J. Z. Krasni

https://orcid.org/0000-0002-5192-8007
✉ *jan.krasni@uni-marburg.de*
Марбургский университет (Германия, Марбург)

Towards an ethical strategy for research data infrastructures: Digitalizing archives of historical hate

Аннотация. В статье рассматриваются этические сложности, связанные с разработкой инфраструктур исследовательских данных (Research Data Infrastructures, RDI) для оцифрованных архивов с акцентом на материалы, содержащие исторический контент, касающийся вражды. Рассматривается напряжение, с одной стороны, между принципами открытого доступа и массовой оцифровки, которые направлены на повышение доступности знаний, и, с другой стороны, с этическим императивом предотвратить распространение вредного контента, который может увековечить предвзятые идеологии или вредные стереотипы. Для решения этих проблем автор предлагает комплексную этическую стратегию, которая объединяет подход трансэпистемического осмысления с принципами организационного обучения (в понимании Organisations-pädagogik). В этой стратегии особое внимание уделяется сотрудничеству между различными заинтересованными сторонами — архивистами, исследователями, экспертами в области информационных технологий и заинтересованными сообществами — для создания как этически надежных, так и практически жизнеспособных решений. Выходя за пределы исключительно технических или правовых рамок, предложенный подход стремится сбалансировать доступность исторических документов для исследовательских целей и необходимость снижения рисков, связанных с распространением ненавистнического контента. В статье рассматриваются такие важные вопросы, как алгоритмическая предвзятость, которая может непреднамеренно усилить вредные стереотипы, и потенциал двойного назначения технологий искусственного интеллекта, когда технологии, созданные для повышения эффективности архивного дела, могут быть использованы не по назначению. Автор также обсуждает противоречие между принципами открытой науки и ограниченным доступом к чувствительным материалам, выступая за тонко настроенные модели контроля доступа как для человеческих пользователей, так и для систем искусственного интеллекта. Благодаря трансэпистемическому подходу стратегия поддерживает междисциплинарный диалог таким образом, чтобы инфраструктуры исследовательских данных были этичными, сохраняли исторические знания и защищали их от вреда, способствуя формированию стандартов ответственного цифрового архивирования.

EDN: RIDITQ

Ключевые слова: этическая стратегия, инфраструктуры исследовательских данных, цифровые архивы, ненавистнический контент (контент вражды), алгоритмическая предвзятость, трансэпистемический подход к проектированию, открытый доступ, этика искусственного интеллекта

Для цитирования: *Krasni J. Z.* Towards an ethical strategy for research data infrastructures: Digitalizing archives of historical hate // Шаги/Steps. Т. 11. № 4. 2025. С. 205–221. EDN: RIDITQ.

Поступило 13 сентября 2024 г.; принято 9 октября 2025 г.

Shagi / Steps. Vol. 11. No. 4. 2025
Articles

J. Z. Krasni

https://orcid.org/0000-0002-5192-8007
✉ *jan.krasni@uni-marburg.de*

University of Marburg
(Germany, Marburg)

Towards an ethical strategy for research data infrastructures: Digitalizing archives of historical hate

Abstract. This paper explores the ethical complexities of developing research data infrastructures (RDIs) for digitalized archives, with a focus on materials containing historical hateful content. It examines the tension between the principles of open access and mass digitization, which aim to enhance knowledge accessibility, and the ethical imperative to prevent the dissemination of harmful content that could perpetuate biased ideologies or harmful stereotypes. The author proposes a comprehensive ethical strategy that integrates a trans-epistemic design approach with principles of organizational learning and development to address these challenges. This strategy emphasizes collaboration among diverse stakeholders — archivists, researchers, IT experts, and affected communities — to create solutions that are both ethically robust and practically viable. By moving beyond solely technical or legal frameworks, the approach seeks to balance the accessibility of historical records for research purposes with the need to mitigate risks associated with the spread of hateful content. The paper delves into critical issues such as algorithmic bias, which can inadvertently amplify harmful stereotypes, and the dual-use potential of AI, where technologies designed for archival efficiency might be misused. It also addresses the conflict between open science principles and restricted access to sensitive materials, advocating for nuanced access controls for both human users and AI systems. Through a trans-epistemic

EDN: RIDITQ

lens, the strategy fosters interdisciplinary dialogue to ensure RDIs serve as ethical infrastructures that preserve historical knowledge while safeguarding against harm, contributing to a framework for responsible digital archiving.

Keywords: ethical strategy, Research Data Infrastructures (RDI), digital archives, hateful content, algorithmic bias, trans-epistemic design approach, open access, AI ethics

To cite this article: Krasni, J. Z. (2025). Towards an ethical strategy for research data infrastructures: Digitalizing archives of historical hate. *Shagi / Steps, 11*(4), 205–221. EDN: RIDITQ.

Received September 13, 2024; accepted October 9, 2025

Introduction

This paper discusses the ethical issues related to the integration of AI in Research Data Infrastructures (RDI), the digitalization of historical archives, and the challenges of dealing with historically problematic content: various materials promoting hatred and discrimination. RDI is understood as a broad technological framework which, nowadays, is being built to be both interoperable (among each other) and compatible with data analysis and data-processing tools. These tools are based on machine learning technology, i. e., AI, which is intended to help automate many research processes. RDI and AI open partly overlapping but different sets of ethical risks that this paper reflects upon.

The field of research libraries and archives introduces a series of peculiar contradictions in ethical discussions concerning AI and RDI. There, the fundamental principles of open science and the imperative for open research data directly conflict with the ethical need to moderate (i. e. restrict) access to historiographic and archival knowledge — even to those materials for which copyright protection does not apply. These paradoxes challenge the existing (mostly utopian) paradigm of ethical discussions about the future of knowledge apparatuses, and about the application of AI technologies in their broad variety of use, which both enables access and raises concerns regarding uncontrolled spread and re-creation of hateful materials. Furthermore, the possibility of algorithmic bias reinforcing such content further complicates the matter.[1]

Therefore, this article suggests an ethical strategy — currently only on a theoretical level — as a way to create an ethical intervention by regarding both sets of principles, but also the needs, technology, and the opinions of various stakeholders with a shared interest in the field. The trans-epistemic design approach

[1] For example, there is a large body of scientific work from the past which develops and supports racist theory or antisemitism. It resides in historical and research libraries. This scientifically supported discrimination could potentially serve for AI-driven generation of new discriminatory material and for justification of discriminatory policies, etc.

would consider various perspectives, scientific traditions, institutional (bodies of) knowledge of practices and procedures, and a unique setting of shared interest where new value-based apparatus would be formed.

This paper supports the idea that mass digitalization is necessary for the protection and accessibility of historical material, but that accessibility should be nuancedly restricted both for algorithms and for a human audience, making the application of novel research still possible.

Theoretical approach

This paper is written in the spirit of discourse analysis and follows on the concept of apparatus (dispositif) in Agamben's sense of the term [Agamben 2009]. In other words, it refers to the terminology and the analytical reflective framework of this paradigm. It therefore considers the main relevant epistemic fields – ethical and some of the technical discussions concerning the RDI and AI – and based on those suggests its own solution. Following Agamben's methodological approach, this paper argues that the solution lies in a trans-disciplinary understanding and engagement with diverse epistemic fields and stakeholders, which forms the basis of the proposed ethical strategy. The specificity of this case is, however, that the apparatus itself still does not exist, as it is in the making – and it is just a matter of time when it will emerge.

Discussing the solution for the ethical modus operandi of a RDI-apparatus (i. e., its role in society, and within the human-technology network) may help us practically leave the Foucauldian dystopian properties of the oppressive apparatus. The prospect of an ethical infrastructure should be therefore understood as a strategic, productive and creative process of apparatus construction that takes into consideration analysis of relevant discourses of its emergence, thereby offering ground for the implementation of our proposed ethical strategy.

The following literature overview will, however, show us various discourse topics within the diverse field of AI and RDI ethics.

AI ethics

Central themes in AI ethics — such as bias, fairness, accountability, transparency, and the embedding of values — are critical not only in AI [Novelli et al. 2024], but also in broader contexts of technological infrastructures, including research data infrastructures (RDIs). This discussion addresses some fundamental ethical concerns and lays foundation for understanding how they inform and relate to ethical considerations in RDIs.

Addressing bias in domains like healthcare, criminal justice, and employment (to mention some of them) requires diverse datasets and robust frameworks for responsible deployment. Ferrara [2023] stresses the importance of equity in federated AI systems, where balancing privacy and fairness is particularly challenging. Similarly, Mehrabi et al. [2021] underscore the societal risks of biased AI systems, emphasizing the need for interventions to prevent the amplification of inequalities. These principles, while rooted in AI ethics, resonate strongly with

concerns surrounding bias in RDIs, especially in the case of historical collections of hateful material where data equity cannot be ensured (think of Nazi files on what we today call vulnerable groups).

Accountability ensures that AI systems operate within ethical bounds and remain subject to oversight (by, e. g. implementing specific filters and safety measures). Kroll [2020] links accountability to transparency, highlighting the importance of creating systems understandable to users and stakeholders. Gualdi and Cordella [2021] address the human-technology control dilemma, illustrating how accountability mechanisms can mediate the balance between human oversight and technological autonomy [Nickel 2022]. Raji et al. [2020] call for actionable measures to implement accountability in AI, a challenge that also emerges in the governance of RDIs, where ethical oversight is vital. Accountability in our case can be understood in case of human control of access and depth of approach to the collections.

Transparency underpins trust in AI systems, particularly in high-stakes applications like healthcare. Subías-Beltrán et al. [2024] highlight the role of transparency in fostering patient confidence in AI-driven radiology tools. Hemphill et al. [2023] further identify clear regulations and scientific validation as essential for building trust in AI systems. These insights directly inform discussions around RDIs — and potentially suggest a possible model for AI-supported archives — where transparency in data collection, usage, and governance is equally significant for maintaining trust among stakeholders.

The embedding of ethical values such as fairness, autonomy, and non-maleficence is foundational to ethical AI design. Poel [2020] advocates for the alignment of AI systems with societal values, ensuring they serve the common good. Jobin and Ienca [2019] provide a global overview of AI ethics guidelines, demonstrating a consensus on value-driven AI development. However, Whittlestone et al. [2019] caution that while defining ethical principles is essential, practical implementation remains a substantial challenge — a theme that parallels the ethical integration efforts required in RDIs. In this context, it is crucial to note that the idea of "trans-epistemic design," as defined by Keller and Weber [2020], offers a valuable framework for addressing this challenge. It refers to a process of design that aims to bring together different perspectives from diverse scientific fields and knowledge systems, allowing for the emergence of new solutions that are both ethically and practically grounded.

Research data infrastructures: bridging ethics and governance

Research data infrastructures (RDIs) are systems — or rather apparatuses — developed to store, analyze, and cross-reference scientific research data, mostly within specific disciplines. Their development, a global endeavor, must reconcile competing priorities, including compatibility of data and metadata standards, safety, accessibility, legal frameworks, and adherence to ethical principles such as FAIR (Findability, Accessibility, Interoperability, Reusability) in various fields, from biomedicine and life sciences to art history and ethnographic collections. Rooted in the emergence of big data and data mining, RDIs are deeply connect-

ed to the history of data management and governance, encompassing practices for organizing, generating, and managing data. This section reviews key ethical considerations in RDIs, emphasizing their relevance to the evolving research landscape. While these ethical concerns are crucial for all RDIs, they are particularly acute when considering research data infrastructures for digitized archives, especially those that include historically loaded and potentially harmful material.

The treatment of human subjects in big data research is a pressing ethical concern. Metcalf & Crawford [2016] discuss the "ethical divide" in big data, where traditional ethical frameworks often fail to adequately protect individuals. The authors highlight how the anonymity and informed consent processes foundational to ethical research frequently compromise privacy and autonomy. These considerations extend also to depictions of victims which are often part of the historical visual collections concerning the slave trade, colonization in Africa, and the Holocaust. Ethics committees, tasked with overseeing research practices, face challenges in adapting to the complexities of big data. Favaretto et al. [2020] explore researchers' experiences with ethics review processes, finding that existing guidelines may inadequately address novel ethical challenges. They advocate for committees to move beyond gatekeeping and engage collaboratively with researchers to foster a culture of ethical awareness. This aligns, on the one hand, with Finn & Shilton [2023], who emphasize the need for updated ethics codes and governance structures within scientific communities to address the ethical implications of data sharing and research practices. On the other hand, it follows the line of our research that supports working groups whose design supports interdisciplinary, interinstitutional, and even more trans-epistemic exchange.

Rantanen et al. [2019] argue — considering how to balance accessibility with ethical considerations — that effective governance should avoid exploitative or harmful outcomes. Without such frameworks, data ecosystems risk undermining trust and perpetuating inequities, thereby harming to individuals and communities. The philosophical underpinnings of data ethics are critical to shaping ethical (and especially AI-driven) RDIs [Ryan 2020]. Floridi & Taddeo [2016] provide a comprehensive analysis of data ethics, addressing dilemmas related to data collection and any types of its usage. They contend that ethical frameworks must be context-specific, reflecting the diverse applications of data in such fields as biomedical research and social sciences.

Balancing broad access to data with the rights of individuals is another significant ethical challenge. Shabani et al. [2021] outline principles for data access governance, aiming to protect data subjects while ensuring researchers can access the data needed for scientific progress. Similarly, Ochang et al. [2022] examine the ethical and legal governance of brain data, advocating for clear guidelines that safeguard individuals while enabling innovation. The last two points are crucial for our argument, as they emphasize equally important principles of open science with its research approaches (which is usually impossible in archives), and protection of individuals (and groups) from harms of possible manipulation.

Archives and research libraries as e-infrastructures

Our interest is focused on the application of AI tools in digitalized archives and specialized research libraries. While this niche topic is closely connected with the aforementioned (better researched) fields, it still overlaps only partly with points discussed in the general literature on digital ethics, and remains an area that has been insufficiently addressed in existing literature.

The implementation of AI is closely tied to the very practical problems of "technological enveloping" of the analogue process in order to make it conform to digital (and therefore AI accessible) environment [Floridi 2011, Floridi 2023: 39-42]. This "technological enveloping," in the case at hand, would entail the transformation of analogue archival processes into automated digital workflows — which would certainly further lead to disruption of traditional archival work. Furthermore, there is an unaddressed issue of the dual use of machine learning technologies, which is often overlooked in discussions of algorithmic bias (e. g. recognizing patterns and not imposing them where they do not fit; applying opposing rules depending on the situation). Both of these problems (dual use and technological enveloping) are further linked to the principles of Open Science[2] and Open Data policy[3] — the very concepts that guide the global initiative for building RDIs, and are particularly strong in Europe through the European Strategy on Research Infrastructures.[4] It is surprising, therefore, that the principles of open science are rarely mentioned in the literature on AI ethics. These principles represent baseline concepts and the true justification for the process of global datafication and mass digitization. They give meaning to and enable the emergence of RDIs for storing and analysis of research data, as well as movements like citizen science, concerned with using open data to repeat scientific results and bring science into everyday life. As Holterhoff [2017] argues, the transformation of archival materials into a digital format may require an explicit critique of the archive's power dynamics (i. e. traditional, conservative role and the sets of rules connected to it) that are often hidden in the structure of the catalogue and the metadata themselves.

There are numerous projects engaged in mass digitalization of cultural and scientific heritage. Although these projects are practically focused on creating RDIs, they were often not explicitly framed as such. Only recently, within the European context, have consortia emerged to pursue large RDI projects for cultural institutions, research in the humanities and social sciences — and archives. For a long time, archives were mainly just subject to (mass) digitization, with less attention to the infrastructural aspect. The infrastructures created rely on specialized software solutions for archives that emulate traditional structures. Only

[2] See UNESCO official document (URL: https://unesdoc.unesco.org/ark:/48223/pf0000379949).

[3] See the information hub on open data policy (URL: https://opendatapolicylab.org), and the Open Data Directive by the EU (URL: https://eur-lex.europa.eu/legal-content/EN/TXT/ ?qid=1561563110433&uri=CELEX:32019L1024).

[4] See the roadmap for constitution of Research Infrastructures in Europe (URL: https://roadmap2021.esfri.eu/media/1295/esfri-roadmap-2021.pdf).

recently are there organized infrastructural experiments to develop tools enabling navigation and research.

In recent years, there has been a growing number of discussions about applying AI-driven tools in archives to innovate older structures [Colavizza et al. 2021; Cushing et al. 2023], as well as some early projects and experiments [Randby, Marciano 2020].[5] All discussions about creating digital copies (or "digital twins") [Marchello et al. 2023] rely on the idea of opening knowledge pools, thus enabling insights from all of the collective heritage. In archival practice, traditional ethical questions are often tied to the problem of selection, preparation, and digitization [Manžuch 2017]. This also represents a way to institutionalize public, national, or collective memory through the process of selecting material for digitization (to be remembered or forgotten).

In North American and UK discussions, there is growing awareness of the importance to also consider how the choices of selection and description, as well as lack of context, can shape how collections are understood, particularly when they contain hateful (e. g. racist) material [Chilcott 2019]. This process discusses archiving everything to a kind of memory politics through the politics of digitization [Zaagsma 2022]. Furthermore, in this context, discussions about ethical use of digitized collections are closely linked to discussions about ownership and the right to interpretation of the archive, as shown by works on racist memorabilia, such as the "Jim Crow Museum" [Pilgrim 2005].

Toxic bias — patterns of social condition

The most prevalent and widely discussed problems concerning AI and ML are the so-called algorithmic biases. If the training of algorithms is based on datasets that "do not reflect the world as it is" — in other words, if they produce a partisan view of a specific phenomenon — this systemic error will be repeated whenever the algorithm is applied, or, more accurately, whenever AI is used to perform a task for which it was designed (e.g. generate content based on trained data). The usual concern is that flawed data will produce unfair outcomes and reinforce existing, socially conditioned injustices [O'Neil 2016]. There are numerous examples of algorithmic bias across a broad range of fields, including medical and healthcare data [Jones et al. 2024], complex autonomous technological systems [Yang et al. 2024], business decision-making processes [Ghasemaghaei, Kordzadeh 2024], and education [Dieterle 2024]. Furthermore, many socially conditioned injustices [Kordzadeh, Ghasemaghaei 2022], such as racism, antisemitism, misogyny, white supremacism, and sexism — along with a plethora of other forms of prejudice-based social hate that reinforce real forms of social suffering — can be emulated and reinforced through AI [Abramson et al. 2024]. All of these problems are, of course, reflected in the data on which the algorithms are trained.

[5] The project Aureka.ai is one of the startups coming from a FU Berlin project. However, it is a commercial project. There are plenty of public projects such as Transkribus or eSkriptorium that are perfectioning OCR and HTR through LLMs.

One should be aware of the fact that machine learning is based on the property of recognizing patterns. The study of social hate, within antisemitism studies, shows that hurtful emotions are often crystallized in specific prejudices [Jensen, Schüler-Springorum 2013; Ascone et al. 2023], in stereotypes representing a communicative pattern of the relation towards a given group, thus constructing the given group as vulnerable and unworthy of being regarded as equal. In other words, the recognition and reinforcement of such biases are not "wrong" from a purely technical point of view: they represent the very task machine learning is designed for. If we accept that data is not "flawed" but that they merely mirror our society, it should be our humanistic education and values that guide us in rejecting injustice — also on the level of AI. The core ethical issue here is not the existence of these patterns, but how to stop them from being perpetuated through technological means. Dual use of AI in this case does not mean omitting harmful material, but recognizing it and, for example, warning, marking, and manage access to the harmful content.

This situation is more complex when considering various archival collections. Imagine a historical library in an ethnographic or anthropological museum in a country with a colonial past. Historical scientific books and papers in the field of anthropology or ethnography from the time before the Second World War often represent a rather problematic worldview. They contain hurtful stereotypes of various races, ethnicities, religious communities, and other groups we today recognize as vulnerable. An interested researcher will always have access to such historical collections. However, it is entirely different if such a collection is made available online without the proper historical context provided by historians. In that case, this kind of historical literature would stand symbolically for the whole institution — and vice versa — in the present day. The very same problem, however, appears also in other types of historical archives that can be understood as the "toxic chambers" of our society, such as the material from research institutions dealing with the Holocaust, slavery, and other types of historically loaded material.

Toxic chambers of history and contemporaneity

While the previous discussion has focused on ethical challenges in RDI and AI in general, this section now turns to a more specific area: specialized institutions and collections dealing with historical injustices. These include institutions like museums and memorial centers dedicated to genocides, slavery, racism worldwide addressing various forms of social hate, discrimination, and suffering. The historical material these institutions collect and store is of utmost importance for research, memorialization, rituals of remembrance, and for understanding the social development of various groups in society. However, these primary sources also reveal a persistent pattern — that of hate and horror. Such archives are often large, and frequently not entirely indexed for researchers. Such material, if found within general public institutions, is typically stored in specialized departments (called "Giftkammer" in the jargon of librarians), accessible only to researchers with a specialized permit.

Furthermore, we are confronted with online repositories of research data in linguistics, sociology and IT (or in an interdisciplinary setting) focused on contemporary hateful material. This type of research is quite widespread in recent years and has also a practical side as it helps to develop automated moderation. The research data in that case are usually annotated (and not always anonymized) linguistic or multimodal corpora from social media. After the research, the material should be stored according to most Research Data Management guidelines for a period of 10 years. Such material should be accessible in order for other researchers to verify or challenge the results of the given research. However, most repositories with such material do not make it conform to the FAIR principles nor make it openly accessible (i. e. without elaborate access management which includes human evaluation).

In many countries, policies supporting the digitalization of historical collections often require institutions to make these resources available through Open Access.[6] This aligns with the principles of Open Science. Projects like Europeana[7] represent a good example of how the need for common heritage to be studied in the digital sphere is met. Europeana enables remote and open access to collections, facilitating new combined research approaches (in the spirit of "distant reading") to better understand the collective past. However, such policy would be problematic for the digitalization of archives with hateful content, as it would enable the spread of harmful material.

There are ongoing discussions about the complexities of mass digitalization and Open Access when it comes to scientific heritage, legacies, and intellectual property,[8] but archives storing collections of hateful, problematic, and inappropriate content are still not widely discussed. This gap leads to problematic situations that are not covered by any legal frameworks, established ethical or any best practices. Instead, they are addressed through the approximate and often subjective feelings of the institutions (or the person) responsible for digitalization. For example, the German Federal Archive may enable online viewing of Nazi propaganda videos, while the republication and commercial sale of books such as Hitler's *Mein Kampf* (or other hateful propaganda) are forbidden or permitted only with historical context and commentary. In contrast, a joint project by the Big Ten Academic Alliance in the United States has made a collection of racist and white supremacist newspapers accessible for paying members, citing the importance of preserving even disturbing historical records to prevent the forgetting of a shameful past.[9] These different approaches show that there is no clear consensus internationally on how to provide access to sensitive and symbolically

[6] Compare for example National Library of Finland's policy that draws on Open Science (URL: https://www.doria.fi/handle/10024/180341) or German Science Society's (DFG) guidelines for digitization and indexing (URL: https://www.dfg.de/en/research-funding/funding-opportunities/programmes/infrastructure/lis/funding-opportunities/digitisation-cataloguing).

[7] See more about it on: (URL: https://www.europeana.eu/en/about-us).

[8] See the discussion that deals with exactly this problem (URL: https://www.dfg.de/de/aktuelles/neuigkeiten-themen/info-wissenschaft/2021/info-wissenschaft-21-108).

[9] See more about it on: (URL: https://dwso.revealdigital.org).

problematic historical content, despite the clear role of public archives to ensure that historical documents are accessible to the public and that the historical context should be provided for the problematic collections.[10]

However, the use of these digitalized objects to train algorithms presents a significant challenge. Applying AI and training algorithms on mass-digitalized material from archives and specialized libraries is obviously a step that awaits these institutions in order to make them searchable and easily accessible by the researchers, students and other interested parties. The problem becomes alarming when algorithms are trained exclusively on hateful material. With the implementation of AI into the RDIs, the algorithms will be trained also on this data. The dual use of AI would then consider technical tasks (e. g., OCR and HTR, automated generation of metadata and indexing) and far less the (re)creative tasks associated with generative AI by training Large Language Models for generative AI.

The ethical dilemma can, therefore, be formulated as follows: how can we digitize and apply new ML and NLP technologies to problematic and hateful material, and how can we open up data and its AI derivatives for research, while simultaneously avoiding the sharing of historical hateful content on other social media platforms? This dilemma highlights the conflict between the open science principle and the need to prevent the spread of propaganda online against the backdrop of the historical hateful material.

Factors for ethical strategy — legal and value systems

Depending on the country, the handling of hateful material while digitalizing and publishing varies significantly. In the US, for example, the presence of hateful language from the past does not seem to present a barrier to the dissemination of historical material. The First Amendment, which protects freedom of speech, stands against any form of institutional censoring. If we consider the JFK archive, for instance, we can see that racist language in the collections is made available without censorship, though a context warning is typically provided.[11]

As Naumann [2023] reminds us, making archival records available online is regulated by national and regional laws, and internal institutional rules. While the specific legal situation in Germany is very strict in this regard, other countries in Europe also have similar laws. Specifically, there are restrictions that target the content of archival records and prevent their publication. For example, German laws prohibit making archival content publicly accessible online if it promotes group-based enmity. Furthermore, even after the record's protection period has ended (which can be between 70 and 110 years), protection can be prolonged if the content displays group-based enmity or other unconstitutional elements [Naumann 2023: 240–241]. This is particularly tied to the protection of human col-

[10] Even though there is little chance that the old propaganda videos, books or historic material would seduce any soul nowadays, the risks are that this material could be used in the process of radicalization of society. E.g. instructions on torturing from the past times.

[11] See, for example, the collection Racist literature (BIBPP-012-001) at the JFK Archive website (URL: https://www.jfklibrary.org/asset-viewer/archives/bibpp-012-001).

lectives. Any content that disseminates propaganda material of unconstitutional organizations, incites hatred, graphically depicts violence, or glorifies war should not be made available online, as it contravenes a range of specific German laws.

> Accordingly, even records that are many decades or even centuries old could be excluded from free access. This is not necessary, for example, for interrogation protocols under torture from witch trials, as the language and writing are only accessible to a few knowledgeable people. For typewritten texts with corresponding content, however, a waiver of free online access on this basis is entirely appropriate [Naumann 2023: 249].

One may argue that Naumann interprets these regulations in the strictest manner possible — but one cannot rule them out. Furthermore, as all digitalization is usually tied to Open Access, this means that large and relevant historical, yet controversial, collections in Europe may be excluded from any kind of indexing and algorithmic processing (except for metadata), which would further render them unsearchable and unfindable within the archival RDIs. The underlying reason for this exclusion would be a fear of legal consequences. While the process of digitalization could easily be dissociated from Open Access, it remains to be seen whether access to AI tools will be differentiated from public dissemination.

Beyond the challenge of hateful content, there is ongoing debate regarding the use of problematic, though historically accurate, terms in archival metadata that are now considered unacceptable due to their discriminatory implications [Doğtaş et al. 2022; Naß 2020; Modest, Lelijveld 2018]. This concerns the use of such terminology in metadata, catalogues, normed files, and controlled vocabularies for classifying archived objects [Strickert 2021]. For example, terminology related to racial categories created and institutionalized in specific historical contexts often features in the metadata of historical archives, thereby reinforcing outdated and harmful concepts. How to meet this ethical demand while maintaining the discoverability of relevant records remains an open question.

Ethical strategy within a trans-epistemic setting

Developing an ethical strategy is crucial for addressing anticipated changes in the future, both in the digitalization of archives and in research infrastructures. As is already evident, the use of LLMs impacts work-intensive practices like OCR and HTR. The next logical step, already being widely tested, is automated annotation and metadata generation. Subsequently, archival search and research functions will be expanded through various ML-driven tools or appropriate APIs, which would enable researchers to create tailored research tools. To consider both technical and value system inputs for ensuring ethical RDI development, we need a trans-epistemic approach, forming working groups that represent varying views.

An "ethical strategy," as formulated by Luciano Floridi [2023: 90–91], should prevent undesirable (future) developments and their implementation during the innovation process. It comprises four tasks to ensure ethical design and governance: critically questioning the project, signaling ethical problems, engaging

stakeholders affected by the issue, and designing and implementing shareable solutions. In the context of archive digitalization, this collaboration plays a vital role in engaging with public and collective memory, acknowledging historical facts, and fostering transgenerational engagement between the descendants of perpetrators and victims to support a shared and sustainable future.[12]

Several recent projects explore the intersection of trans-epistemic design (TED) and organizational-pedagogical approaches for producing human-computer and cognitive assistance systems [Keller et al. 2024] — processes that resemble the creation of an ethical archive-RDI. These projects use participatory methods within an interdisciplinary framework, creating access points for stakeholders to ensure that technological solutions are not only technically sound, but also ethically grounded and pedagogically effective [Keller, Weber 2020; Heidelmann, Weber 2020, Keller et al. 2021]. For example, research on mobile digital assistance systems has explored how cognitive ergonomics (assistive technology) can be achieved with a trans-epistemic approach [Ebert et al. 2022]. Furthermore, prototype-oriented projects using advanced technologies such as VR / AR in workplace contexts, highlight the need for learning-theoretical foundations in their design [Haase et al. 2020].[13]

As argued by Keller and Weber, the TED process enables new forms of knowledge production or reflection during the creation of cyber-physical systems and cognitive assisting systems [Keller, Weber 2020: 627–629]. This focus on bringing stakeholders together and enabling them to understand each other's perspectives, which is at the core of trans-epistemic design, corresponds both with ethical deliberation and the necessary participative co-creation of the cognitive assistant system that an RDI-archive will undoubtedly become. In other words, having an ethical board "in the loop" is a solution-oriented way to avoid judicial, ethical, and technological pitfalls associated with an archive-RDI. Finally, only in this setting would it be possible to jointly formulate and implement ethical principles and guidelines in technological solutions and infrastructure.

By combining a trans-epistemic design process with an organizational-pedagogical framework, the ethical strategy would involve creating a system where various perspectives on the problem are heard, and where the solutions are generated through a process of learning and co-creation that enables sustainable and ethically informed outcomes in archival digitization and the use of AI.

Discussion

As we have seen, the ethical challenges posed by artificial intelligence, research data infrastructures, and the principles of open science are not always directly aligned. This misalignment significantly complicates the creation of RDI-archives, especially given the existence of historically hateful material and the

[12] This is of highest importance not only for social cohesion, but also for prevention of future similar wrongdoings.

[13] These interdisciplinary teams work together on projects situated at various the R&D institutes, e. g. EverAssist project at Fraunhofer Institute for Factory Operation and Automation IFF (URL: https://www.everassist.de).

foreseeable, and often desired, application of AI tools within these contexts. This shows that the issue of digitalized archives is particularly complex, as it brings together two equally important, but often conflicting sets of values – the value of knowledge sharing and open access on the one side, and the value of protecting specific social groups from harm, on the other.

To address this complexity, it seems necessary to integrate an ethical research organization within the RDI-apparatus. A trans-epistemic exchange between interested and affected stakeholders is essential to enable an interdisciplinary approach, ensuring that both ethical risks and relevant theoretical insights from diverse domains are considered. As seen in our previous discussion, traditional approaches to RDI governance, focusing on technical issues and legal requirements, are not always capable of dealing with the ethical questions presented here. Such an exchange of perspectives, therefore, has the potential to lead to a paradigm shift in technological innovation, allowing for the development of systems that do not only provide new possibilities for research and access to historical material, but are also based on a strong ethical foundation.

References

Abramson, C. M., Li, Z., Prendergast, T., & Sánchez-Jankowski, M. (2024). Inequality in the origins and experiences of pain: What "big (qualitative) data" reveal about social suffering in the United States. *RSF: The Russell Sage Foundation Journal of the Social Sciences, 10*(5), 34–65. https://doi.org/10.7758/RSF.2024.10.5.02.

Agamben, G. (2009). *What is an apparatus? and other essays*. Stanford Univ. Press. https://doi.org/10.1515/9781503600041.

Ascone, L., Becker, M. J., Bolton, M., Chapelan, A., Haupeltshofer, P., Krasni, J., Krugel, A., Mihaljević, H., Placzynta, K., Pustet, M., Scheiber, M., Steffen, E., Troschke, H., Tschiskale, V., & Vincent, Ch. (2023). *Decoding antisemitism: An AI-driven study on hate speech and imagery online*. Technische Universität Berlin. Centre for Research on Antisemitism. https://doi.org/10.14279/depositonce-17105.

Chilcott, A. (2019). Towards protocols for describing racially offensive language in UK public archives. *Archival Science, 19*, 359–376. https://doi.org/10.1007/s10502-019-09314-y.

Colavizza, G., Blanke, T., Jeurgens, C., & Noordegraaf, J. (2021). Archives and AI: An overview of current debates and future perspectives. *Journal on Computing and Cultural Heritage, 15*(1), Article 4. https://doi.org/10.1145/3479010.

Cushing, A. L., & Osti, G. (2023). "So how do we balance all of these needs?": How the concept of AI technology impacts digital archival expertise. *Journal of Documentation, 79*(7), 12–29. https://doi.org/10.1108/JD-08-2022-0170.

Dieterle, E., Dede, C., & Walker, M. (2024). The cyclical ethical effects of using artificial intelligence in education. *AI & Society, 39*, 633–643. https://doi.org/10.1007/s00146-022-01497-w.

Dogtas, G., Ibitz, M.-P., Jonitz, F., Kocher, V., Poyer, A., & Stapf, L. (2022). Kritik an rassifizierenden und diskriminierenden Titeln und Metadaten – Praxisorientierte Lösungsansätze. *Critical Library Perspectives, 9*(4), 1–14. https://doi.org/10.21428/1bfadeb6.abe15b5e.

Ebert, K., Bode, M., Haase, T., & Keller, A. (2022). Mobile digitale Assistenzsysteme in der Weberei — Anforderungen an die kognitiv ergonomische Gestaltung. In *Technologie und Bildung in hybriden Arbeitswelten. 68. GfA-Frühjahrskongress 2022* (pp. 1–6) (n. p.).

Favaretto, M., Clercq, E., Briel, M., & Elger, B. (2020). Working through ethics review of big data research projects: an investigation into the experiences of Swiss and American researchers. *Journal of Empirical Research on Human Research Ethics, 15*(4), 339–354. https://doi.org/10.1177/1556264620935223.

Ferrara, E. (2023). *Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies* (preprint). https://doi.org/10.2196/preprints.48399.

Finn, M., & Shilton, K. (2023). Ethics governance development: the case of the Menlo Report. *Social Studies of Science, 53*(3), 315–340. https://doi.org/10.1177/03063127231151708.

Floridi, L. (2011). Enveloping the world for AI. *The Philosophers' Magazine, 54*(54), 20–21. https://doi.org/10.5840/tpm20115437.

Floridi, L. (2023). *The ethics of artificial intelligence: Principles, challenges, and opportunities*. Oxford Univ. Press.

Floridi, L., & Taddeo, M. (2016). What is data ethics? *Philosophical Transactions of the Royal Society a Mathematical Physical and Engineering Sciences, 374*(2083), 20160360. https://doi.org/10.1098/rsta.2016.0360.

Ghasemaghaei, M., & Kordzadeh, N. (2024). Understanding how algorithmic injustice leads to making discriminatory decisions: An obedience to authority perspective. *Information & Management, 61*(2), 1–14. https://doi.org/10.1016/j.im.2024.103921.

Gualdi, F., & Cordella, A. (2021). Artificial intelligence and decision-making: The question of accountability. In *Proceedings of the 54th Hawaii International Conference on System Sciences* (pp. 2297–2306) (n. p.). https://doi.org/10.24251/hicss.2021.281.

Haase, T., Keller, A., Radde, J., Berndt, D., Fredrich, H., & Dick, M. (2020). Anforderungen an die lerntheoretische Gestaltung arbeitsplatzintegrierter VR-/AR-Anwendungen. In GfA Dortmund (Ed.). *Digitaler Wandel, digitale Arbeit, digitaler Mensch?* B.16.1., 1–7.

Heidelmann, M. A., Weber, S. M. (2022). Eine Haltung ausbilden — Organisationen und Netzwerke beraten lernen. Mit symbolischen Ordnungen der Beratung zur Organisationspädagogischen Professionalisierung. In J. Elven, & S. M. Weber (Eds.). *Beratung in symbolischen Ordnungen. Organisationspädagogische Analysen sozialer Beratungspraxis* (pp. 325–356). Springer. https://doi.org/10.1007/978-3-658-13090-9_17.

Hemphill, S., Jackson, K., Bradley, S., & Bhartia, B. (2023). The implementation of artificial intelligence in radiology: a narrative review of patient perspectives. *Future Healthcare Journal, 10*(1), 63–68. https://doi.org/10.7861/fhj.2022-0097.

Holterhoff, K. (2017). From disclaimer to critique: Race and the digital image archivist. *Digital Humanities Quarterly, 11*(3). http://www.digitalhumanities.org/dhq/vol/11/3/000324/000324.html.

Jensen, U., & Schüler-Springorum, S. (2013). Einführung: Gefühle gegen Juden. Die Emotionsgeschichte des modernen Antisemitismus. *Geschichte Und Gesellschaft, 39*(4), 413–442. https://doi.org/10.13109/gege.2013.39.4.413.

Jobin, A., & Ienca, M. (2019). The global landscape of ai ethics guidelines. *Nature Machine Intelligence, 1*, 389–399. https://doi.org/10.1038/s42256-019-0088-2.

Jones, C., Castro, D. C., De Sousa Ribeiro, F., et al. (2024). A causal perspective on dataset bias in machine learning for medical imaging. *Natural Machine Intelligence, 6*, 138–146. https://doi.org/10.1038/s42256-024-00797-8.

Keller, A., Selinski, A., Vuong, T. H. C., & Haase, T. (2024). Stakeholderspezifische Zugänge zu arbeitsgestalterischen Inhalten — technisch-didaktische Konzeption und erste Erkenntnisse. *Arbeitswissenschaft in-the-loop. Mensch-Technologie-Integration und ihre Auswirkung auf Mensch, Arbeit und Arbeitsgestaltung*, I.1.4, 1–6. https://doi.org/10.24406/publica-3888.

Keller, A., & Weber, S. M. (2020). Trans-epistemic design-(research): Theorizing design within industry 4.0 and cognitive assistive systems. In *Proceedings of the Design Society: DESIGN Conference* (Vol. 1, pp. 627–636). Oxford Univ. Press. https://doi.org/10.1017/dsd.2020.173.

Keller, A., Weber, S.M., Rentzsch, M. Haase, T. (2021). Lern- und Assistenzsysteme partizipativ integrieren – Entwicklung einer Systematik zur Prozessgestaltung auf Basis eines organisationspädagogischen Ansatzes. *Zeitschrift für Arbeitswissenschaft, 75*, 455–469. https://doi.org/10.1007/s41449-021-00279-2.

Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems, 31*(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212.

Kroll, J. A. (2020). Accountability in computer systems. In M. D. Dubber, F. Pasquale, & S. Das (Eds.). *The Oxford handbook of ethics of AI* (pp. 180–196). Oxford Univ. Press. https://doi.org/10.1093/oxfordhb/9780190067397.013.10.

Manžuch, Z. (2017). Ethical issues in digitization of cultural heritage. *Journal of Contemporary Archival Studies, 4*, Article 4. 1–17.

Marchello, G., Giovanelli, R., Fontana, E., Cannella, F., & Traviglia, A. (2023). Cultural heritage digital preservation through AI-driven robotics. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (Vol. XLVIII-M-2-2023, pp. 995–1000) (n. p.). https://doi.org/10.5194/isprs-archives-XLVIII-M-2-2023-995-2023.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), 1–35. Article 115. https://doi.org/10.1145/3457607.

Metcalf, J., & Crawford, K. (2016). Where are human subjects in big data research? The emerging ethics divide. *Big Data & Society, 3*(1), 1–14. https://doi.org/10.1177/2053951716650211.

Modest, W. (2018). Words matter. In P. Lelijveld (Ed.). *Words Matter. An unfinished guide to word choices in the cultural sector* (E-book, pp. 13–17). Tropen Museum, Afrika Museum, Museum Volkenkunde, Wereled Museum. URL: https://www.materialculture.nl/en/publications/words-matter.

Naß, M. A. (2020). Was darf die Kunst(institution)? Zwischen dem white cube als safe space und Zensur als Neurechter Kampfbegriff. In C. M. Ruederer (Ed.). *Infrastructures — Online Reader*. Kunstverein München e. V. URL: https://www.kunstverein-muenchen.de/de/programm/programmreihen/2020/infrastructures/online-reader.

Naumann, K. (2023). Ethische Grundlagen der Onlinestellung von digitalisierem Archivgut und deren Umsetzung. *Recht und Zugang, 4*(3), 237–252. https://doi.org/10.5771/2699-1284-2023-3.

Nickel, P. (2022). Trust in medical artificial intelligence: A discretionary account. *Ethics and Information Technology*, *24*(1), Article 7. https://doi.org/10.1007/s10676-022-09630-5.

Novelli, C., Taddeo, M., & Floridi, L. (2024). Accountability in artificial intelligence: what it is and how it works. *AI & Society, 39*, 1871–1882. https://doi.org/10.1007/s00146-023-01635-y.

Ochang, P., Stahl, B., & Eke, D. (2022). The ethical and legal landscape of brain data governance. *Plos One, 17*(12), e0273473. https://doi.org/10.1371/journal.pone.0273473.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Pilgrim, D. (n. d., retrieved 2005). The garbage man: Why I collect racist objects. In *Jim Crow Museum*. https://jimcrowmuseum.ferris.edu/collect.htm.

Poel, I. (2020). Embedding values in artificial intelligence (AI) systems. *Minds and Machines, 30*(3), 385–409. https://doi.org/10.1007/s11023-020-09537-4.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In M. Hildebrandt, & C. Castillo (Eds.). *FAT*'20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 33–44). Association for Computing Machinery. https://doi.org/10.1145/3351095.3372873.

Randby, T., & Marciano, R. (2020). Digital Curation and Machine Learning Experimentation in Archives. In Xintao Wu et al. (Eds.). *2020 IEEE International Conference on Big Data (Big Data)* (pp. 1904–1913). IEEE. https://doi.org/10.1109/BigData50022.2020.9377788.

Rantanen, M., Hyrynsalmi, S., & Hyrynsalmi, S. (2019). Towards ethical data ecosystems: A literature study. In *2019 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC): Conference Proceedings ICE/IEEE ITMC 2019 #47383: Co-creating our Future: Scaling-up Innovation Capacities through the Design and Engineering of Immersive, Collaborative, Empathic and Cognitive Systems* (pp. 1–9). IEEE. https://doi.org/10.1109/ice.2019.8792599.

Ryan, M. (2020). In AI we trust: Ethics, artificial intelligence, and reliability. *Science and Engineering Ethics, 26*(5), 2749–2767. https://doi.org/10.1007/s11948-020-00228-y.

Shabani, M., Thorogood, A., & Murtagh, M. (2021). Data access governance. In G. Laurie et al. (Eds.). *The Cambridge handbook of health research regulation* (pp. 187–196). Cambridge Univ. Press. https://doi.org/10.1017/9781108620024.023.

Strickert, M. (2021). Zwischen Normierung und Offenheit – Potenziale und offene Fragen bezüglich kontrollierter Vokabulare und Normdateien. *LIBREAS. Library Ideas, 40*, 1–19. https://doi.org/10.18452/23807.

Subías-Beltrán, P., Pitarch, C., Migliorelli, C., Marte, L., Galofré, M., & Orte, S. (2024). The role of transparency in AI-driven technologies: Targeting healthcare. In E. P. Dados (Ed.). *Artificial intelligence — Ethical and legal challenges* (forthcoming; online first). 1–21. http://dx.doi.org/10.5772/intechopen.1007444.

Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019). The role and limits of principles in AI ethics: Towards a focus on tensions. In V. Conitzer et al. (Eds.). *AIES'1: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 195–200). Association for Computing Machinery. https://doi.org/10.1145/3306618.3314289.

Yang, H. F., Zhao, Y., Cai, J., Zhu, M., Hwang, J. -N., & Chen, Y. (2024). Mitigating bias of deep neural networks for trustworthy traffic perception in autonomous systems. In *2024 IEEE Intelligent Vehicles Symposium (IV)* (pp. 633–638). IEEE. https://doi.org/10.1109/IV55156.2024.10588805.

Zaagsma, G. (2023). Digital history and the politics of digitization. *Digital Scholarship in the Humanities*, *38*(2), 830–851. https://doi.org/10.1093/llc/fqac050.

Информация об авторе

Ян Златкович Красни
PhD
научный сотрудник, Департамент образования, Марбургский университет,
Bunsenstr. 3, 35032, Marburg, Germany
✉ *jan.krasni@uni-marburg.de*

Information about the author

Jan Zlatković Krasni
PhD
Research Fellow, Department of Education, University of Marburg
Bunsenstr. 3, 35032, Marburg, Germany
✉ *jan.krasni@uni-marburg.de*