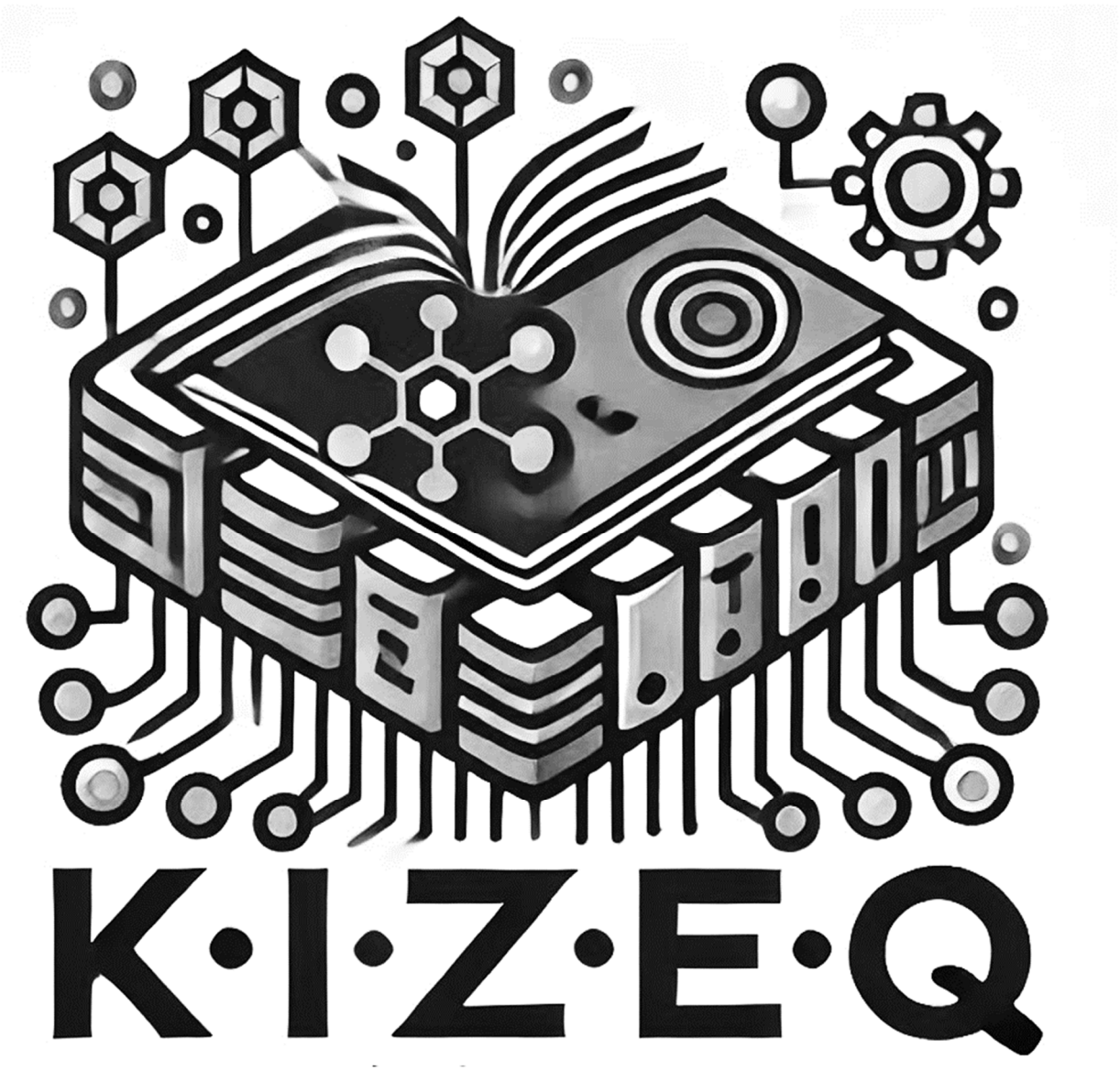
Projektphase I - KIZEQ/CORE:

Corpus-based analysis, Ontological modelling, Responsible anonymisation, Explainable AI Applications

v. 2.1 1/2026



Einrichtung: Universität Trier, Forschungsstelle SEAL
Projektkoordination: PD Dr. Massimiliano Livi (Geschichte)
Projektteam: Prof. Dr. Achim Rettinger (Computerlinguistik), Prof. Dr. Benjamin Weyers (Informatik), Dr. Thomas Grotum (Geschichte**)**

Project Description

KIZEQ/CORE aims to create a curated, semantically enriched, and FAIR/CARE-compliant corpus based on approximately 10,000 digitised documents from the AfW. The project addresses two strategic goals: first, the **creation of a high-quality dataset for training novel machine learning models for ethically responsible AI applications**; and second, **the development of a sustainable, standards-compliant data foundation designed for integration into existing national infrastructures and long-term reuse across disciplines**.
KIZEQ/**CORE** fills a critical gap in existing digital research environments by offering a corpus designed to support: **C**orpus-based analysis, enabling AI systems to reconstruct fragmented life histories from testimonial material; **O**ntological modelling, including the identification of legal, medical, and institutional entities for use in semantic graphs and digital knowledge infrastructures; **R**esponsible anonymisation, through the development and testing of privacy-preserving methods such as role abstraction and reversible pseudonymisation; **E**xplainable AI Applications: providing data for training AI models on historically embedded, multi-register, multilingual materials in a way that ensures transparency and ethical accountability.

Communities Demand and Preparatory Work: Despite ongoing efforts in Digital History to digitise archival holdings and harmonise metadata, a significant gap remains: curated, semantically annotated, and ethically governed corpora suitable for AI-based historical research are still scarce. This is particularly true in sensitive domains such as reparation, persecution, and post-conflict justice, where conventional archival systems are often inadequate for semantic modelling, narrative reconstruction, or responsible data reuse. Scholars increasingly agree that large-scale archival holdings like the AfW files must be made digitally accessible to enable critical and ethically informed historical analysis. Researchers in fields such as Holocaust studies, transitional justice, and digital memory are also calling for alternative, research-driven access models. Recent strategic frameworks further underscore this need (Makhortykh and Mann, 2024; Richardson-Walden and Marrison, 2023), especially in order to unlock ethically sensitive and currently under-utilised archival holdings (Barak-Cheney and Maman, 2024). The AfW corpus is also highly relevant beyond historical research. Its complexity aligns with the needs of communities working in Digital Humanities, Archival Science, corpus and computational linguistics, and applied AI (Walden-Richard and Makhortykh, 2024). These fields have consistently called for interoperable, ethically robust historical corpora to support both interdisciplinary research and computational model training.

Creation of a High-Quality Dataset for Training AI Systems: In direct response, KIZEQ/CORE delivers a research-driven corpus architecture featuring layered annotations (TEI/XML, JSON-LD), multilingual processing, and ontological modelling of legal, medical, and institutional discourse across time and context. A key deliverable is a high-quality dataset combining primary documents with a manually validated knowledge graph containing metadata, named entities, and relational structures. This enables the training of AI systems capable of temporally and semantically contextualised information extraction, particularly concerning personal histories across diverse temporal, institutional, and linguistic contexts. The AfW corpus includes extensive biographical and institutional data (e.g., multilingual person and place names, roles, affiliations) that support life-path modelling, discourse structure

detection, and historical entity linking. These requirements inform the design of annotation pipelines, knowledge representation layers, and evaluation metrics, structured around two concrete use cases.

The **composition, selection, and quality criteria** of the KIZEQ/CORE dataset follows structure, terminology and methodological recommendations based on established documentation frameworks for dataset transparency and reusability for Digital Cultural Heritage Datasets (Alkemade et al. 2023; Gebru et al. 2021) and machine learning usability (Oberbichler 2024a). The corpus is distinguished by its multi-register composition: individual files combine everyday language (e.g. autobiographical statements), bureaucratic and legal terminology (e.g. compensation decisions), as well as medical and psychiatric discourse (e.g. expert opinions, diagnoses). **This convergence of linguistic registers within a single case file providing different perspectives on the same individual is rare and provides novel challenges to language modelling approaches.** So far, the register of documents or any other metadata is not explicitly provided to and retained by language models and cannot transparently be accessed from the parametric memory of the model.

We contend that establishing an empirically grounded training corpus capable of handling diverse domain-specific registers is essential for advancing robust natural language understanding, knowledge representation and generation. Without such corpora, language models lack the necessary contextual awareness to accurately capture linguistic variation and register-induced biases. Additionally, the AfW source material is semi-structured. It includes genre-specific formats such as forms, tables, petitions, and legal rulings. These formats encode stable semantic relationships and hierarchical structures. While these recurring patterns pose challenges for language models, they provide critical signals that enhance training for entity and relation extraction, document structuring, and information retrieval from heterogeneous textual inputs.

Beyond annotation, **the dataset generated by KIZEQ/CORE actively enables the development of not only language models but also ontology-based AI methods for research on historical corpora.** The data includes clearly defined entities (e.g. persons, institutions, legal categories) and their interrelations across time, enabling the construction of semantically rich knowledge graphs with diachronic depth. The ontology developed in KIZEQ/CORE serves to formalise the conceptual and relational structures embedded in the reparation documents. It includes (1) the definition of core concepts such as persecution categories, recognition types, and evaluative institutions (cf. Vafaie 2024); (2) the mapping of semantic and procedural relationships between these concepts, such as taxonomic (e.g. "racial persecution" as a subcategory of "Nazi persecution"), causal, or procedural links (e.g. between application, medical review, and decision) (Vafaie/Bruns/Pilz 2023); (3) the development of a standardised, cross-disciplinary vocabulary that ensures semantic coherence; and (4) the disambiguation of context-dependent or polysemous terms within individual case files. This formal structure forms the prerequisite for advanced AI applications such as semantic search tools and trustworthy dialogue-oriented interfaces (e.g. research chatbots).

KIZEQ/CORE constitutes the foundational phase of the broader long-term initiatives KIZEQ/Engine and KIZEQ/Dialogue that the team aims to develop within the framework of the DFG e-Research Technologies programme. The overarching project, AI-based Infrastructure for Contemporary Witnessing, Memory, and Source Research (**KI**-gestützte **I**nfrastruktur zur **Z**eitzeugen-, **E**rinnerungs- und **Q**uellenforschung, **KIZEQ**), aims to unlock and structurally

process the AfW holdings through a sustainable, AI-enabled research and data infrastructure. Its ambition is not only to render these documents machine-readable, but to make them semantically accessible and contextually explorable, both for academic research and for civic participation in the “post-witness era” (Popescu and Schult 2015).

Strategic Alignment and Integration into Existing National Infrastructures: Rather than establishing a new infrastructure from scratch, the project enhances and complements existing frameworks by delivering a scalable, research-oriented, semantically enriched corpus optimised for scholarly analysis and AI training. KIZEQ/CORE adopts a vertically and horizontally integrated design, connecting archival data curation, domain-specific semantic modelling, and infrastructural interoperability. The resulting datasets will adhere to FAIR and CARE principles and fully comply with established repository and metadata standards. Integration is planned with relevant components of the NFDI landscape, particularly NFDI4Memory, CLARIN, and Text+.

KIZEQ/CORE complements the federal *Transformationsprogramm zur Sicherung und Digitalisierung der Wiedergutmachungsdokumente*, coordinated by the Federal Ministry of Finance. The project implements the standards defined in §4 of the corresponding framework agreement (Rahmenvereinbarung, 2022) for the ***Themenportal Wiedergutmachung*** **in the Archivportal-D**, as set out by the *Konferenz der Leiterinnen und Leiter der Archivverwaltungen des Bundes und der Länder* (KLA). This includes the metadata catalogue and descriptors specified in Annexes 1–3 of the KLA guideline *Erschließung von Entschädigungsakten gemäß BEG v. 18.9.1953*, version 2.0 (KLA-AG 2024; see Meyer/Götze 2025 for details). To ensure compliance and quality in the implementation of these standards the project will also draw on expert advice from the **Leibniz Institute FIZ Karlsruhe**. Further strategic collaborations are envisaged with the ***Landecker Digital Memory Lab***, an international, transdisciplinary hub investigating AI-based memory technologies and public history interfaces, and with the ***European Holocaust Research Infrastructure*** **(EHRI)**, the EU's flagship platform for integrated Holocaust archival research and digital access.

By connecting to these infrastructures, KIZEQ/CORE ensures both the future reusability of its corpus and its interoperability within a federated research ecosystem spanning historical, archival, and computational domains. The project is thus embedded in and strategically coordinated with broader national data initiatives and stands to play a key role in shaping future standards in the digital transformation of historically sensitive documents for training AI systems. A key priority is to ensure that the data architecture is modular, extensible, and compatible with future corpus integration, for example, via linked data or knowledge graphs.

In this respect, the project prepares the technical and methodological foundations for a federated, cross-institutional infrastructure for historical data. To support this, a dedicated Task T5.2 will initiate the “KIZEQ Forum”, a structured expert platform for coordinating terminology development, knowledge modelling, and standards alignment across different corpus holders.

Long-Term Reuse Across Disciplines: The KIZEQ/CORE corpus offers exceptional interdisciplinary value across law, psychology, and medicine. It comprises thousands of legal rulings, medical reports, and psychological assessments submitted in the context of individual reparation procedures. These documents provide unique insight into how institutions codified experiences of persecution, suffering, and trauma in the post-war period. **In legal scholarship**, the corpus supports analyses of how evidentiary categories, legal norms, and recognition criteria evolved under transitional justice regimes. It enables empirical study of bureaucratic

practices and the normative architecture of symbolic compensation. **In psychology and psychiatry**, the files document long-term trauma from the perspective of state recognition. They include psychiatric evaluations, neurological diagnoses, and impairment assessments that reflect both clinical discourse and survivors' efforts to articulate their suffering. The structured extraction of these materials allows for AI-supported modelling of institutionalised trauma narratives and retrospective mental health research. **In medicine**, especially orthopaedics and internal medicine, the corpus provides longitudinal data on physical harm resulting from forced labour, detention, malnutrition, and torture. These records form an evidence base for studies in historical health outcomes and administrative biopolitics.

Given this sensitive content, the project places a strong emphasis on **responsible data handling**. KIZEQ/CORE addresses this by developing domain-specific protocols based on reversible pseudonymisation, role-based generalisation, and semantic masking. These methods ensure data protection without compromising annotation depth or analytical value. An **advisory board of experts in law, ethics, archival science, and memory cultures** will oversee implementation, ensuring alignment with GDPR, FAIR/CARE, and archival ethics. Their expertise will support the development of a proof-of-concept for the ethically compliant use of personal documents in AI-driven research contexts and provide ongoing guidance to align data processing workflows with legal and ethical standards, fostering trustworthiness and compliance throughout the project lifecycle.

The public version of the corpus will be released, in line with FAIR/CARE principles, under a non-commercial academic licence (CC-BY 4.0). This ensures legal clarity, maximal scientific reusability, and alignment with DFG funding conditions. The restricted version (with reversible pseudonyms) will follow controlled access protocols, governed by institutional and legal oversight.

Project Readiness: Key preparatory work over the past three years has established the framework for corpus design, data sampling, and annotation strategy, supported by interdisciplinary workshops and early coordination with institutional partners (Service Center eSciences, University Library) and external collaborators (e.g. University of Bielefeld, NFDI4Memory, FIZ Karlsruhe). **A digitised sample corpus of approximately 3,000 case files (no longer subject to data protection embargoes) is currently being prepared in compliance with applicable archival and legal regulations**. This initial dataset, cleared for semantic processing and enriched with metadata, will serve as the foundation for KIZEQ/CORE and enables pilot annotation, machine learning workflows, and semantic enrichment to commence at the very start of the funding period.

Objectives

The **overall objective** during the funding period is the construction of a high-quality, machine-readable corpus. This will be enriched through multiple layers of semantic annotation and made accessible in interoperable formats (TEI/XML, JSON-LD), accompanied by full documentation, metadata, and persistent identifiers.
Thus, the objectives of KIZEQ/CORE are threefold:

- **To develop a representative corpus of AfW documents** that meets the highest standards of semantic quality, data protection, and technical interoperability. Rather than

aiming for volume, KIZEQ/CORE focuses on functional design, prioritising semantic utility, cross-domain adaptability, and scalable integration potential.

- **To support the training and evaluation of AI systems** in the Digital Humanities, particularly in the analysis of biographical, institutional, and discursive complexity in historical contexts.
- **To establish a reusable and expandable dataset** that serves as a corpus nucleus for the future integration of related collections and infrastructure initiatives.

The objectives are exemplified through two central use cases: (1) “Making life stories machine-readable”, focusing on the semantic modelling of fragmented biographical data; and (2) “From semantic fields to moral categories”, analysing how institutional language codified suffering. These cases inform annotation protocols, data structuring, and AI training workflows, and exemplify the cross-disciplinary relevance of the corpus.

The KIZEQ/CORE corpus is structured as a versioned, extensible dataset designed for iterative deployment and long-term reuse. The **release strategy** is aligned with key project phases and ensures both early demonstrability and methodological scalability:

- **Pilot Release (by Month 6)**: 50 fully annotated documents (including entity markup, discourse segmentation, and metadata linkage) drawn from the pool of 3,000 case files available at the start of the project for immediate integration and schema refinement.
- **Mid-Term Release (by Month 12)**: An additional 2,000 documents from the same pool, with semi-automatic annotation and quality-controlled metadata, supporting early model training and stakeholder feedback.
- **Final Release (by Month 24)**: Over 10,000 documents in fully structured formats, validated through automated quality benchmarks and reproducible annotation pipelines.

To ensure sustainability and transferability, the final release will include the annotation software and processing workflows, enabling external reuse and scalable corpus extension in line with FAIR/CARE principles. Notably, two version of the dataset will be released. The public one under a CC-BY 4.0 licence to guarantee open, legally compliant, and reproducible scientific use. Restricted versions with reversible pseudonyms will follow controlled access protocols under institutional oversight.

Relevance in Societal, Political and Science Policy Contexts: By enabling AI-compatible access to historically sensitive documents, the project supports the digital transformation of historical memory work and contributes to public education, critical digital literacy, and democratic engagement. In this regard, KIZEQ/CORE will contribute to a key debate in Holocaust and memory studies: not only how institutions can apply AI technologies, but how they should critically engage with AI-generated content and representations in a fragmented and polarised digital public sphere (Makhortykh and Mann, 2024). Then KIZEQ/CORE addresses an urgent need for high-quality, ethically grounded, and practically usable training data in a domain that is historically significant, politically sensitive, and pedagogically essential (IHRA Policy Briefing, 2024).

General ethical aspects: The KIZEQ/CORE project involves digitising, annotating, and processing historical case files from the *Amt für Wiedergutmachung* (AfW Saarburg), which document compensation claims by survivors of National Socialist persecution. **These files are no longer subject to data protection restrictions and can therefore be processed in accordance with applicable archival regulations**. However, these sources contain highly

sensitive personal information, including medical and psychiatric records, biographical data, and sworn testimonies that go far beyond the formal expiration of embargo periods. The semantic richness of the material is inseparably tied to its ethical sensitivity.
To ensure compliance with ethical, legal, and archival standards, KIZEQ/CORE embeds data protection and responsible handling of personal data at the core of its design. The project develops a domain-specific anonymisation protocol based on three complementary techniques:

- Role-based generalisation, replacing personal identifiers (e.g. names, addresses) with functional roles (e.g. claimant, medical expert);
- Reversible pseudonymisation, assigning cryptographically managed identifiers that enable re-linking under strict legal and ethical conditions;
- Context-sensitive semantic masking, which protects particularly sensitive content (e.g. psychiatric diagnoses) while maintaining the document's interpretability.

This hybrid anonymisation framework is **aligned with GDPR (especially Art. 89), the FAIR and CARE principles, and the requirements of German archival law (e.g. Landesarchivgesetz RLP)**. To balance data protection and research utility, the anonymisation framework is designed with two distinct access levels. First, a public dataset version will be published with full anonymisation and generalisation, suitable for public engagement and downstream reuse under academic licences (e.g. CC-BY 4.0). Second, a restricted research version will retain reversible pseudonyms and semantically sensitive details in controlled access environments. This tiered structure allows for use-case-specific access pathways under ethical oversight and ensures that sensitive historical materials remain accessible for qualified scientific purposes, while protecting individual privacy and complying with GDPR and archival law (e.g. Landesarchivgesetz RLP). Both versions will be governed by clearly defined access protocols and accompanied by provenance documentation and ethical usage guidelines.
All processing workflows are designed to ensure transparency and reproducibility. The anonymisation and pseudonymisation protocols are developed, documented in detail, version-controlled, and reviewed by an interdisciplinary ethics and legal advisory board. These protocols include:

- clear documentation of transformation rules,
- provenance metadata indicating anonymisation steps and decisions,
- and compliance records verifying alignment with GDPR (Art. 89) and Landesarchivgesetz RLP.

The protocols will be published under an open licence and designed for re-use in comparable domains, such as Holocaust archives, psychiatric records, or transitional justice datasets.
Access for general users and researchers will be provided under non-commercial, academic-use licences (e.g. CC-BY 4.0). Each data release is accompanied by:

- machine-readable documentation of data provenance and processing steps,
- persistent identifiers for stable referencing (e.g. DOI, URN),
- and clear versioning to ensure traceability of changes and updates.

The combination of these measures ensures that KIZEQ/CORE achieves both high ethical accountability and semantic integrity, enabling its safe use for reproducible, explainable AI applications and interdisciplinary historical research.

Project- and subject-related list of publications

- Alkemade, H., Candela, G., Claeyssens, S., Colavizza, G., Eren, S., Freire, N., Irollo, A., Isaac, A., Lehmann, J., Neudecker, C., Osti, G., van Strien, D., & Wevers, M. (2025): Datasheets for Digital Cultural Heritage Datasets (V. 2). Zenodo. https://doi.org/10.5281/zenodo.15828222
- Barak-Cheney, Y. & Maman, H. (2024): Exploring AI's Role in Holocaust Education: Powers and Pitfalls. In: Bricman, A., Jikeli, G., Baumkotter, A., Pook, L. K. et al. (Hg.): The Fragility of Freedom: Online Holocaust Denial and Distortion, Institute for Strategic Dialogue, Amman/Berlin. https://www.isdglobal.org/isd-publications/the-fragility-of-freedom-online-holocaust-denial-and-distortion/
- Brunner, J., Goschler, C., & Frei, N. (Hg.) (2013): Die Globalisierung der Wiedergutmachung. Politik, Moral, Moralpolitik. Göttingen: Wallstein.
- de Vita, L. & Goschler, C. (Hg.) (2025): Redefining Reparation. Wassenaar 1952 and the Global Politics of Repair. London: Routledge.
- Frei, N., Brunner, J., & Goschler, C. (2009): Komplizierte Lernprozesse. Zur Geschichte und Aktualität der Wiedergutmachung. In: Dies. (Hg.): Die Praxis der Wiedergutmachung, Göttingen: Wallstein, S. 9–47.
- Gebru, T., Morgenstern, J., et al. (2021): Datasheets for Datasets. In: Communications of the ACM, 64(12), S. 86–92. https://doi.org/10.1145/3458723
- Goschler, C. (1992): Wiedergutmachung. Westdeutschland und die Verfolgten des Nationalsozialismus 1945–1954. München: Oldenbourg.
- Harrower, N. (2020): Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities. ALLEA Report. https://www.ouvrirlascience.fr/sustainable-and-fair-data-sharing-in-the-humanities-recommendations-of-the-allea-working-group-e-humanities-2/
- IHRA – International Holocaust Remembrance Alliance (2024): Does AI Have a Place in the Future of Holocaust Memory? Policy Briefing, November 2024. https://www.digitalmemorylab.com/policy-briefing-ai-24/
- KLA-AG (2023): UAG Erschließung der KLA-AG „Transformation der Wiedergutmachung". Erschlieflung von Entschädigungsakten gemäß BEG v. 18.9.1953 - MetadatenkatalogVersion 2.0 (Stand:12.12.2024). https://wiki.deutsche-digitale-bibliothek.de/download/attachments/126518213/Metadatenkatalog_V2.0.pdf
- König, M. (2025): FAIR und grün: Empfehlungen für einen ressourcenschonenden Umgang mit Forschungsdaten. In: Digital Humanities am DHIP, 2.5.2025. https://doi.org/10.58079/13uu6
- Lehmann-Richter, A. (2010): Auf der Suche nach den Grenzen der Wiedergutmachung. Die Rechtsprechung zur Entschädigung für Opfer der nationalsozialistischen Verfolgung. Berlin: Berliner Wissenschafts-Verlag.
- Luthra, M., & Zijlma, A. (2025): From Bias Paralysis to Bias as a Category of Analysis: Introducing the Bias-Aware Framework. In: Rebora, S., Rospocher, M., Bazzaco, S. (Hg.): Diversità, Equità e Inclusione: Sfide e Opportunità per l'Informatica Umanistica nell'Era dell'Intelligenza Artificiale, AIUCD2025, S. 86–93. https://doi.org/10.6092/unibo/amsacta/8380
- Makhortykh, M. & Mann, H. (2024): AI and the Holocaust: Rewriting History? UNESCO & World Jewish Congress. https://doi.org/10.54675/ZHJC6844
- Mauermann, J. & Oberbichler, S. (2025): LLM Biases: Expected and Unexpected Model Design Effects in Historical Newspaper Article Extraction on the Messina Earthquake. DH Lab. https://doi.org/10.58079/137qr
- Meyer, N. & Götze, O. (2025): EAD(DDB)-WGM – EAD(DDB) Anwendungsprofil für personenbezogenes Schriftgut im Themenportal Wiedergutmachung. https://wiki.deutsche-digitale-bibliothek.de/spaces/DFD/pages/126518213/EAD
- Oberbichler, S. (2024a): AI Model Research Documentation Sheet (AIRDocS). Zenodo. https://doi.org/10.5281/zenodo.14550113
- Oberbichler, S. (2024b): Large-Scale Research with Historical Newspapers: A Turning Point through Generative AI. Digital Humanities Lab, Leibniz-Institut für Europäische Geschichte. https://doi.org/10.58079/11v9i
- Oberbichler, S. & Petz, C. (2025): Evaluating Bias within an Epistemological Framework for AI-Based Research in the Humanities. In: Rebora, S., Rospocher, M., Bazzaco, S. (Hg.), Diversità, Equità e Inclusione, S. 52–59. https://doi.org/10.6092/unibo/amsacta/8380
- Popescu, D. I. & Schult, T. (2015): Revisiting Holocaust Representation in the Post-Witness Era. Palgrave Macmillan. https://doi.org/10.1057/9781137530424

- Rahmenvereinbarung (2022): Rahmenvereinbarung über die Bereitstellung von digitalen Inhalten für ein „Themenportal Wiedergutmachung“. Bundesministerium der Finanzen. https://www.bundesfinanzministerium.de/Content/DE/Downloads/Oeffentliche-Finanzen/Wiedergutmachung/rahmenvereinbarung-themenportal-wiedergutmachung.pdf
- Rebora, S., Rospocher, M. & Bazzaco, S. (Hg.) (2025): Diversità, Equità e Inclusione: Sfide e Opportunità per l’Informatica Umanistica nell’Era dell’Intelligenza Artificiale. Verona: AIUCD. https://doi.org/10.6092/unibo/amsacta/8380
- Richardson-Walden, V. G. & Marrison, K. et al. (2023): Recommendations for Digitising Material Evidence of the Holocaust. Sussex: REFRAME. https://doi.org/10.20919/fiov3702
- Richardson-Walden, V. G. & Makhortykh, M. (2024): Imagining Human-AI Memory Symbiosis: How Re-Remembering the History of Artificial Intelligence Can Inform the Future of Collective Memory. Memory Studies Review, 1(20), S. 1–18.
- Schöch, H. (2001): Wege und Irrwege der Wiedergutmachung im Strafrecht. In: Roxin, C. & Schünemann, B. (Hg.): Festschrift für Claus Roxin zum 70. Geburtstag, Berlin: de Gruyter, S. 1045–1064.
- Soper, E., Fujimoto, S. & Yu, Y.-Y. (2021): BART for Post-Correction of OCR Newspaper Text. In: Proceedings of the 7th Workshop on Noisy User-Generated Text (W-NUT 2021), S. 284–290. https://doi.org/10.18653/v1/2021.wnut-1.31
- Thonke, C. (2004): Hitlers langer Schatten. Der mühevolle Weg zur Entschädigung der NS-Opfer. Wien: Czernin.
- Vafaie, M. (2024): German National Socialist Injustice on the Semantic Web: From Archival Records to a Knowledge Graph. In: ISWC 2024 Doctoral Consortium. Baltimore. https://ceur-ws.org/Vol-3884/paper8.pdf
- Vafaie, M., Bruns, O., Pilz, N., Waitelonis, J. & Sack, H. (2023): CourtDocs Ontology: Towards a Data Model for Representation of Historical Court Proceedings. In: Proceedings of the 12th Knowledge Capture Conference (K-CAP’23), S. 175–179. https://doi.org/10.1145/3587259.3627562
- Volmer-Naumann, J. (2012): Bürokratische Bewältigung. Entschädigung für nationalsozialistisch Verfolgte im Regierungsbezirk Münster. Essen: Klartext.
- Wachter, C. & Jentsch, P. (2025): LLM-Supported Workflow for Processing Faulty OCR. GitHub. https://github.com/Digital-History-Bielefeld/llm-supported-workflow-for-processing-faulty-ocr

Für weitere Informationen kontaktieren Sie bitte:

SEAL
Forschungs- und Dokumentationsstelle
Universität Trier / DM 252
Universitätsring 15
D-54296 Trier
+49 (0) 651/201-3332
seal@uni-trier.de
https://www.forschungsstelle-seal.de/projekte/kizeg-projekt/