being a recent innovation, the systematic inclusion of lay participants has been a defining feature of many large-scale dictionary projects, particularly those concerned with regional and

non-standard varieties. Although general historical dictionaries such as the *Oxford English Dictionary* are often cited as early examples of public participation in lexicography¹, dialect documentation practices represent an even more consistent culturally and structurally embedded form of citizen involvement (cf. Spiekermann 2016).

From the early nineteenth century onwards, dialectological research relied heavily on local speakers (e.g., Schmeller 1821) and large-scale undertakings such as linguistic atlases would have been inconceivable without this distributed network of contributors (cf. Wenker 1889-1923; Gilliéron 1902-1914). Particularly in the context of dialect dictionaries, citizen participation was not merely supportive but constitutive. The empirical basis of these projects depended almost entirely on contributions from non-academic collaborators, who acted as knowledgeable insiders to local speech communities, contributed lexical knowledge, usage examples, and metalinguistic insights through questionnaires, word lists, and correspondence. Their role corresponds closely to what is now described in Citizen Science research as the co-production of knowledge: contributors did not simply deliver raw data but shaped the scope, depth, and regional differentiation of the resulting lexicographic descriptions.

The Hesse-Nassau Dictionary (hereinafter also HNWB) is not merely an example but a paradigmatic condensation of these participatory principles. As one of the most innovative dialect dictionaries of the early twentieth century, it was the first to systematically integrate questionnaire data into individual word entries, thereby establishing the principle of word geography (cf. Berthold 1938; Knoop et al. 1982; Lameli 2014). Information on location thus functions as an indicator of a lemma's geographical scope of validity, foregrounding spatial distribution as an epistemically relevant dimension of lexical meaning. This approach embedded citizen-generated data directly into the analytical structure of the dictionary. Seen from a contemporary perspective, the HNWB archive thus exemplifies an early, large-scale Citizen Science infrastructure *avant la lettre*, in which expert knowledge and local linguistic competence were tightly interwoven.

In today's digital environment, this historical model of participation is being reinterpreted and extended. Citizen contributors increasingly move beyond the role of informants to become active collaborators who participate in the transcription, annotation, interpretation, and contextualisation of dialect data (Dorn et al. 2024; Dücker/Fischer 2025). This shift reflects a broader development in Citizen Science toward more reflexive, sustainable, and mutually beneficial forms of collaboration and highlights the continuing relevance of participatory approaches for dialect lexicography and regional linguistic heritage.

1 As early as 1859 the Philological Society laid the groundwork for the collective enterprise, that would become the *Oxford English Dictionary* (cf. *Proposal for the Publication of a New English Dictionary by The Philological Society*. London: Trübner). Especially Murray (1879) explicitly mobilized the English- and American-speaking public to collect usage evidence; hundreds of volunteers supplied quotation slips that underpinned definitions and preliminary semantic structures.

2 Defining Citizen Science

Against the historical backdrop of sustained public involvement in dialectological and lexicographic research, the concept of Citizen Science provides a differentiated, albeit contested, analytical framework. As Haklay et al. (2021) argue, the field is resistant to a single definition, as criteria often depend on geographical, disciplinary, and political contexts. Generally, however, in the humanities it is understood as the active involvement of non-professional participants in one or more stages of the research process, ranging from data generation to interpretation and dissemination (Pettibone/Ziegler 2016). Crucially, participation is not conceived as an undifferentiated activity but as structured engagement embedded in scientific workflows.

For projects such as the HNWb, Haklay's typology, which distinguishes four levels of Citizen Science according to the depth of citizen involvement (Haklay 2013, 2018), is a useful classification scheme, but comes with limitations. At Level 1 (Crowdsourcing), citizens contribute resources or automatically collected data without being cognitively involved in scientific interpretation. Level 2 (Distributed Intelligence) marks the threshold at which active participation begins: citizens perform intellectually relevant tasks such as observing phenomena, categorising data, or transcribing texts. Level 3 (Participatory Science) involves citizens in shaping research questions and data collection strategies, while Level 4 (Extreme or Collaborative Citizen Science) is characterised by an egalitarian collaboration in which professional researchers and citizen scientists jointly negotiate all stages of the research process.

This model is advantageous for classifying citizen involvement in dialect lexicography, as it allows historical and contemporary practices to be assigned to different levels of participation. Traditional dialect documentation relied heavily on what would today be classified as Level-2-activities: local speakers systematically answered questionnaires, compiled word lists, and provided usage examples, thereby contributing expert linguistic knowledge rooted in everyday practice. In some cases, long-term informants or regional correspondents, many of whom occupied socially and institutionally hybrid positions, such as teachers, acted as knowledgeable insiders and contributed material in a sustained and locally focused manner, thereby shaping the empirical profile of the documented data.² Although dialect informants might superficially resemble Level-1 contributors acting as linguistic 'sensors', the cognitive effort involved in translating lived linguistic experience into structured responses justifies their classification as Level-2 contributors. Rather than constituting participatory research in the sense of Haklay's Level 3, such contributions can be understood as proto-participatory practices whose epistemic relevance remained largely implicit. The decisive

-
- 2 The prominent role of teachers as contributors highlights a hybrid form of participation that challenges contemporary Citizen Science definitions. While teachers were not professional linguists, their involvement was institutionally embedded and partially aligned with their occupational roles. This blurs the often-assumed distinction between 'professional' and 'citizen' participation, which in many modern CS frameworks is implicitly tied to a separation between paid work and voluntary leisure activity. This historical constellation complicates typologies that rely on a strict binary between professional researchers and voluntary citizen contributors (cf. Haklay 2013).

difference from contemporary Citizen Science thus lies not in the principle of participation itself but in its explicit conceptualisation, formalisation, and infrastructural embedding.

Digital infrastructures make historically implicit forms of collaboration explicit by embedding them into transparent, structured, and reproducible workflows. Activities such as transcription, annotation, and categorisation can thus be systematically implemented as Level-2 (Distributed Intelligence) practices, while at the same time creating conditions for more reflexive and sustained forms of engagement (Dücker/Fischer 2025). Contributors are no longer confined to isolated acts of data provision but gain insight into the epistemic context, decision-making processes, and research goals that shape the archive.

Understanding Citizen Science as a graded and historically embedded practice therefore provides a conceptual bridge between analogue dialect documentation and contemporary digital lexicography. The HNWB illustrates that Level-2 activities in the humanities are not an “inferior” preliminary stage but constitute the epistemic backbone of long-term lexicographic research. Haklay’s typology is thus not employed as a normative scale of participatory “success”, but as a descriptive model for analysing the division of labour between academic research, public participation, and digital infrastructures in long-term dialect dictionary projects.

From this perspective, digital Citizen Science in dialect lexicography represents both a continuation and a reconfiguration of earlier participatory practices. Its point of departure is not an abstract research design but the material legacy of historical collaboration. In the case of the Hesse-Nassau Dictionary, the extensive archive of paper slips questionnaires embodies both the outcome of earlier citizen involvement and the foundation for renewed participation under digital conditions.

3 Starting points for Citizen Science: The Archive

As a material legacy of early participatory research, the archive of the Hesse-Nassau Dictionary forms the natural starting point for contemporary Citizen Science initiatives. The archive is not only a repository of data but also a repository of participatory practices. It is fed by two distinct but closely interrelated data classes.

- (1) The **slip archive** constitutes its core component. It comprises more than 300,000 paper slips on which meanings and usage information were excerpted, classified, and lemmatized. These data stem from unsolicited public submissions that were actively encouraged through postcard surveys as well as calls published in journals and calendars between 1914 and 1926 (Sitzungsberichte 1920: 132). A single slip may contain different types of linguistic information; in broad terms, these can be distinguished between meaning descriptions, phonetic notes, contextualised word attestations (sentence examples), form attestations, and morphological or syntactic information.
- (2) The second data class consists of a **questionnaire corpus**, made up of responses provided by local contributors (i.e. Gewährsperson), predominantly schoolteachers. Its core comprises 79 questionnaire rounds in parts of the dictionary area, amounting in total to approximately 19,000 completed questionnaire pages. The

responses cover a wide thematic spectrum, ranging from family life and traditional customs to flora and fauna and local knowledge, and frequently include structured elicitation tasks, such as naming the parts of illustrated objects (e.g. a loom) or listing regional expressions for predefined concepts (Werth et al. 2021, 203). While the questionnaire material represents a substantial and systematically collected body of data, it was only rudimentarily excerpted by the dictionary editors, leaving much of its analytical potential unrealised.

A central motivation behind the questionnaire surveys – and indeed behind the Hesse-Nassau Dictionary as a whole – was the transfer of the dialect-geographical method developed at the Deutscher Sprachatlas to lexicography (the so-called “Marburg School”; cf. Knoop et al. 1982; Lameli 2014). In this tradition, linguistic forms are not primarily conceived as abstract system elements but as spatially distributed phenomena whose geographical patterning constitutes an essential part of their linguistic description. Consequently, information on the place of attestation does not merely function as a source reference or documentary metadata. Rather, it serves as an indicator of the geographical scope of validity of a lexical item and thus becomes an integral component of its semantic and pragmatic characterisation. Lexicographic description, in this sense, is inseparable from the mapping of linguistic variation.³

Together, these materials were gathered with the involvement of more than 1,000 local contributors and amount to roughly 1.5 million dialectal references. The richness of this material stands in marked contrast to its limited integration into the printed dictionary since large parts of the archive remain accessible only through indirect references. The complete digital preservation of the archive therefore not only secures a unique historical resource but also exposes its latent analytical potential. Unlocking this potential requires a systematic transformation of heterogeneous, analogue materials into structured, machine-readable data – a process that cannot be achieved by editorial expertise alone but calls for renewed forms of distributed, participatory engagement. A further obstacle to systematic evaluation lies in the material and graphic characteristics of the sources themselves. Most paper slips and questionnaire answers are written in a range of historical handwriting systems, including cursive scripts such as Kurrent and Sütterlin, which are no longer in everyday use. Owing to their graphic variability and palaeographic complexity, these scripts cannot yet be transcribed efficiently using current Handwritten Text Recognition (HTR) technologies such as Transkribus (<https://www.transkribus.org>) or eScriptorium (<https://github.com/UB-Mannheim/escriptorium>). As a result, large parts of the archive continue to depend on manual transliteration, making renewed forms of distributed, human-based participation not merely desirable but methodologically indispensable.

3 According to the principle of “word geography” proposed by Berthold (1938), the geographical distribution of the keywords described is illustrated with the aid of maps. As the first dictionary project of its kind, the HNWb attempted to consistently apply this principle from the outset (an example that many parallel projects have followed), and its word articles and maps provide a wealth of ideas and results for further research in the field of geographical comparative analysis.

4 From Archival Data to Digital Participation

Against this background, the digital Citizen Science infrastructure of the Hesse-Nassau Dictionary is designed to operationalise the archive's latent analytical potential by transforming historically grown materials into structured, machine-readable data through renewed forms of public participation. Two complementary applications support this process, addressing different core components of the archive and enabling distinct but interrelated modes of participation. In addition to their epistemic relevance, the applications contribute to the preservation of the archival material. Many original questionnaires are in a critical physical condition due to the high wood content of the paper and advanced ageing. The digital workflow reduces the need for physical handling and thus helps to safeguard the material for future research. The two applications presented here are implemented as high-fidelity prototypes. Their functional scope corresponds to that of intended production systems, and parts of the infrastructure are publicly accessible (<https://uni-marburg.de/bN-lc7L>). At the same time, the tools are currently in a pilot phase: they have so far been tested by a limited number of users and do not yet provide access to the entire digitised archive.

From a technical perspective, both applications are embedded in a shared digital infrastructure. Built as a progressive web app (https://developer.mozilla.org/en-US/docs/Web/Progressive_web_apps, <https://quasar.dev>) with Service Endpoint based on FastAPI (<https://fastapi.tiangolo.com>) and backed by a PostgreSQL (<https://www.postgresql.org>) database, the system transforms the heterogeneous archival holdings into structured, analysable data. In addition, the integration of IIIF-compliant image services (<https://iiif.io>) ensures that digitised cultural assets – such as questionnaire pages, handwritten slips, maps, or other archival objects – can be delivered and displayed in a standardised manner via web interfaces. Users can interact with the applications of the website without requiring personal registration. A token-based system ensures that individual contributions can be persistently associated with anonymous contributor profiles, enabling continuity and transparency. This architectural design not only supports the current Citizen Science workflows but also provides a sustainable foundation for future extensions, interoperability, and reuse of the data.

4.1 Document Editor

While questionnaire surveys formed the empirical backbone of early dialect documentation, the smallest and most analytically condensed units of the HNWB archive are the individual paper slips. The document editor is dedicated to their systematic digitisation, structuring, and enrichment.

4.1.1 Concept: From Card Index to Database

The conceptual aim of the document editor is the in-depth annotation of the slip archive. The HNWB is based on around 300,000 of often handwritten slips, each documenting a lexical item together with information on meaning, usage in context, and place of attestation. In their original analogue form, these slips were organised in a physical card index that em-

bodied decades of editorial practice but remained inaccessible to systematic computational analysis.

In contrast to the questionnaire material, which is primarily organised by locality, the slip archive follows a word-oriented logic. The document editor therefore is designed to enable contributors to work on clusters of slips associated with individual lemmas. Its overarching goal is to transform the analogue card index into a digital, searchable corpus that can serve as the empirical foundation for both printed and digital dictionary entries.

4.1.2 Structure and Functionality

The document editor is implemented as a specialised interface optimised for the efficient processing of small-scale data units. At its centre, a digital image window displays the scan of the original slip. As many slips are no larger than a small note card, the viewer is designed for rapid readability and focused inspection.

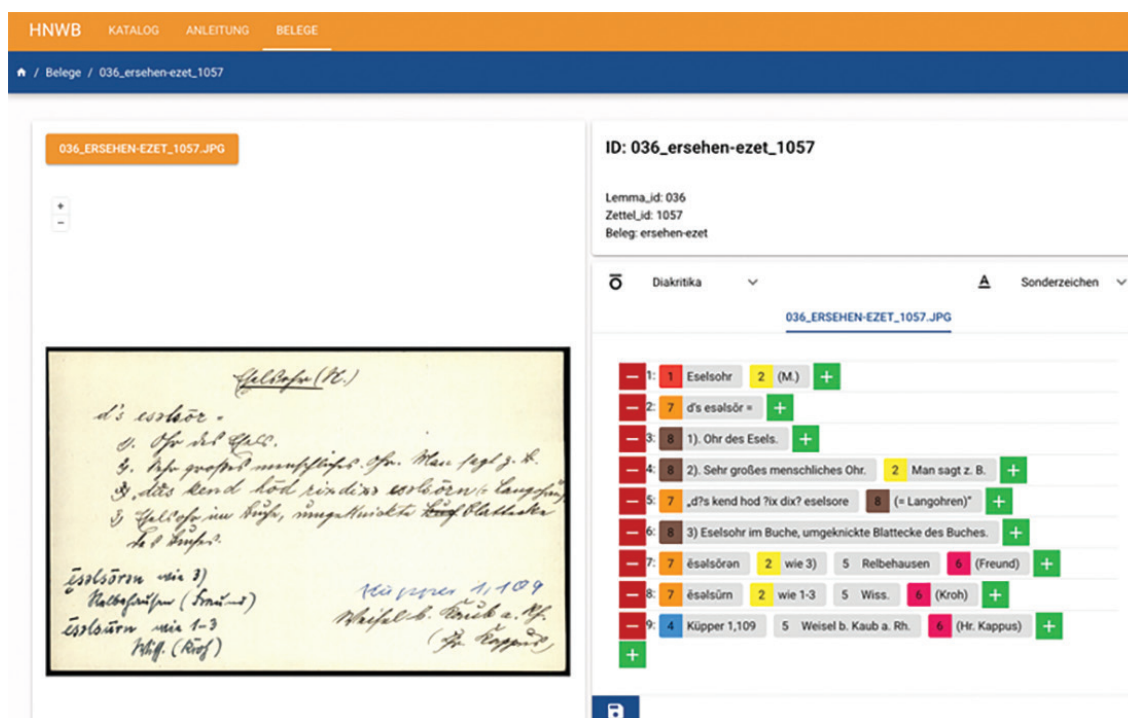


Figure 1: User interface of the document editor. The image shows the digital capture of a handwritten note (ID: 1057) on the lemma “Eselsohr” (dog-ear). On the left is the original document in cursive script, and on the right is the corresponding line-by-line transcription and annotation in the database mask.

Adjacent to the image, the editor interface allows contributors to transliterate the handwritten content and to assign structured metadata using colour-coded and numerical tags. These tags capture information such as lemma assignment, meaning, regional spelling,

place of attestation, and local informant. Users shall be enabled to work through entire sets of slips associated with a given lemma in a continuous workflow, closely resembling – but digitally enhancing – the traditional practice of sorting and annotating paper slips. In its current implementation, the document editor operates on a selected subset of the digitised slip archive. While the core functionalities reflect the design of the intended end product, access to the complete corpus will be enabled in later development stages.

4.1.3 Involvement of Citizen Scientists

The granularity of the document editor allows the processing of single paper slips as bite-sized tasks. This lowers the threshold for participation and allows volunteers to contribute flexibly, even in short intervals. At the same time, the platform supports a degree of thematic and regional specialisation. Citizen scientists may deliberately select lemmas that are of interest and thus contribute situated linguistic knowledge to the annotation process. Local expertise is particularly valuable when interpreting usage notes, regional spellings, or semantically opaque expressions. Quality assurance is supported through review mechanisms that allow entries to be checked by other contributors or by expert editors. Through this form of collaborative philology, citizen scientists participate directly in the transformation of the “raw data” of linguistic history into structured evidence. Their work provides the indispensable groundwork for scholarly analyses of lexical variation, semantic change, and regional distribution.

To ensure the scholarly reliability and analytical usability of citizen-generated data, the HNWB infrastructure implements a multi-layered quality assurance framework that is embedded directly into the digital workflows. At the level of data entry, structural constraints within the interfaces – such as predefined tags, controlled vocabularies, and guided input fields – reduce formatting inconsistencies and limit interpretative variance. In addition, a collaborative validation model is employed: for complex or palaeographically ambiguous documents, multiple independent transliterations or annotations can be created for the same source, allowing discrepancies to be identified through comparative inspection. Cases of significant ambiguity or conflict could also be systematically flagged and subjected to expert review by the editorial team. Crucially, all digital entries remain persistently linked to their original manuscript sources via IIIF-compliant image services, ensuring transparent provenance and verifiability at every stage of the lexicographic process. This combination of structured input, collaborative validation, and expert oversight balances the openness of Citizen Science with the methodological standards of digital lexicography.

4.2 Questionnaire Application

While the document editor operates on the smallest analytical units of the archive, the questionnaire application addresses the large-scale survey material that forms the historical backbone of the HNWB.

4.2.1 Concept and Objectives

The overarching goal of the questionnaire application is the digitisation and transliteration of the historical questionnaire corpus of the Hesse-Nassau Dictionary. At its core lies a collection of approximately 6,500 handwritten questionnaires completed between 1914 and 1926 by voluntary local contributors. These questionnaires represent an extensive and systematically collected source of dialect data in the archive.

From a scientific perspective, the application responds to the fact that, during the original compilation of the dictionary, only selected parts of the questionnaire material were manually excerpted and transferred onto paper slips. Large portions of the original responses were therefore never fully evaluated. By converting the handwritten content into machine-readable text through manual transliteration, the application aims to unlock these previously unexplored linguistic resources and make them available for integration into modern linguistic databases.

4.2.2 Structure and Functionality

Similar to the document editor, the questionnaire application is implemented as a modular web-based environment designed to support focused transliteration work. The entry point is a questionnaire catalogue page that provides an overview of the available material. On the left-hand side, users see a list of questionnaire pages available for processing; on the right-hand side, the corresponding localities are displayed as points on a map. Visual indicators signal the processing status: an orange outline around a location point or an orange background behind a page icon indicates that a transliteration already exists. Users may nevertheless initiate additional transliterations at any time.

Selecting a page icon opens an individual questionnaire page that is uniquely associated with a specific locality and questionnaire series. The working view is again organised as a split-screen interface. On the left-hand side, a digital viewer displays a high-resolution scan of the original handwritten questionnaire. On the right-hand side, the editor interface provides input fields for the letter-by-letter transliteration of the dialect data, supplemented by auxiliary tools such as special-character palettes for phonetic symbols.

Contributors may transliterate an entire questionnaire page or focus on individual questions, for example a prompt eliciting local expressions for *Wacholder* (*Juniperus communis*). A token-based session mechanism stores the progress persistently without requiring a formal registration process, thereby combining usability with data protection. Although the application is technically capable of handling large-scale questionnaire material, it currently provides access only to a defined portion of the digitised corpus. This phased approach allows for iterative testing and refinement of the workflow under real research conditions.

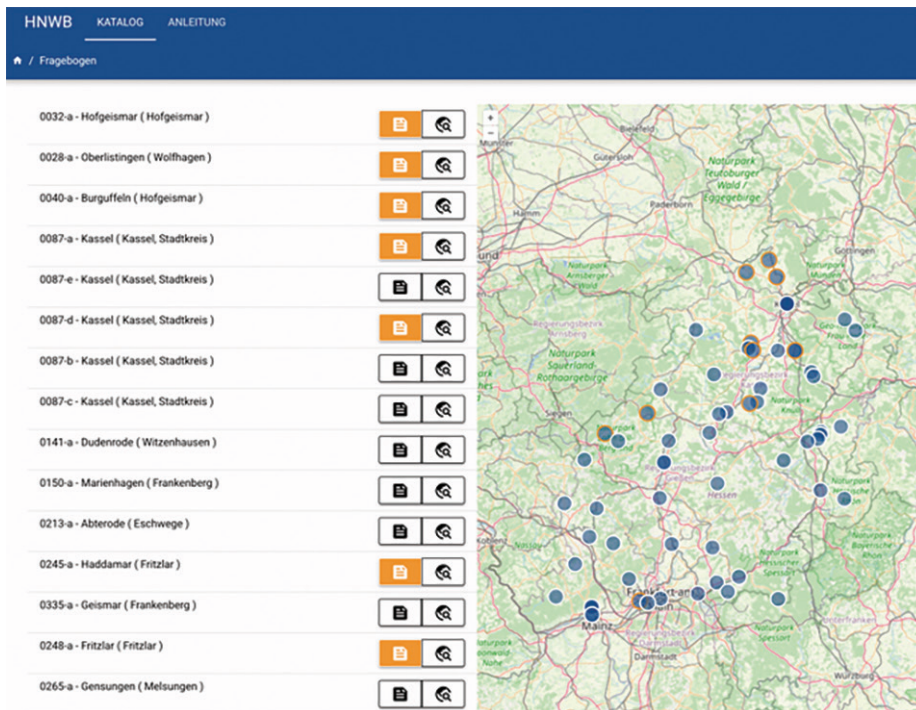


Figure 2: Questionnaire catalog. The interactive overview map and document list enable targeted access to dialect data via a list or direct selection of locations in the Hesse-Nassau region.

4.2.3 Involvement of Citizen Scientists

The questionnaire application represents a classic Citizen Science setting that depends on broad public participation. Its design deliberately lowers barriers to entry in order to engage interested laypersons, dialect speakers, and regional historians. The anonymous token system allows contributors to begin working immediately and to resume their contributions at a later point.

At the methodological level, the platform implements transliterations through distributed intelligence. Given the volume of handwritten material and the palaeographic challenges it presents, systematic processing would not be feasible without the involvement of citizen scientists. Their contributions enable the deciphering and digitisation of handwriting that cannot be reliably processed using automated methods.

At the same time, the platform facilitates knowledge transfer. Contributors engage actively with the linguistic history of their region, while the application provides detailed guidelines and instructional materials to ensure data quality suitable for scholarly analysis. In this way, the questionnaire application functions as a digital bridge between an analogue archival legacy and contemporary linguistic research, extending historical participatory practices into the digital age.

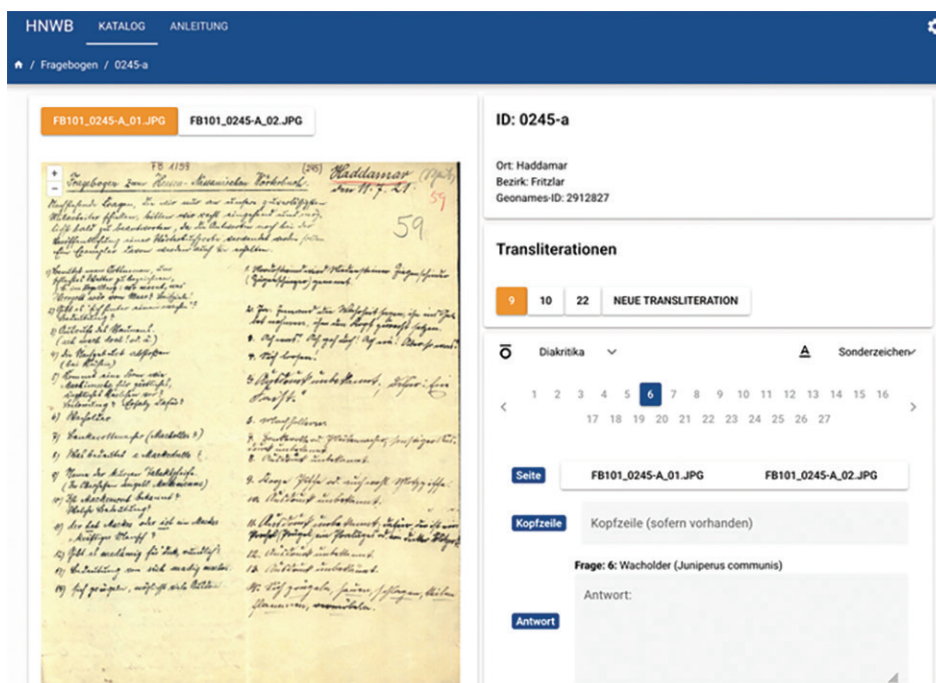


Figure 3: Questionnaire App. The image shows the work interface for transliterating questionnaire 0245-a from Haddamar (Fritzlar district). On the left is the original document, while on the right is the input mask for the transliteration of the answers – here, for example, for question 6 (“juniper”).

5 From Informants to Collaborators

In earlier stages of dialect lexicography, the contributions of local respondents were indispensable to the empirical foundation of dictionary projects, yet largely confined to the phase of data generation. While questionnaires, correspondence, and unsolicited submissions supplied the material basis of dialect lexicographic work, decisions concerning selection, structuring, interpretation, and presentation remained firmly within the domain of academic expertise. Citizen participation was thus constitutive but epistemically backgrounded, operating as an implicit prerequisite or proto-participatory rather than as an explicitly theorised component of knowledge production.

Within the framework of digital Citizen Science, this historically established division of labour is being reconfigured. As demonstrated by the applications presented above, citizen contributors are no longer restricted to providing linguistic material but participate directly in the transformation, structuring, and enrichment of historical sources. Through activities such as the transliteration of handwritten questionnaires and the annotation of archival slips, they intervene at stages that decisively shape the form, granularity, and analytical usability of the data. In this sense, they function as collaborators embedded within clearly defined lexicographic workflows.

Crucially, this shift does not imply a dissolution of established epistemic roles, nor does it amount to participatory research in the strong sense of Haklay's highest levels. At the same time, it invites a reconsideration of where citizen contributions are situated within the lexicographic process as traditionally described in lexicographic theory (Klosa-Kückelhaus/Tiberius 2024). Rather, it reflects a reassessment of tasks that were traditionally classified as preparatory or technical. In the framework of the lexicographic process, such tasks form integral stages linking data collection, data modelling, and the construction of dictionary entries; they are not auxiliary to lexicographic work, but constitutive of how lexical knowledge is selected, structured, and ultimately represented. Decisions taken during transliteration, annotation, and categorisation – such as the interpretation of ambiguous handwriting, the identification of regional variants, or the assignment of annotations – are epistemically consequential precisely because they intervene at these processual interfaces. They pre-structure the organisation of the data and condition the kinds of linguistic generalisations that can subsequently be drawn. Digital Citizen Science thus renders visible a layer of interpretative labour that has always been present in lexicographic practice but rarely acknowledged as such.

From this perspective, Haklay's typology can be productively aligned with the dialect-geographical tradition of the Marburg School. The spatial principle underlying the Hesse-Nassau Dictionary – according to which geographical distribution constitutes a core dimension of lexical meaning – was only realisable through the distributed intelligence of its contributors. The historical reliance on locally embedded actors exemplifies a form of situated expertise: Rather than representing a merely logistical solution, this hybrid constellation constituted an institutionalised epistemic arrangement that prefigures contemporary models of collaborative knowledge production.

The prototype status of the digital applications plays a central methodological role in this re-configuration. As high-fidelity yet experimental systems, they allow collaborative workflows to be observed, evaluated, and iteratively refined. Patterns of participation, interpretative strategies, and quality-related challenges become empirically accessible before large-scale implementation. In this way, prototyping functions not simply as a technical development stage, but as an epistemic instrument for reflecting on the conditions and consequences of Citizen Science in lexicographic research.

Taken together, the transition from informants to collaborators does not mark a rupture with earlier dialectological practices (Fischer 2024). Rather, it makes the epistemic significance of participatory labour that has long underpinned dialect lexicography explicit. By embedding this labour into transparent, structured, and traceable digital workflows, contemporary Citizen Science re-articulates historical practices under new media conditions and establishes them as a reflexive component of lexicographic methodology (cf. Dorn et. al 2024).

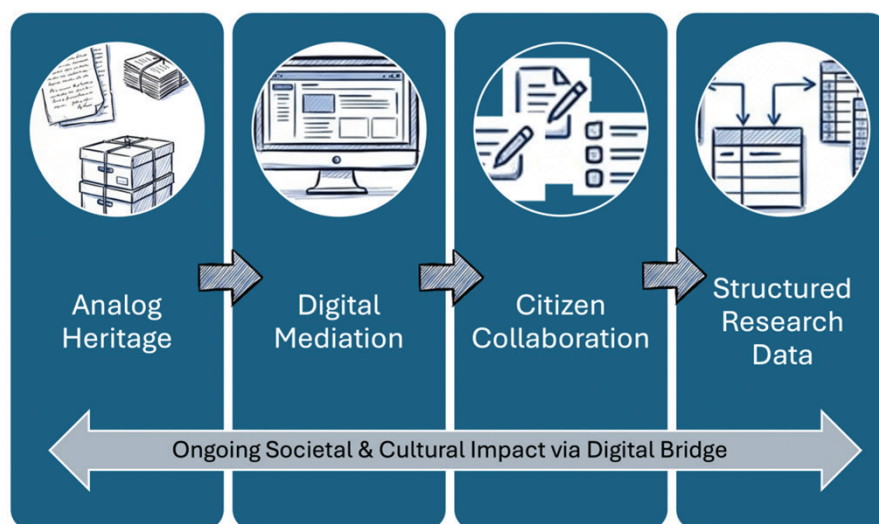


Figure 5: The Digital Bridge model. The diagram illustrates the transformative workflow from analog archival heritage to structured research data, mediated by digital infrastructures and powered by the epistemic collaboration of citizen scientists to ensure ongoing societal impact.

6 Conclusion and outlook

Building on the epistemic reconfiguration outlined in Section 5, this paper has argued that Citizen Science in dialect lexicography is best understood not as an external innovation, but as a reflexive framework for analysing and reorganising long-standing forms of cooperative knowledge production. The case of the Hesse-Nassau Dictionary shows that the participatory logic of dialect documentation predates contemporary Citizen Science discourse by more than a century. What has changed is the media, infrastructures, and conceptual tools through which participation is made explicit, formalised, and sustained.

From a meta-theoretical perspective, the study contributes to current discussions of Citizen Science in the humanities by foregrounding the role of archives, long-term projects, and historically layered data. The HNWB exemplifies a form of distributed and situated expertise that corresponds closely to what is classified as Level-2 participation according to Haklay (2013, 2018), while at the same time challenging normative hierarchies that equate deeper participation exclusively with co-design or full collaboration. In the context of lexicography, distributed intelligence does not represent a preliminary or subordinate mode of engagement, but constitutes the epistemic backbone of the enterprise.

Methodologically, the two digital prototypes presented here – the Document Editor and the Questionnaire Application – demonstrate how participatory workflows can be operationalised within a robust digital infrastructure. These tools function not merely as technical interfaces, but as epistemic environments in which historical sources are transformed into structured, analysable data. By integrating quality assurance, provenance tracking, and collaborative validation directly into the workflows, the infrastructure reconciles the openness

of Citizen Science with the methodological standards of digital lexicography. At the same time, the prototype-based approach enables a reflective mode of implementation in which participatory practices themselves become objects of empirical observation and methodological adjustment.

Looking ahead, the long-term sustainability of dialect lexicography will depend on the successful integration of such participatory infrastructures. This includes not only the maintenance of technical systems, but also the cultivation of enduring relationships with citizen contributors and regional communities. Beyond their immediate scientific value, digital Citizen Science initiatives foster shared ownership of linguistic heritage and create opportunities for cultural rediscovery that extend well beyond academic contexts.

Ultimately, the transformation of the Hesse-Nassau Dictionary's archive into a digital, participatory infrastructure underscores a central insight of this study: the "citizen" has always been at the heart of dialectology. By making participatory labour explicit and epistemically accountable, digital tools do not replace historical practices, but enable past linguistic knowledge to be reinterpreted, recontextualised, and integrated into contemporary research. In this sense, Citizen Science provides not only a practical framework for data processing, but a conceptual lens through which the continuity, adaptability, and societal embeddedness of dialect lexicography can be understood anew.

References

- Berthold, L. (1938). Die Wortkarte im Dienste der Bedeutungslehre. In *Zeitschrift für Mundartforschung* 14, Stuttgart: Franz Steiner, pp. 101–106.
- Dorn, A., Stocker, R. & Stöckle, P. (2024). Old dialect words through the ages – the ABCs of dialect project. In *ARPHA Proceedings* 6, pp. 213–216. Accessed at: <https://doi.org/10.3897/ap.e126110> [19/01/2026].
- Dücker, L. & Fischer, H. (2025). Sprachvariation dokumentieren, transkribieren, reflektieren – Citizen Science in der Regionalsprachenforschung. In E. Wyss, et al. *Gruß & Kuss – Briefe digital. Bürger*innen erhalten Liebesbriefe. Studien zur Liebesbriefforschung und zu Citizen Science in den Geistes- und Kulturwissenschaften. Digital Philology | Working Papers in Digital Philology* 04/2025. Darmstadt: TUPrints. Accessed at <https://doi.org/10.26083/tuprints-00030732> [19/01/2026].
- Fischer, H. (2024): Dialektologie als Citizen Science. Perspektiven und Tools für eine bürger-nahe Wissenschaft. In Ch. Ferstl, P. Kaspar, L. Zehetner (eds.): *Dialekt unterwegs. Varietäten im Zeichen von Globalisierung und Migration*. Tirschenreuth: Johann-Andreas-Schmeller-Gesellschaft.
- Gilliéron, J. & Edmont, E. (1902–1920). *Atlas linguistique de la France*. Paris: Champion.
- Haklay, M., Dörler, D., Heigl, F., Manzoni, M., Hecker, S. & Vohland, K. (2021). What Is Citizen Science? The Challenges of Definition. In K. Vohland, et al. *The Science of Citizen Science*. Cham: Springer, pp. 13–33. Accessed at: https://doi.org/10.1007/978-3-030-58278-4_2 [19/01/2026].

- Haklay, M. (2013). Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation. In D. Sui, S. Elwood, M. Goodchild (eds.) *Crowdsourcing Geographic Knowledge*. Dordrecht: Springer. Accessed at: https://doi.org/10.1007/978-94-007-4587-2_7 [19/01/2026].
- Haklay, M. (2018). Participatory citizen science. In S. Hecker, et al. (eds.) *Citizen Science: Innovation in Open Science, Society and Policy*. London: UCL Press, pp. 52-62. Accessed at: <https://doi.org/10.14324/111.9781787352339> [19/01/2026].
- Klosa-Kückelhaus, A. & Tiberius, C. (2024). 3 The Lexicographic Process. In A. Klosa-Kückelhaus (ed.) *Internet Lexicography: An Introduction*. Berlin, Boston: De Gruyter, pp. 61-96. Accessed at: <https://doi.org/10.1515/9783111233758-004> [19/01/2026].
- Knoop, U., Putschke, W. & Wiegand, H. E. (1982). Die Marburger Schule: Entstehung und frühe Entwicklung der Dialektgeographie. In W. Besch, U. Knoop, W. Putschke, H. E. Wiegand (eds.) *Dialektologie. Ein internationales Handbuch zeitgenössischer Forschung*. Bd. 1. Berlin: de Gruyter, pp. 38-92.
- Lameli, A. (2014): *Schriften zum Sprachatlas des Deutschen Reichs*. Bd. 3. Erläuterungen und Erschließungsmittel zu Georg Wenkers Schriften. Hildesheim: Olms.
- Murray, J. A. H. (1879). *An appeal to the English-speaking and English-reading public to read books and make extracts for the Philological Society's new English Dictionary*. Oxford: Clarendon Press.
- Pettibone, L. & Ziegler, D. (2016). Citizen Science: Bürgerforschung in den Geistes- und Kulturwissenschaften. In K. Oswald, R. Smolarski (eds.) *Bürger Künste Wissenschaft: Citizen Science in Kultur und Geisteswissenschaften*. Gutenberg: Computus Druck Satz & Verlag, pp. 57-69.
- Proposal for the Publication of a New English Dictionary by The Philological Society* (1859). London: Trübner.
- Schmeller, J. A. (1821). *Die Mundarten Bayerns grammatisch dargestellt*. München: Thiene-mann.
- Sitzungsberichte der Preussischen Akademie der Wissenschaften. Philosophisch-historische Klasse* (1922-1938). Berlin: Verlag der Akademie der Wissenschaften.
- Spiekermann, H. H. (2016). 73. Kulturwissenschaftliche Orientierung in Dialektologie / Sprachgeographie. In L. Jäger, et al. (eds.) *Sprache - Kultur - Kommunikation / Language - Culture - Communication: Ein internationales Handbuch zu Linguistik als Kulturwissenschaft*. Berlin, Boston: De Gruyter Mouton, pp. 701-707. Accessed at: <https://doi.org/10.1515/9783110224504-074> [19/01/2026].
- Wenker, G. (1889-1923). Sprachatlas des Deutschen Reichs. Marburg: Forschungszentrum Deutscher Sprachatlas. Published as Digitaler Wenker-Atlas (DiWA) at <https://regional-sprache.de/atlanten-und-karten.aspx> [19/01/2026].
- Werth, A., Vielsmeier, B. & Aumann, St. (2021). Hessen-Nassauisches Wörterbuch (HNWb). In A. N. Lenz, Ph. Stöckle (eds.). *Germanistische Dialektlexikografie zu Be-*

ginn des 21. Jahrhunderts. Stuttgart: Steiner, pp. 201-221. Accessed at: <https://doi.org/10.25162/9783515129206> [19/01/2026].

Acknowledgements

Thanks are due to Tinatin Margalitadze, organiser of the II International Conference on *Lexicography in the XXI Century*, for insightful feedback and constructive discussion. We also would like to thank Selina Berus for her careful reading of the manuscript; any remaining errors are our own.