

Digitization as an opportunity for collaboration: digitizing personal correspondence from World War II at the intersection of history, archival science, and the digital humanities

van Lange, Milan Mikolaj

m.van.lange@niod.knaw.nl
 NIOD Institute for War, Holocaust, and Genocide Studies,
 Amsterdam

van Nispen, Annelies

a.van.nispen@niod.knaw.nl
 NIOD Institute for War, Holocaust, and Genocide Studies,
 Amsterdam

Keijzer, Carlijn

c.keijzer@niod.knaw.nl
 NIOD Institute for War, Holocaust, and Genocide Studies,
 Amsterdam

The project ‘First-Hand Accounts of War: War letters (1935-1950) from NIOD digitized’ (2020-2023) is concerned with the digitization and so-called ‘datafication’ of historical – mostly handwritten – wartime letters. The original paper documents are part of the archival collection of the NIOD Institute for War, Holocaust, and Genocide Studies in Amsterdam, The Netherlands. The collection known as ‘Collection 247: Correspondence’ contains personal letters from the period before, during, and after World War II and the War of Independence in Indonesia (1945-1950). The letters give a first-hand account of experiences of a wide range of different people: soldiers at the front, Jewish children in hiding, or people in forced labour camps – to name but a few (Van Lange et al. 2022; Keijzer 2023; Keijzer et al. 2023). ‘First-Hand Accounts of War’ preserves and scans the original documents, but also aims to take digitization a step further.

All the scanned letters are transcribed to transform them into machine-readable texts (e.g., plain text (.txt) and Extensible Markup Language (.xml)). These transcriptions are partly made by a group of dedicated *citizen scientists* affiliated to the NIOD, and partly by the project team members (Romein et al. 2022; Bouman 2023). The resulting ‘Ground Truth’ transcriptions are then fed to the computer to train models for Handwritten Text Recognition (HTR) using READ-COOP’s Transkribus software environment (Liakos et al. 2019; Muehlberger et al. 2019; Hodel et al. 2021). These models are already made available to the public (Van Lange et al. 2023). They were also used within the project to digitize the bulk of the collections by automated transcription.

In addition, an enriched machine-readable digital dataset was created and will be made available online (e.g., through the websites Oorlogsbronnen.nl (general public), Archieven.nl (historical researchers, professionals), and Zenodo (digital humanists, researchers, developers, etc.)). This ensures that the collection is not only better preserved for future generations, but also generates new possibilities for digital humanities or digital history research. In addition to the more tangible deliverables, ‘First-Hand Accounts of War’ aimed from the start to bring together experiences, perspectives, and contributions of information specialists, archivists, and (digital) history scholars. The project is as much about the envisioned end products, as about learning from and reflecting on the collaborative process along the way (Van Lange et al. 2022). This relates to the multidisciplinary project team itself, as to the collaborations with researchers, archivists, students, and colleagues involved in similar projects (e.g., Taskinen et al. 2021; Janz / Verduyze 2023).

This poster shows how multidisciplinary conversations (both in person and in the blogosphere) and collaborative endeavours in ‘First-Hand Accounts of War’ generated reflexive insights on processes of digitization and the transformations the archival collection underwent. This includes, amongst other things, the choices and decisions made in creating a digital representation of a physical historical archive (Keijzer / Van Nispen 2021; Schijvenaars 2022; Bouman 2023; Van Lange 2023). The poster invites, for example, to (re)consider how the generation of transcriptions and metadata shapes the knowledge one can retrieve from digitized historical data (Kennedy 2017; Tames 2020) or the reception of such data (Rawson 2018). Take, for example, the descriptions in (digitized) archival collections: The concepts, labels, or categories used are not only functional, they also have an epistemological function and meaning. In addition, they ‘(...) [rely] on language that can change the meaning of historical materials; it is created to suit particular audiences and needs (...)’ (Rawson 2018). Terms used in describing sources are often the first and foremost framework of the historical collection with which researchers engage – whether in a ‘paper archive’, or engaging with the sources in a digital context. This is important because archives are (and have been) always about power, as they both reflect and constitute power relations (Cook / Schwartz 2002; Rawson 2018; Zaagsma 2022).

To conclude, the poster displays how the divergent professional and academic backgrounds of the different team members in ‘First-Hand Accounts of War’ are reflected in the resulting project outcomes and deliverables by briefly addressing also the following topics: 1) legal issues (e.g., related to the EU’s GDPR), 2) the (re)use of Ground Truth data and HTR-models, and 3) the acknowledgement of contributions of partners, project members, and *citizen scientists* (Romein et al. 2022).

Bibliography

Bouman, Muriël (2023): “‘Is God a person?’ The questions and choices that arise when transcribing wartime letters”, in: *NIOD Blog* <https://www.niod.nl/en/blog/god-person-questions-and-choices-arise-when-transcribing-wartime-letters> [21.04.2023].

Cook, Terry / Schwartz, Joan. M. (2002): “Archives, Records, and Power: From (Postmodern) Theory to (Archival) Performance”, in: *Archival Science* 2: 171-185.

Hodel, Tobias et al. (2021): “General Models for Handwritten Text Recognition: Feasibility and State-of-the Art. German Kurrent as an Example”, in: *Journal of Open Humanities Data* 7. DOI: <https://doi.org/10.5334/johd.46>.

Janz, Nina / Vercruyse, Sarah Maya (2023): “WARLUX - Soldiers and their communities in WWII: The impact and legacy of war experiences in Luxembourg”. <https://www.c2dh.uni.lu/projects/warlux-soldiers-and-their-communities-wwii-impact-and-legacy-war-experiences-luxembourg> [08.09.2022].

Keijzer, Carlijn (2023): “Hidden treasure and papers strewn with cigar ash: on the archival documents that find their way to NIOD”, in: *NIOD Blog* <https://www.niod.nl/en/blog/hidden-treasure-and-papers-strewn-cigar-ash-archival-documents-find-their-way-niod> [19.04.2023].

Keijzer, Carlijn / Lange, Milan van / Nispen, Annelies van (2023): “Introduction: Wartime letters and the “Postbus NIOD” campaign week”, in: *NIOD Blog* <https://www.niod.nl/en/blog/introduction-wartime-letters-and-postbus-niod-campaign-week> [21.04.2023].

Keijzer, Carlijn / Nispen, Annelies van (2021): “Over digitalisering & keuzes”, in: *NIOD Blog* <https://niodbibliotheek.blogspot.com/2021/03/oorlog-uit-eerste-hand-4.html> [10.06.2022].

Kennedy, Kara (2017): “A Long-Belated Welcome: Accepting Digital Humanities Methods into Non-DH Classrooms”, in: *Digital Humanities Quarterly (DHQ)* 11, 3.

Lange, Milan van et al. (2023): “NIOD_WarLet_1945-1950_NoBasemodel“. Amsterdam and Innsbruck: READ COOP SCE, NIOD. https://readcoop.eu/model/niod_warlet_1945-1950_nobasemodel/ [21.04.2023].

Lange, Milan van (2023): “Written, cherished, scanned: the many lives of a wartime letter“, in: *NIOD Blog* <https://www.niod.nl/en/blog/written-cherished-scanned-many-lives-wartime-letter> [21.04.2023].

Lange, Milan van / Nispen, Annelies van / Keijzer, Carlijn (2022): “Handgeschreven “egodocumenten” automatisch transcriberen: Oorlog uit Eerste Hand: Oorlogsbrieven van het NIOD (1935-1950) digitaal”, in: *E-DATA & RESEARCH*, 17, 1: 8.

Liakos, Filotas et al. (2019): “Digitization Initiatives For Preservation & Access#: The Transkribus Project As a Stepping Stone For a New Archival Era”. <https://openaccess.leidenuniv.nl/handle/1887/73887> [21.10.2020].

Muehlberger, Guenter et al. (2019): “Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study”, in: *Journal of Documentation* 75, 5: 954–976. DOI: <https://doi.org/10.1108/JD-07-2018-0114>.

Rawson, Ken (2018): “The Rhetorical Power of Archival Description: Classifying Images of Gender Transgression”, in: *Rhetoric Society Quarterly* 48, 4: 327–351.

Romein, Annemieke et al. (2022): “Exploring Data Provenance in Handwritten Text Recognition Infrastructure: Sharing and Reusing Ground Truth Data, Referencing Models, and Acknowledging Contributions. Starting the Conversation on How We Could Get It Done”. DOI: <https://doi.org/10.5281/zenodo.7267245>.

Schijvenaars, Tess (2022): “Het transcriberen en annoteren van oorlogsbrieven. Hoe kan het ons helpen en wat gaat er verloren?”, in: *NIOD Blog* <https://www.niod.nl/nl/blog/oorlog-uit-eerste-hand-5> [10.06.2022].

Tames, Ismee (2020): “Digitale ontsluiting van het Centraal Archief Bijzondere Rechtspleging: Mogelijkheden en onmogelijkheden”, in: *Tijdschrift voor geschiedenis* 133, 2: 303–324.

Taskinen, Ilari et al. (2021): “Religion, Patriotism and War Experience in Digitized Wartime Letters in Finland, 1939–44”, in: *Journal of Contemporary History* 57, 3: 577–596.

Zaagsma, Gerben (2022): “Digital History and the Politics of Digitization”, in: *Digital Scholarship in the Humanities* 1–22.