

Article

Not peer-reviewed version

---

# Advancing Historical Research Through AI and Data-Centric Approaches

---

[Wolfgang Göderle](#)<sup>\*,†</sup>, [Malte Rehbein](#)<sup>†</sup>, Markus Gerstmeier<sup>†</sup>

Posted Date: 25 February 2026

doi: 10.20944/preprints202602.1333.v1

Keywords: digital source criticism; computational history; computer vision





Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Advancing Historical Research Through AI and Data-Centric Approaches

Wolfgang Göderle <sup>1,2,3,\*</sup> , Malte Rehbein <sup>2,3,†</sup>  and Markus Gerstmeier <sup>2,†</sup> 

<sup>1</sup> University of Graz, Department of Digital Humanities, Austria

<sup>2</sup> University of Passau, Department of Computational Humanities, Germany

<sup>3</sup> Max Planck Institute of Geoanthropology, Department Structural Changes of the Technosphere, Germany

\* Correspondence: wolfgang.goederle@uni-graz.at

† These authors contributed equally to this work.

## Abstract

The rapid digitization of large source collections in the humanities over the last three decades has comprehensively transformed the discipline: The accessibility of primary sources has improved drastically, the pre-processing of research data has been revolutionized in some areas, and new transdisciplinary approaches have emerged and become possible. However, the theoretical grounding of these developments has not kept pace with the changed realities of the research process in many respects: most critically, the concept of “information” — central to computer science and computational methods — has so far been insufficiently received and theorized within historical methodology. In this contribution, we employ a concept from Science and Technology Studies — Bruno Latour’s “circulating reference” — to analyze and render describable the processes of historical research within a digitized research environment. Through three case studies — AI-supported segmentation of Habsburg cadastral maps (1817–1861), computational analysis of the Hof- und Staatsschematismus (1702–1918), and the datafication of the Munich Special Court archive inventory — we demonstrate how and at which specific points historical research benefits from this framework, and what new insights it enables.

**Keywords:** digital source criticism; computational history; computer vision

## 1. Introduction

Massive digitization efforts in the past two decades have fundamentally transformed the conditions of historical research. The increasing availability of machine learning for automating indexing tasks — particularly layout detection and OCR/HTR — as well as the widespread accessibility of generative AI, will continue to reshape research conditions (Eberle et al. 2024; Siebold and Valleriani 2022). Whether these technical innovations will transform research practices on the broadest level remains to be seen; the development of digital methods in Digital and Computational Humanities and the discourse about Digital History have not yet led to their methodological innovations being translated into historical research at large (Hiltmann 2024). However, we argue that the most profound impact goes beyond mere accessibility of sources or computational power.

Digitization and AI are catalyzing an epistemological shift that reveals a fundamental tension between traditional conceptualizations of historical sources and emerging data-centric approaches. For over a century, historical methodology has conceptualized sources not as “information containers” but through a rich vocabulary of epistemological concepts: in the German tradition as *Quellen* distinguished into *Überreste* (unintentional remains) or *Tradition* (intentional testimonies) (Droysen 1868); in the Anglo-American world as “evidence” requiring authentication and interpretation (Collingwood 1946); and in French historiography as *documents* and *témoignages* demanding rigorous *critique externe* and *critique interne* (Langlois and Seignobos 1898). Notably absent from this vocabulary was any systematic conceptualization of sources as containing “information” in a technical sense. Sources

were understood as fragments requiring reconstruction, testimonies demanding evaluation, or traces needing interpretive contextualization – but not as information repositories awaiting extraction.

The emergence of “information” as a category for understanding historical sources is a comparatively recent development, introduced by Manfred Thaller in the context of Historical Information Science (Thaller 1993). It has since appeared primarily in contexts of digitization and computational analysis, introducing a conceptual framework borrowed from information science rather than emerging from historical methodology itself. This shift is not merely terminological. When sources undergo digital preprocessing – through OCR, database structuring, georeferencing, or AI-powered analysis – they are transformed into “information objects.” More significantly, these processes reveal a granular layer beneath the documentary surface: data. Digital tools, particularly AI-powered instruments, allow us to decompose what traditional methodology understood as holistic traces or testimonies into constituent data elements that can be manipulated, recombined, and analyzed at scale.

This transformation raises fundamental questions: What happens when we reconceptualize sources from traces requiring interpretive reconstruction to information awaiting extraction? What is lost – and what new analytical possibilities emerge – when the hermeneutical tradition meets the computational paradigm of data processing? Can computational methods be productively applied to historical research when the discipline’s theoretical foundations provide no systematic framework for conceptualizing information, data, or their relationship to historical knowledge (Hiltmann 2024)?

This article examines this largely untheorized conceptual shift and its implications for historical epistemology. We argue that historians are increasingly adopting computational methods without the theoretical grounding necessary to justify their epistemological claims. Drawing on the methodological traditions of German *Quellenkritik*, Anglo-American empiricism, and French critical approaches, we first demonstrate that historical methodology has systematically avoided theorizing “information” as a category, instead privileging selective, exemplary source work justified through hermeneutical interpretation. We then propose that frameworks from Science and Technology Studies, particularly Bruno Latour’s concept of “circulating reference,” offer productive pathways for bridging the gap between traditional source criticism and computational analysis.

Our analysis proceeds in five parts. First, we examine how traditional historical methodology has conceptualized sources, demonstrating the absence of systematic information theory and the methodological legitimacy historically accorded to selective source work. Second, we explore how Latour’s work on knowledge construction might provide theoretical foundations for understanding the relationship between data, information, and historical knowledge. Third, fourth, and fifth, we demonstrate these theoretical considerations through three case studies: AI-supported segmentation of Habsburg cadastral maps from the long 19th century, computational analysis of the *Schematismus*, a comprehensive directory of officeholders spanning 1700–1918, and the datafication of the archive inventory of the Munich Special Court. Through this investigation, we aim to contribute to what several scholars have called “digital hermeneutics”: an approach that maintains the critical rigor of traditional historical methodology while embracing the transformative potential of digital technologies (Fickers and Zaagsma 2022; Siebold and Valleriani 2022).

## 2. History as Data Science? The Absence of Information Theory in Historical Methodology

The increasing application of computational methods to historical research confronts a fundamental problem: the discipline’s theoretical foundations provide no systematic framework for understanding what “information” is, how it relates to historical sources, or what distinguishes “data” from “knowledge.” This absence is not merely an oversight but reflects deeper methodological commitments that have shaped historical practice for over a century.

When we examine foundational textbooks that have shaped generations of historians – such as *Grundlagen des Studiums der Geschichte* by Boshof, Düwell, and Kloft (Boshof et al. 1973), *Werkzeug des Historikers* by Ahasver von Brandt (Brandt 2012), *The Pursuit of History* by John Tosh (Tosh 2021),

From *Reliable Sources* by Howell and Prevenier (Howell and Prevenier 2001), or *The Historian's Craft* by Marc Bloch (Bloch 1953) – a consistent pattern emerges. While these works demonstrate sophisticated approaches to source criticism and interpretation, none develops a systematic theory of what constitutes “information” in historical sources or how such information relates to the data from which it is constructed or the knowledge it produces. The current generation of textbooks that explicitly address digital and computational approaches – such as *Digital Humanities: Eine Einführung* (Jannidis et al. 2017), *Digital Humanities in den Geschichtswissenschaften* (Antenhofer et al. 2024), and *Digital History: Konzepte, Methoden und Kritiken* (Döring et al. 2022) – has made significant strides in establishing the practical and methodological foundations for computational work with historical sources, and has importantly begun to articulate the epistemological challenges that arise from this transformation. Building on these contributions, the present article seeks to advance the theoretical integration of information and data concepts into historical epistemology – a task that these works identify as a desideratum but that, given their broader scope, they necessarily leave to more specialized investigations.

Traditional historical methodology treats “information” as recoverable content within sources, requiring proper methodological application through *Quellenkritik* for extraction and interpretation. *Quellenkritik* provides sophisticated tools for establishing authenticity and context; hermeneutics guides interpretation; but neither tradition requires historians to theorize what “information” itself might be in a technical or formal sense. Information is simply assumed to be directly accessible through established hermeneutic procedures.

This theoretical underdevelopment becomes striking in comparison with adjacent fields. In Science and Technology Studies, scholars like Bruno Latour have systematically theorized the construction and circulation of scientific knowledge. In information science, concepts like data granularity, knowledge representation, and the relationships between these categories have been elaborated into comprehensive theoretical frameworks. Within historical scholarship itself, Manfred Thaller's pioneering work since the 1980s on “Historical Information Science” explicitly theorized information structures in historical sources (Thaller 1993), and subsequent overviews have documented this sub-discipline's development (Boonstra et al. 2004). However, these contributions remained confined to a specialized community and were never integrated into mainstream historical methodology. Historical methodology has largely remained in what might be called a pre-theoretical relationship to its fundamental epistemic objects.

This absence reflects the discipline's long-standing methodological privilege of selective source work – the epistemological legitimacy of drawing historical insights from carefully chosen examples without systematic assessment of their representativeness. Traditional methodology has operated with what might be called a “principle of exemplary interpretation”: the assumption that individual sources, properly contextualized through *Quellenkritik*, can yield valid insights without requiring assessment of their statistical representativeness.

This approach reflects the discipline's hermeneutical foundations. The German tradition explicitly distinguished between *verstehen* (understanding) and *erklären* (explaining), with the former requiring interpretive depth rather than quantitative breadth. Anglo-American empiricism similarly privileged the historian's judgment in selecting sources over systematic sampling. French historiography, from the *école méthodique* through the Annales School, emphasized critical evaluation of documents but not the systematic theorization of information as such. Even Marc Bloch argued that historians must “select and sort” facts as a necessary form of abstraction, without developing a systematic theory of how this selection relates to representativeness or information completeness.

This deficit produces a persistent methodological dichotomy between the micro-perspective – deep, context-rich, but statistically vulnerable – and the macro-perspective – broad, structural, but often detached from primary sources. We argue that the integration of AI and data-centric approaches provides the epistemological infrastructure to reconcile these perspectives. By subjecting massive corpora to computational analysis (referencing the “Totum” rather than the “Pars”), we no longer rely on the *Pars pro Toto* fallacy alone. Instead, data-centric history enables a “scalable hermeneutics”:

the ability to toggle between the distant reading of structural patterns and the close reading of the individual anomaly. In this framework, the general truth is derived from the comprehensive data layer, which allows us to identify the specific case as either representative or exceptional with statistical confidence. The “Pedocomparator” of the digital age does not flatten the micro-perspective; it validates it by placing it within an empirically secured macro-context.

Even quantitatively oriented traditions engaged with representativeness only selectively. The Annales School’s emphasis on serial sources and *longue durée* introduced quantitative methods but within specific research agendas rather than as a general methodological requirement. The cliometric revolution, pioneered by scholars like Robert Fogel in the 1960s, developed sophisticated statistical methods but remained a specialized subfield (Fogel and Engerman 1974). Its methodological innovations were never translated into broader historical methodology; most historians continued to work with selective sources justified through hermeneutical interpretation (Hudson 2000).

The result is a striking lacuna at the heart of historical methodology. Despite generations of reflection on source criticism, narrative construction, and historical explanation, historical scholarship has not produced a canonical theory of information in sources. This works reasonably well for close reading and hermeneutical interpretation, where the trained historian’s tacit knowledge guides effective source work without requiring explicit theorization. However, this pre-theoretical approach becomes untenable in the context of digital transformation (Fickers et al. 2022; Hiltmann 2024). When we work with machine learning, natural language processing, or database systems, we necessarily operate with “data,” “information,” “metadata,” and “knowledge” as technical categories with specific meanings. These are functional components of computational systems, not mere metaphors. Digital humanities projects must specify what constitutes a “data point,” how “information” is structured and encoded, how “knowledge” is represented and queried – conceptual precision that historical methodology has not traditionally provided.

As datafication transforms historical sources into structured datasets amenable to computational analysis at scale, the methodological privilege of selective source work becomes harder to sustain. When cadastral records spanning entire territories become computationally accessible, when parish registers can be analyzed in their totality rather than through selected samples – the question of representativeness shifts from a specialized concern to a fundamental methodological challenge. AI-supported research, by making comprehensive analysis technically feasible, exposes what was previously a pragmatic necessity as a methodological choice requiring explicit justification.

In attempting to bridge this gap, many digital historians have turned to theoretical frameworks from information science. The DIKW (Data-Information-Knowledge-Wisdom) hierarchy has proven particularly attractive, appearing frequently in digital humanities discourse as an apparently ready-made solution. Despite being what Rowley (Rowley 2007) calls a ‘fundamental, widely recognized and taken-for-granted model,’ the DIKW hierarchy has been criticized as ‘strongly depend[ing] on a subjective viewpoint’ (Rehbein 2021, p. 161) or ‘unsound and methodologically undesirable’ (Frické 2009). Applied to historical sources, these problems become even more pronounced: sources resist hierarchical ordering, simultaneously embodying data points, conveying information, representing knowledge claims, and requiring wisdom for interpretation.

The uncritical adoption of such frameworks risks obscuring what makes historical sources distinctive: their fragmentary nature, their embeddedness in specific contexts, their resistance to standardization, and their demand for interpretive rather than extractive approaches. Yet the digital transformation also offers unprecedented opportunities. The challenge lies in developing theoretical frameworks that preserve the hermeneutical depth of traditional source criticism while embracing the analytical possibilities of computational methods. We argue that such frameworks exist, but must be sought not in information science’s hierarchical models but in Science and Technology Studies’ more nuanced accounts of how data, inscriptions, and knowledge circulate and transform.

### 3. Bridging the Gap: Circulating Reference and Historical Sources

The theoretical lacuna identified in the previous section—the absence of a systematic framework for understanding information in historical sources—cannot be filled by simply importing concepts from information science. The DIKW hierarchy, as we have seen, imposes a linear structure fundamentally at odds with the iterative, hermeneutical nature of historical interpretation (Drucker 2011). Yet the digital transformation of historical research demands conceptual frameworks that can bridge traditional source criticism and computational analysis. We propose that such a framework exists, not in information science, but in Science and Technology Studies, specifically in Bruno Latour's concept of "circulating reference" (Latour 1999a). While Latour's Actor-Network Theory has faced substantial critique—particularly regarding human agency (Spiegel 2005), epistemological relativism (Bloor 1999), and ahistoricism (Finzsch 2004), we adopt his transformation chain framework pragmatically as a heuristic for analyzing digitization processes rather than endorsing ANT's broader ontological claims. Latour's work has seen similar pragmatic use in historical research on other occasions (Goederle 2015).

#### 3.1. Latour's Circulating Reference: From Matter to Form

In his ethnographic study of soil scientists working in the Amazon forest of Boa Vista, Latour (Latour 1999b) traces how observations of physical phenomena—the forest-savanna transition—become scientific knowledge through a series of transformations.<sup>1</sup> His account begins not with data or information but with the material world itself: soil, plants, earthworms.

Latour concluded from his participant observation that – at each stage of the scientist's research work – something is lost and something gained: When botanist Edileusa extracts plant specimens from the forest, she loses the living ecosystem but gains portable, preservable samples. When these specimens are dried, pressed, and labeled, they lose their three-dimensionality and color variation but gain stability and compatibility with taxonomic systems. When pedologist René takes soil samples using standardized protocols, he loses the continuous variation of the soil profile but gains discrete, comparable data points. When these samples are encoded using the Munsell color system, they lose their material presence but gain universal commensurability—any soil scientist anywhere can understand "10YR3/2." This principle of simultaneous loss and gain through transformation has parallels in historical epistemology. Daston and Galison (Daston and Galison 2007) trace how mechanical reproduction technologies in 19th-century science—photography, lithography—similarly involved trade-offs between interpretive richness and reproducible standardization, which they term "mechanical objectivity." The transformation chains we observe in digital history represent a contemporary instantiation of this ongoing tension between the singular and the mobile, the interpretive and the systematic.

Crucially, Latour emphasizes that reference is not a correspondence between words and world, between language and reality, but rather a property of the entire chain of transformations. Truth and accuracy depend not on eliminating mediations but on maintaining their traceability. The pedocomparator (the device that allows soil samples to be systematically compared and encoded) does not distort reality—it constructs a reality that can be known, manipulated, and communicated. As Latour writes: "To the metaphysical question, 'Does the world exist independently of our representation?' one can only answer: 'Yes, of course,' but this question immediately gives way to another: 'How many mediations are necessary for us to gain access to it?'" (Latour 1999b, p. 69). This formulation directly challenges the representationalist epistemology underlying much traditional source criticism, which seeks to minimize mediation between historian and past. Latour's framework instead valorizes mediation as being constitutive of knowledge-making (Rheinberger 1997).

Reference circulates in both directions. The expedition can return "upstream" from the final scientific diagram to the forest site because each transformation preserved essential features while making them mobile. The numbered tags nailed to trees, the coordinates inscribed in notebooks,

<sup>1</sup> The Boa Vista expedition is recounted in Chapter 2 ("Circulating Reference") of Latour (Latour 1999b, pp. 24-79). For earlier formulations of these ideas on "immutable mobiles" and "centers of calculation," see Latour (1986).

the soil samples in plastic bags—these create a reversible path that allows movement back and forth between matter and form, between the forest and the laboratory, between observation and inscription. This bidirectionality is crucial for historical research: unlike purely extractive data models, where “raw data” flows unidirectionally into increasingly abstract representations (critiqued by Gitelman (2013)), Latour’s chain preserves what we might call “epistemic reversibility” – the possibility of return. Computational transformations remain valid only insofar as they maintain traceable connections back to source materials, allowing historians to move between aggregate patterns and individual documents.

### 3.2. *Tracing the Hybrid Chain: Transformations in a Blended Workflow*

Latour’s conceptual framework is particularly valuable because it allows us to dissect the historical research process without falling into false dichotomies between “analog” and “digital” scholarship. Most contemporary research takes place in a blended setting, where traditional hermeneutic practices are inextricably interwoven with digital tools. Latour’s model of the chain of reference enables us to pinpoint exactly *where* and *how* digital technologies intervene in this workflow.

Rather than treating digitization as a rupture—a division between “traditional” and “digital” history—Latour’s framework reveals continuity in practices of transformation, inscription, and mobility. This perspective aligns with recent critiques of digital exceptionalism in the humanities (Hitchcock 2014; Putnam 2016), which emphasize that computational methods represent intensifications and accelerations of existing scholarly practices rather than fundamental breaks. As Putnam (Putnam 2016) demonstrates, full-text database searching systematically biases research toward sources that were already privileged—digitized first, English-language, institutionally preserved—creating what she calls “shadows” that reshape historical practice.

Yet this observation, while important, should not be mistaken for a uniquely digital phenomenon. Traditional historical practice has always been shaped by comparable selection biases: the privileging of printed editions over archival manuscripts, the authority of editorial committees determining what merits publication, and the institutional hierarchies governing access to collections. In this sense, the “shadows” Putnam identifies in digital repositories mirror long-standing patterns in which scholarly gatekeepers have determined which sources circulate widely and which remain buried in archives. What digital transformation does change, however, is the *scale* at which these biases can be both reproduced and challenged. AI-driven approaches to mass digitization and computational analysis hold the potential to democratize access to previously marginalized sources and to disrupt established canons – not by eliminating selection, but by making its mechanisms visible and contestable (Rehbein 2024).

Latour’s model makes visible that these digital interventions are not neutral supports. Every time we insert a digital link into the chain—be it a search algorithm or a text recognition model—we alter the trajectory of the reference. The aim of this article is to utilize this framework to identify these exact insertion points and to subject the implicit decisions made there (often by software developers rather than historians) to the same critical scrutiny we apply to the sources themselves. This attention to infrastructure echoes the work in Science and Technology Studies on how technical systems encode particular epistemologies and value commitments (Bowker and Star 2000). Classification systems, database schemas, and algorithmic parameters are never merely technical—they embody decisions about what counts as similar, what features matter, what can be ignored.

### 3.3. *The Data Layer as Intermediate Transformation*

Latour’s framework further illuminates the specific nature of the steps labeled “Digital Intervention” above. It helps us to understand the emergence of a “data layer”—a phenomenon often misunderstood in practice. When we apply AI-powered segmentation to cadastral maps, for instance, we are not simply “mining” pre-existing information. Instead, we are extending the chain of reference by inserting a new, highly interpretative transformation.

When we apply AI-powered segmentation to cadastral maps, for instance, we are not simply “mining” pre-existing information. Instead, we are extending the chain of reference by inserting a new, highly interpretative transformation.<sup>2</sup> The process of generating this layer can be broken down into specific operations of abstraction:

1. **Input (Visual representation):** The digitized map is treated as a matrix of pixels. At this stage, it preserves the visual complexity and the “noise” of the original archival document.
2. **Operation (Algorithmic segmentation):** An AI model predicts pixel classes (e.g., “building” vs. “background”). This is the act of *discretization*—forcing unambiguous categories onto a continuous or ambiguous reality.
3. **Output (Data layer):** The probability maps are converted into vector polygons. These are immutable mobiles: distinct, countable objects that can be moved into databases, separated entirely from the visual context of the map.

It is crucial to underscore that this generation of data is a fundamentally contingent process. The resulting dataset is never the inevitable extract of the source, but always one of many possible outcomes dependent on variable technical parameters. For instance, a shift in the optical resolution (DPI) during the initial digitization fundamentally alters the “materiality” available to the machine. A map scanned at 300 dpi may yield a generalized set of building footprints, while the same map scanned at 600 dpi might allow the algorithm to distinguish between solid walls and garden fences. Similarly, choices regarding color depth or contrast thresholds in the pre-processing stage dictate what the algorithm “sees” and what remains invisible. Thus, the data we analyze are not fixed properties of the past, but artifacts of a specific configuration of the technical apparatus. This contingency of data generation aligns with arguments in critical data studies emphasizing the performative rather than descriptive nature of datasets (Gitelman 2013; Schöch 2013). Data do not describe a pre-existing reality so much as they enact particular ways of engaging with sources. The 300 DPI scan of a cadastral map is not “more accurate” than the 600 DPI scan in any absolute sense – each resolution enacts a different reality, preserving different features, enabling different questions. The choice of resolution is thus not merely technical but fundamentally epistemological.

This operational sequence is never neutral; it is shaped by a negotiation process between historiographical intent and technical feasibility. Historians and computer scientists must jointly define the “ontology of the model”: How many nuances can be preserved? Should the algorithm distinguish between distinct types of buildings (e.g., residential vs. utility), or is a generic “structure” label sufficient? In the Computational Humanities, these epistemological questions are frequently determined by resource constraints (“pragmatic modeling”). The cost of creating ground truth (training data) and the limitations of accessible computing power often dictate the final accuracy. Thus, the resulting data layer always represents a compromise between the ideal historical question and the material reality of the project’s resources. Consequently, these decisions define the horizon of the research outcomes; they catalyze the specific form of knowledge that can be produced. Although factors like annotation budgets or algorithmic architectures may initially appear extrinsic to the intellectual process of historical inquiry, they are effectively reconfigured here. They cease to be mere logistical details and become constitutive elements of the methodological apparatus itself, dictating the granularity and the reliability of the historical narrative that follows.

What we call “data”—the polygons representing buildings, roads, forests, and fields, or the pixels classified as water—are not atomic units that existed “in” the map waiting to be extracted. Rather, these are generated through a process of discretization; they are created through a transformation that loses certain features (the aesthetic qualities of the original cartography, the variations in how

<sup>2</sup> This directly challenges what Drucker (2011) terms the “data/capta” distinction—the notion that humanities scholars should recognize their materials as “taken” (capta) rather than “given” (data). While Drucker’s intervention usefully denaturalizes data, Latour’s transformation chain framework goes further: it shows that all data—in sciences and humanities alike—are generated through transformation processes. The question is not whether data are “given” or “taken,” but whether transformation chains remain traceable and preserve features relevant to research questions.

different surveyors represented similar features) while gaining others (machine-readability, statistical analyzability, systematic comparability across map sheets). This reframing has profound implications. It dissolves the false dichotomy between “information extraction” (which implies information pre-exists in sources) and “data creation” (which implies data is invented by researchers). Instead, we recognize that all data are created through transformations, but these transformations are not arbitrary—they preserve certain features while making them mobile, comparable, and manipulable. The question is not whether data are “really there” or “constructed,” but whether the transformation chain is traceable, whether it preserves features relevant to our research questions, and whether it allows us to move back and forth between different levels of abstraction.

### 3.4. From Source to Data: The Chain of Circulating Reference

Latour’s framework allows us to dismantle the static hierarchy of the DIKW pyramid in favor of a dynamic model of circulation. The “data layer” described above is not a distinct foundational stratum upon which information and knowledge sit; rather, it represents a specific node in a bi-directional chain of transformations. This reconceptualization has precedent in information science itself: Frické (Frické 2009) demonstrates that the DIKW hierarchy lacks coherent definitions, logical structure, or empirical support. More fundamentally, the hierarchy’s linear progression contradicts actual knowledge-making practices in all fields. Rehbein (Rehbein 2021) advocates for the analytical utility of categorization, specifically challenging the frequent amalgamation of data, information, and knowledge into the singular rubric of “information.” He posits that while information constitutes a necessary condition for understanding, it is not a sufficient one. Instead, he argues for a model of multi-directional interdependence in which these three categories remain distinguishable yet functionally inseparable. Within the context of historical inquiry, this framework implies that the comprehension of any single element is fundamentally predicated upon its relationship to the others. Scientific knowledge, as Rheinberger (Rheinberger 1997) shows through laboratory ethnography, emerges through iterative cycles where “epistemic things” (objects under investigation) and “technical objects” (instruments and methods) co-evolve without clear hierarchy.

Instead of a linear ascent from raw data to wisdom, digital historical research relies on the constant capability of *reversible movement*: downstream, from the specific material source to abstract, mobile data that enables calculation and comparison across thousands of records; and upstream, back from the statistical pattern to the original pixel or archival context to validate the meaning of the data point.

This bidirectionality is the hallmark of digital scholarship. We zoom from aggregate statistics to individual map sheets. We compare segmented layers with original rasters to validate model performance. However, this movement is no longer solely under the control of the historian. It is mediated by the technical dispositive itself. The ability to “zoom back in”—to traverse the chain upstream—depends on interface designs, database architectures, and storage decisions made by computer scientists and engineers. If a data pipeline discards the link to the original image filename to save storage, the chain is broken; the reference ceases to circulate.

Therefore, the agency in this epistemological process is distributed. It is not just the historian interpreting sources, but a collaborative interplay involving the historian’s questions, the engineer’s architectural decisions, the algorithm’s pattern recognition, and the hardware’s limitations.

This reconfigures our definitions:

- **Data** are not atomic facts, but highly processed, mobile transformations optimized for manipulation.
- **Information** is not “contextualized data,” but the active maintenance of the link between the abstraction (data) and its origin (source).
- **Knowledge** implies the mastery of the entire transformation chain—understanding not just what the data signify, but how to retrace the steps of their production and how to account for the decisions of all actors (human and non-human) involved in the process.

These redefinitions resonate with Rheinberger's (Rheinberger 1997) concept of "experimental systems" in which knowledge emerges not from extracting information from objects but from constructing systems that make objects answerable to questions. For historical research, digital transformation chains constitute experimental systems—arrangements of sources, tools, protocols, and questions that co-produce historical knowledge. The historian's task is not to "extract information" from sources but to design and maintain transformation chains that preserve epistemic reversibility while enabling new forms of engagement.

### 3.5. Implications for Data-Driven Historical Research

We argue that this theoretical reframing has immediate practical implications for data-driven historical research. It suggests that the value of computational methods lies not in "extracting information" from sources but in creating new transformations that make certain features more visible, comparable, and analyzable while preserving traceability back to original sources. This aligns with arguments in digital humanities for what might be called "explicit hermeneutics" (Capurro 2010) — making visible the interpretive decisions encoded in computational workflows. Berry and Fagerjord (Berry and Fagerjord 2017) similarly argue for "critical digital humanities" that interrogates rather than naturalizes computational methods. Our contribution extends these calls by providing a concrete framework (transformation chains) for documenting and analyzing interpretive decisions.

When we use CNNs to segment cadastral maps, we are not revealing "the data in the maps" but creating a new inscription—vector layer that preserves spatial relationships and feature classifications while losing artistic style and individual cartographic choices. The methodological question is not "Is this accurate?" (in the sense of perfect correspondence), but rather: "Does this transformation preserve features relevant to our research questions? Can we trace the transformation chain? Can we move bidirectionally between segmented layers and original maps?"

Similarly, when we use transformer models to extract structured data from historical directories like the *Schematismus*, we create inscriptions (database records) that preserve hierarchical relationships and temporal patterns while losing the layout, typography, and organizational logic of the original volumes. The research value depends not on eliminating these losses but on making them explicit and on maintaining the ability to return to original sources when questions arise that require features lost in transformation.

This approach also suggests new forms of methodological rigor for digital history. Rather than pursuing impossible goals of "complete" or "objective" digitization, we should focus on:

1. **Transparency about transformations:** Explicitly documenting what is preserved and lost at each stage
2. **Traceability:** Maintaining reversible links between transformed data and source materials
3. **Multiplicity:** Creating multiple transformations of the same sources to preserve different features
4. **Bidirectionality:** Developing workflows that allow movement both toward abstraction (statistics, models) and toward specificity (individual sources, contexts)

These principles find support in recent methodological reflections on computational historiography. Guldi and Armitage (Guldi and Armitage 2014), despite their controversial advocacy for "longue durée" digital history, correctly emphasize that computational scale must be balanced with interpretive depth. More critically, Putnam (Putnam 2016) demonstrates how digitization's "shadows" – systematic biases in what gets digitized and how – require historians to maintain awareness of transformation chains, understanding not just what digital sources reveal but what they obscure. Our framework provides conceptual tools for navigating these tensions: by understanding digitization as transformation rather than reproduction, historians can critically assess both the possibilities and limitations of computational methods.

## 4. Case Study 1: The *Franziszische Kataster*

The *Franziszische Kataster* (1817–1861) represents one of the most extensive cadastral mapping projects of the long nineteenth century (Kain and Baigent 1992; Scharf 2017). Initiated by Emperor Franz I through the Land Tax Patent of 1817, this survey originally covered the Monarchy's Cisleithanian territories—a total of 300,082 square kilometers. The resulting inventory consists of 164,357 hand-drawn, colored sheets (approx. 53 × 66 cm) documenting roughly 50 million parcels across 30,000 municipalities (BEV – Bundesamt für Eich- und Vermessungswesen 2017; Feucht 2008). Present-day Austria alone is represented on 53,212 sheets (Schlögl 2008).

### 4.1. The Cadastre as Research Resource

The *Franziszische Kataster* constitutes an invaluable resource for understanding land use, settlement patterns, and property relationships at the threshold of industrialization. Its fine scale—typically 1:2,880, increasing to 1:1,440 or 1:720 in populated areas—and comprehensive coverage make it theoretically possible to reconstruct spatial and socio-economic conditions with unprecedented detail. The cadastre systematically records not only topographic features but also property boundaries, building structures, infrastructure, and detailed land use categories, down to individual trees of economic significance (Feucht 2008, pp. 7–8).

### 4.2. Georeferencing as First Transformation

Since the early 2000s, the Styrian Provincial Archive (*Steiermärkisches Landesarchiv*) has digitized and georeferenced the *Franziszische Kataster* for Styria (Baumann et al. 2025). As outlined in Section 3, this marks a crucial transformation in the chain of reference: physical sheets—already abstractions of territorial realities—are converted into raster images and aligned with modern coordinate systems (see Figure 1).



Figure 1. Scan of an original sheet of the *Franziszische Kataster*.

Georeferencing enables previously impossible systematic comparisons (Knowles 2008). With mean errors below 2.9 meters (Baumann et al. 2025), the historical cadastre can now be overlaid with LiDAR models, modern orthophotos, and contemporary cadastral data. This integration allows for precise analyses of settlement morphology, landscape transformation, and infrastructural development against the backdrop of 200 years of change.

Critically, this transforms the cadastre's epistemic status. By shifting the maps from their original Habsburg spatial framework to a modern digital inscription, content becomes mobile across temporal and technical contexts. While involving loss (materiality, texture), the process grants digital manip-

ulability and quantitative comparability. Following Latour, the chain extends: territorial reality → field survey → measurement protocols → cartographic representation → digitization → georeferencing → spatially referenced raster images. Each step makes the data more mobile and amenable to computation.

#### 4.3. AI-Supported Segmentation: Decomposing Information into Data

Building on the georeferenced foundation, we apply a UNet++ convolutional neural network (Zhou et al. 2018) to automatically segment buildings, roads, water bodies, and forests. This process transforms what traditional methodology views as documentary information into discrete, computationally manipulable data (see Figure 2).

Operating on  $512 \times 512$  pixel tiles, the model was trained on eleven cadastral municipalities in the Murtal district (province of Styria, Austria) using a combination of manual and model-assisted annotation. Despite source inconsistencies—such as varying drawing styles, color bleeding, and changing representational conventions—the model achieves high accuracy, with an overall Intersection over Union of 96.28% and F1 scores above 95% for all feature classes (Baumann et al. 2025).

Theoretically, this process is not merely “mining” pre-existing information but generating data through systematic discretization. Consider buildings: The original maps depict heterogeneous hand-drawn polygons varying in color, line quality, and architectural detail. The model reduces this cartographic complexity to a binary classification—pixel X is “building” or “not building.” These classified pixels are then converted into vector polygons with precise coordinates and measurable properties.



**Figure 2.** Georeferenced scan of a sheet of the *Franziszeische Kataster* after automatic segmentation.

This transformation involves distinct trade-offs. Lost are aesthetic features, individual surveyor styles, and subtle cues regarding construction materials or perspective. Gained are systematic comparability, quantitative manipulability, and scalability. Every building is identified using consistent criteria, turning them into countable objects amenable to GIS operations and allowing the same analytical procedure to be applied across entire territories.

#### 4.4. Scaling Analysis: From Exemplary to Comprehensive

The most significant epistemological implication of AI-supported segmentation concerns scale. Traditional research on the *Franziszeische Kataster* was necessarily selective—limited to specific villages or valleys due to the enormous labor required for manual extraction. This limitation reflected not

merely pragmatism but an epistemological foundation that privileged micro-regional analysis and hermeneutical depth over statistical coverage, often leaving the representativeness of case studies unaddressed (Hudson 2000).

AI-supported segmentation fundamentally alters this constraint. Once a model achieves adequate accuracy, it can be applied systematically across entire territories. For Styria, approximately 60% of the provincial territory—where georeferencing is complete—can now be analyzed regarding building distribution, road networks, and land use for the 1817–1861 period (Baumann et al. 2025). This enables research into historical provincial settlement density, correlations between forest coverage and topography, and infrastructure patterns that were previously inaccessible.

Moreover, comprehensive coverage allows for unprecedented temporal comparison. By overlaying historical layers with contemporary spatial data (e.g., modern building footprints or LiDAR-derived forest cover), transformation patterns emerge that remain invisible to selective case studies. In this sense, the integrated research environment functions as a “historian’s microscope” (Graham et al. 2022): a computational framework allowing researchers to toggle between the large-scale detection of structural patterns and the close inspection of individual cases. This approach aligns with calls for “scalable reading” (Manovich 2020) or “blended reading” (Stulpe and Lemke 2016), combining systematic breadth with local depth to investigate settlement expansion, infrastructure persistence, and environmental change.

#### 4.5. Insights into Historical Source Production

An unexpected benefit of AI-supported segmentation lies in revealing the internal logics and construction processes of the source material. During model training, segmentation errors frequently highlighted discontinuities at map sheet boundaries, where forest edges, roads, or streams were represented differently by adjacent surveyors. These inconsistencies reveal that the cadastre was not a uniform, standardized representation, but a composite of individual judgments and local adaptations—aligning with STS findings on the interpretive flexibility inherent in standardization projects (Bowker and Star 2000).

These stylistic variations open a further analytical dimension: computational style analysis. Systematic differences in line weight, color palette, and feature representation constitute “cartographic fingerprints” that may enable authorship attribution for individual surveyors. Beyond attribution, classification discrepancies often reflect material ambiguities in the built environment. For instance, inconsistencies in distinguishing wood from stone structures may capture vernacular practices that resisted the cadastre’s classificatory grid. Reading these discrepancies not as errors but as evidence transforms the map into a dynamic source for material culture.

Buildings provide particularly revealing examples. Where the model struggled to distinguish structures in dense clusters, it often mirrored difficulties faced by the original cartographers. Instances of mid-sheet recoloring suggest changes in surveying teams during the production process—insights into the workflow difficult to gain through traditional close reading. Similarly, overlaying segmented buildings with modern footprints yields empirical data on measurement precision, demonstrating high accuracy in flat terrain but noticeable triangulation difficulties in mountainous or forested regions.

#### 4.6. Theoretical Analysis: Data, Information, and the Transformation Chain

How do we theoretically conceptualize AI-supported segmentation? Returning to our framework (Section 3), this process constitutes a new transformation in Latour’s chain of circulating reference. The specific chain for our cadastral research proceeds as follows:

- **Territorial reality:** The physical landscape (villages, fields, forests) observed by surveyors.
- **Field survey:** Measurement and field book recording. *Loss:* Continuous landscape becomes discrete points. *Gain:* Quantifiable, transportable data.
- **Cartographic production:** Translation into graphic representation. *Loss:* 3D terrain becomes 2D map; precision reduced to resolution. *Gain:* Visual comprehensibility and preservation.

- **Digitization & Georeferencing:** Scanning and coordinate alignment. *Loss:* Materiality and original spatial framework. *Gain:* Digital reproducibility and GIS integration.
- **AI segmentation:** Pixel classification via CNN. *Loss:* Cartographic style, aesthetics, and ambiguity. *Gain:* Systematic classification, scalability, and statistical analyzability.
- **Vector conversion:** Creation of geometric polygons. *Loss:* Pixel-level granularity. *Gain:* Topological relationships and precise measurements.
- **Spatial analysis:** GIS operations. *Loss:* Individual feature uniqueness. *Gain:* Pattern recognition and hypothesis testing.

This chain exemplifies Latour's "preservation through displacement" (Latour 1999a, pp. 69-70). It demonstrates that "data" are not intrinsic to the maps but depend on the technical parameters of transformation. Furthermore, the chain is bidirectional. Reference circulates from territory to database and back; statistical patterns in the data can prompt close reading of specific map sheets, akin to zooming in with a "macroscope." This validates the concept of "distant reading" (Moretti 2013) not as a replacement for, but a complement to, hermeneutic engagement.

Crucially, this framework dissolves the dichotomy between "information extraction" and "data creation." AI segmentation is neither pure extraction (finding pre-existing data) nor arbitrary creation. It is a systematic transformation that translates spatial relationships into computational forms. The success of this pipeline rested on aligning historical necessity with algorithmic logic, illustrated by three technical configurations:

1. **Resolution as Context (The Receptive Field):** We downsampled images not merely for efficiency, but to expand the CNN's "receptive field." This allowed the network to perceive sufficient spatial context to distinguish complex entities like buildings, rather than just analyzing texture at full resolution.
2. **Contextual Redundancy (Overlapping Inference):** To counter edge errors, we implemented a "quadruple inference" strategy. Each map section was classified four times with shifting input windows. The final result is a consensus, mimicking a rigorous source critique where observations are verified from multiple angles.
3. **The Ontology of "Streets" (Hermeneutics of Classes):** Initial segmentation of streets failed due to heterogeneity. Historical analysis revealed that maps depicted "common spaces" (*Gemeinflächen*) rather than modern infrastructure. Adjusting the target class from a visual/modern category to a historical/legal one allowed the algorithm to converge.

These examples demonstrate that the resulting data are actively constructed through a dialogue between domain knowledge and machine vision. This iterative alignment parallels Rheinberger's (Rheinberger 1997) co-evolution of "epistemic things" and "technical objects," where the research question and the instrument develop in unison.

## 5. Case Study 2: Structuring the Social Body – The Hof- und Staatsschematismus

While the *Franzische Kataster* organized the physical space of the Monarchy, our second case study turns to the organization of its social and administrative body. The *Hof- und Staatsschematismus* serves as the primary source: a directory of individuals employed at the Habsburg court and state, offering a longitudinal dataset of titles, offices, and salaries from 1702 to 1918 (Noflatscher 2004).

Comprising 153 editions and over 150,000 pages available as PDF scans (Österreichische Nationalbibliothek 2011), this series requires a methodological shift from image segmentation to layout analysis. The transformation from source to data here is not about separating foreground from background, but about decoding the visual grammar of bureaucracy to understand structural hierarchies over time (see Figure 3).



636 Kärnten. — Landwehr-Behörden. — Kirchliche Behörden.		
<b>Gill</b>	Kallab Otto, $\Phi$ C, J. Dr., Ob.-Berg-Kmar.	<b>In Dalmatien.</b> Cupani Igor, $\Phi$ & $\Phi$ C, Berg.
<b>Gratz</b>	Klasi Rudolf, $\Phi$ & $\Phi$ C, J. Dr., Ob. Berg.	<b>K. K. Berg-Verwaltung in Raibl.</b> (Bleiers-, Zinkblende- und Galmei-Bergbau und Blei-Schmelzhütte.) Schnitzler Alois, FJO-R., Bst., Berg. R. Gussmühl, Ob. Ob. Berg-Verw.
<b>Looben</b>	Bussen Felix, $\Phi$ C, J. Dr., Ob. Berg-Kmar. Santo-Passio Otto, $\Phi$ C, Ob. Berg-Kmar.	<b>K. K. Forst- und Domänen-Verwaltungen.</b> Die im Herzogthum Kärnten bestehende Forst- und Domänen-Verwaltung untersteht der k. k. Forst- und Domänen-Direktion in Görz, und kommt in der Abtheilung „Küstenland“ vor. <b>K. K. Forsttechnische Abteilung für Wildbachverbauung.</b> Sektionschef für Kärnten, Krain, und das Küstenland: Innes in Füllsch. Wenediker Ferdinand, FJO-R., GVK, Ob. Forst-R., Sektions-Adjunkt. Schnitzler Alois, $\Phi$ & $\Phi$ C, Ob. Berg-Kmar. Maltweger Silvester, $\Phi$ & $\Phi$ C, Forst-Inspekt., Krain. Stain Anton, $\Phi$ & $\Phi$ C, I. Kl.
<b>Laibach</b>	Strgar Vinzenz, $\Phi$ & $\Phi$ C, Ob. Berg-Kmar.	
<b>Hall</b>	Obsthaas Karl, $\Phi$ & $\Phi$ C, J. Dr., Ob. Berg-Kmar.	
<b>VII. Landwehr-Behörden.</b>		
<b>Landwehr-Truppen.</b>		
Die Kommanden der aus Kärnten ergänzten Landwehr-Truppenkörper unterstehen dem k. k. Landwehr-Kommando in Graz und kommen in der Abtheilung „Steiermark“ vor.		
<b>K. K. Landes-Gendarmarie-Kommando Nr. 14 für Kärnten in Klagenfurt.</b>		
<b>Witsch Ignaz</b> , FJO-R., MVN., Dr., $\Phi$ & $\Phi$ , Oberstlt.	<b>Beinhornführer</b> Urbik Alois, Dr., $\Phi$ & $\Phi$ , Hptm., Richter.	<b>Abtheilungs-Kommandanten</b> Goswametzner, Dr., $\Phi$ & $\Phi$ , Rittm., in Klagenfurt. Goswametzner Alfred, Ob. St., Oberlt., in Klagenfurt. Zwörzack Stanislaus, $\Phi$ & $\Phi$ , Rittm., in Villach. Zwörzack Jrodler, $\Phi$ , Leut., des Train-Regt. Nr. 9 (angelt.), zweiter Offizier in Villach.
<b>May Oskar</b> , $\Phi$ & $\Phi$ , Leut.		
<b>Jansschitz Richard</b> , $\Phi$ , Oberlt.		
<b>C. Kirchliche Behörden.</b>		
<b>Bistum Gurk.</b>		
(Sitz des Fürst-Bischofs: Klagenfurt.)		
<b>Fürst-Bischof</b>		
Seine fürstbischöflichen Gnaden Erzellenz Josef Kahn, FJO-GK., Th. Dr., Sr. päpstl. Heil. Haus-Prälat u. Thron-Assist., röm. Graf, Gh. R., Mitgl. des Herren-u. des kärntn. Landt.		
<b>Dom-Kapitel</b>	<b>Fürstbischöfliches Konseistorium</b>	<b>Dassler Johann</b> , S. J., Prof. der Kirchenrechtl. u. Ph. Dr., Dozent des Kirchenrechtes u. der christl. Archäologie in Rom.
<b>Dom-Probst</b> Müller Anton, FJO-Kl., Th. Dr., Konsist. R., Prosyn. Exam.	<b>Präsident</b> Der jeweilige Fürst-Bischof von Gurk.	<b>Weth Wilhelm</b> , S. J., Prof. der Pastoraltheologie u. Pädagogik.
<b>Dom-Dechant</b> Bitter Guido, Konsist. R., Prosyn. Exam.	<b>Stiftsbischöfliche Hof-Kaplanen</b> Gutt Johann, Th. & Ph. Dr., f. b. geistl. R., Prosyn. Exam., Kanonikus des Kollegiat-Kapitels Straburg, Doz. des Kirchenrechtes u. der Kunstgeschichte.	<b>Barter Johann</b> , S. J., Prof. der spezialtheol.
<b>Dom-Scholaster</b> Wappals Ferdinand, Konsist. R., Prosyn. Exam., inf. Propst von Straburg.	<b>Erlich Lambert</b> , Th. Dr., Kanonikus des Kollegiat-Kapitels Straburg.	<b>Spadil Theophil</b> , S. J., Dozent der Fundamentaldogmatik u. der theol. Philo- studium N. B.
<b>Dom-Kyriakon</b> Sommer Josef, Th. Dr., Konsist. R., Prosyn. Exam.	<b>Theologische Bistums-Lehranstalt in Klagenfurt.</b> Direktor. Der jeweilige Fürst-Bischof von Gurk.	<b>Müller Gregor</b> , S. J., Prof. des Bibeldienstes N. B.
<b>Größer Mathias</b> , EKO-R. S., FJO-R., Prosyn. Exam.	<b>Direktor-Sachverständiger</b> Sommer Josef, Th. Dr., Dom-Kapitular.	<b>Fürstbischöfliches Priester-Seminarium in Klagenfurt.</b> Direktor. Widmer Augustin, S. J.
<b>Widmer Johann</b> , Konsist. R., Prosyn. Exam., f. b. Kapler.	<b>Professoren</b> Goswametzner Ignaz, S. J., Prof. der Moral.	<b>Fürstbischöfliches Knaben-Seminar (Klagenfurt).</b> Direktor. Pirker Johann, Dom-Kapitular.
<b>Pirker Johann</b> , Konsist. R., Prosyn. Exam., Dir. des Knaben-Seminars.	<b>Reinsprecht Ludwig</b> , Th. Dr., Prof. des Bibeldienstes A. B. u. der unml. Dialekte.	
<b>Kovar Martin</b> , Konsist. R., Dom-Verw., Prosyn. Exam.		

Figure 4. Schematismus-page after layout detection.

### 5.3. Defining the Scope: A Strategic Reduction

Applying the logic of “contingent data generation” (Section 3), we restricted the project’s temporal scope to handle the source’s heterogeneity. Automated layout analysis relies on recurring patterns; a model trained on 1750s single-columns fails on 1900s three-column tables. Consequently, the project focused on the 1876–1918 series, representing the longest continuous period of consolidated layout structure (Fleischhacker et al. 2025)<sup>3</sup>.

By accepting a reduction in historical breadth to achieve extraction accuracy, we underscore a crucial constraint: the “digital dataset” is inextricably linked to the visual uniformity of the analog source. Where the source’s visual logic breaks down, the data pipeline effectively halts.

### 5.4. The Synthetic Bridge: Manufacturing Data to Read Data

A fundamental bottleneck in historical Deep Learning is the scarcity of “Ground Truth.” Training a model to recognize the complex layout of the *Schematismus* requires thousands of annotated pages—a task exceeding project resources.

Our solution involved a significant shift: instead of annotating existing sources, we manufactured them (see Figure 5). Using LuaTeX, we generated 3,766 pages of synthetic training data (Fleischhacker et al. 2025). While content was randomized, these pages strictly adhered to historical layout rules (fonts, indentations). This adapts synthetic data strategies (Journet et al. 2017) to specific administrative conventions.

This approach inverts the hermeneutic process. The algorithm learns from a formalized model created by researchers, not the source itself. The “digital eye” is trained on a Platonic ideal—perfectly

<sup>3</sup> Note that the workflow in (Fleischhacker et al. 2025) is a developmental milestone and differs from the current project workflow.



1. **The Ontology of the “Curly” (Handling Hierarchy):** The curly bracket (*Geschweifte Klammer*) is a pervasive visual element in the *Schematismus*, grouping officials under shared institutional headings. Technically, however, it is merely a graphical symbol. To preserve its hierarchical function for computational processing, we defined a dedicated “Curly” class within a three-tiered layout ontology: at the lowest level, *Paragraphs* capture individual text blocks; *Headers* operate across all three levels to mark organizational units; and *Curly*, as a container class at the intermediate level, groups *Paragraphs* and *Headers* into semantically coherent units. This decision elevated a graphical punctuation mark to a structural entity, encoding the insight that in bureaucratic documents, visual syntax carries as much institutional meaning as the text itself.
2. **The Simulation of Variance (Synthetic Data Strategy):** Using synthetic data required encoding our understanding into generator code (e.g., distinguishing indentation types). Consequently, the algorithm’s subsequent “discovery” of structures is tautological: it finds what it was told to look for. However, the strictly rule-based synthetic datasets are augmented during training with a comparatively small proportion (between 10 and 20%) of annotated original pages, which proves sufficient to enable the model to handle the diversity and heterogeneity of the historical source material with confidence. The extraction’s “objectivity” is thus bound by the historian’s initial formalization of layout rules, yet tempered by the model’s empirical exposure to the irregularities of the actual documents – a productive tension between idealized structure and archival reality.
3. **Resolution as an Epistemological Dialectic:** Comparing the case studies reveals a dialectic of resolution. We downsampled maps (Section 4) to capture context (“forest over trees”), but upsampled text (Section 5) to capture structural “fences” like separators. There is no “correct” digital resolution; technical parameters depend on the hermeneutic goal—land use requires blur, administrative hierarchy requires sharpness. Moreover, the choice of model architecture itself imposes additional constraints. In our preliminary experiments with YOLO for layout detection, we found that it operated optimally within different parameter ranges than the Faster R-CNN architecture ultimately adopted, particularly regarding input dimensions and aspect ratios. Depending on the model, it can prove worthwhile to invest considerable effort in pre-processing—adjusting crop dimensions to achieve different aspect ratios, or up- and downscaling resolution—before the source material ever enters the detection pipeline. The “correct” resolution is thus doubly contingent: determined not only by the historical question but also by the computational affordances of the chosen architecture.

The resulting dataset is not a raw copy but a structured interpretation, capturing the 19th-century “bureaucratic gaze.” This exemplifies the transformation of “capta” into data (Drucker 2011), reinforcing that data is bound to its generation context (Gitelman 2013). Strictly speaking, the data did not exist until the algorithm—trained on our premises—generated it. Using this dataset therefore requires a specific “digital source criticism” (Fickers 2012): treating the pipeline not as a neutral conduit, but as a transformative layer shaping historical evidence.

## 6. Case Study 3: The Datafication of everyday Resistance in Nazi Germany – a Digital “Upcycling” of an Archive Inventory about 10,000 + criminal proceedings of the Munich Special Court

Special Courts in 20th century German History were no invention of the National Socialists, but had been introduced in 1932 by the first German democracy – at the time not least as an instrument against violent attacks on the Republic, that means against Nazis in particular. However, the quality and quantity with that Nazi regime implemented the special criminal jurisdiction in its system of repression and persecution right from the beginning of the “seizure of power” illustrates the multi-layered relationship between historically grown institutions of the modern, democratic constitutional and legal state on the one hand and Nazi ideology on the other. Special Courts primarily dealt with “offenses” that had been criminalized by the Nazi regime for ideological reasons, especially during the war, but also particularly serious crimes such as murder.

The original court records of more than 10,000 cases heard by the Munich Special Court between 1933 and 1945 are almost entirely preserved in the Bavarian State Archives. In the 1970s, in the context of reappraising the Nazi era and the historiographical turn of the “modern social history,” key data on the defendants and their crimes was systematically recorded by hand from all these files in a structured form ([Generaldirektion der Staatlichen Archive Bayerns 1977](#)). Each of the historical court records was thus converted into a regest with 12 metadata entries, characteristics such as national and religious backgrounds, party affiliation, or profession.

Since 2021, this “secondary source” from the 1970s was digitized as part of a collaborative project between the Passau Department of Computational Humanities and the Bavarian State Archives, and all of the regesta it contains were automatically transferred to a database using OCR and parsing. While research into resistance against National Socialism has traditionally focused on significant individual resistance fighters, computer-assisted analyses of this data set can provide new macroscopic insights into repression and resistance as phenomena affecting the society as a whole ([Ernst et al. 2023](#); [Gerstmeier et al. 2022](#)).

Even though the compilation work carried out by professional historians in the 1970s was meticulous and thus offers the possibility of accessing the content on a systematic level, there are nevertheless some inconsistencies and historical desiderata. The sometimes poor print quality of the archive inventory has a limiting effect on OCR processing. In addition, the respective legal bases were only recorded for proceedings since 1939. Metadata on the prosecutors and judges involved were not even part of the data model underlying the archive inventory, which is symptomatic of the nature of the state-funded reappraisal of Nazi injustice in the 1970s. Quite a few of the lawyers who became perpetrators during the “Third Reich” were still alive thirty years after the end of the war or were even still in office. These inconsistencies pose a challenge for digital “upcycling” in the sense of automated processes.

## 7. Conclusions

This paper began with the assertion that the digitization of historical sources is not a mechanical reproduction, but a complex process of translation. By comparing three distinct extraction pipelines — the AI-supported segmentation of the *Franziszeischer Kataster* (cadastral maps), the layout analysis of the *Hof- und Staatsschematismus* (Administrative Directories), and the datafication of the Munich Special Court archive inventory — we have demonstrated that the transformation from ‘source’ to ‘data’ requires a continuous negotiation between historical logic and algorithmic constraints. Each case study reveals a different facet of this negotiation: the *Kataster* foregrounds the interplay between image resolution and spatial semantics; the *Schematismus* exposes how visual layout encodes institutional hierarchies that must be formalized before they can be computationally parsed; and the Special Court inventory illustrates how the transformation chain extends across historiographical generations, as a secondary source produced in the 1970s — itself already an interpretive reduction of the original court records — becomes the object of a further digital ‘upcycling’ whose possibilities and limitations are shaped by the epistemological choices of its earlier creators.

### 7.1. Parameters as a Transdisciplinary Negotiation

Our findings challenge the notion of the “black box” in Digital Humanities. Across all three case studies, the quality of the output depended not on the raw power of the algorithms, but on specific manual adjustments — such as the deliberate downsampling of maps, the generation of synthetic layout templates, or the parsing strategies adapted to the idiosyncrasies of a 1970s print publication — that functioned as hermeneutic interventions.

This observation bears a crucial consequence for the practice of digital history: setting these parameters correctly requires a fundamental transdisciplinary negotiation. It is insufficient for historians to merely deliver data and for computer scientists to merely process it. Rather, both parties are forced to acquire a functional literacy in the other’s discipline. This resonates with Hitchcock’s call for historians to ‘confront the digital’ by interrogating the underlying code and algorithms with the same

rigor usually applied to archival sources (Hitchcock 2013). The historian must grasp the concept of the “receptive field” to understand why the machine misinterprets a map, while the engineer must understand the legal concept of “public space” to understand why a street is not just a gap between buildings. The code, therefore, becomes a trading zone where the premises of historical critique and engineering logic must be explicitly aligned – what we might term “collaborative hermeneutics” in the spirit of recent calls for interdisciplinary digital humanities practice (Berry and Fagerjord 2017).

### 7.2. *The Machine as a Stress Test for Source Quality*

Furthermore, the training process generated a “hermeneutic feedback loop” that offered deep insights into the sources themselves. The resistance of the machine—its inability to classify certain segments—often pointed not to algorithmic failure, but to inconsistencies within the historical material. In the case of the Cadastre, the model’s error maps effectively visualized the “cracks” in the surveyor’s logic; in the *Schematismus*, the layout analysis revealed shifting editorial standards that a human reader might have glossed over; in the Special Court inventory, the automated processing exposed inconsistencies in the 1970s compilation — from varying print quality affecting OCR accuracy to the systematic absence of metadata on prosecutors and judges, a lacuna that reflects not archival negligence but the political constraints governing the state-funded reappraisal of Nazi injustice at the time. In this capacity, the Computer Vision model acts as a stress test for the archive. It allows us to assess the internal consistency and quality of the original source material with a precision that the naked eye could rarely achieve. The “artificial” gaze thus sharpens the “historical” gaze, revealing the fragility of the bureaucratic order the sources claim to represent.

### 7.3. *Scaling the Hermeneutic Circle*

Ultimately, the value of this “operationalized hermeneutics” lies in its scalability. Traditional historical scholarship is often trapped in a tension between the deep analysis of individual cases (micro-history) and the generalization of systemic trends (macro-history). While (Guldi and Armitage 2014) famously called for a return to the *longue durée* to address this, they offered little guidance on how to technically achieve it with fragmented archives. The methods presented here offer a bridge across this divide. By training models that encode expert domain knowledge—specifically tailored to the logic of the source—we can process entire archives with a consistency that no human reader could maintain. The Special Court case illustrates this most vividly: where research on resistance against National Socialism has traditionally centered on prominent individual figures, the comprehensive datafication of over 10,000 criminal proceedings enables a macroscopic perspective on repression and everyday resistance as societal phenomena — precisely the shift from exemplary to comprehensive analysis that our theoretical framework demands. This does not replace the historian’s judgment but scales it. Ideally, this leads to a “statistically secured hermeneutics,” where the interpretation of the past is supported by the comprehensive analysis of its material traces.

**Author Contributions:** Conceptualization and methodology Wolfgang Göderle and Malte Rehbein; writing – original draft preparation, Wolfgang Göderle; writing – review and editing, Malte Rehbein and Markus Gerstmeier; visualization, X.X.; project administration, Wolfgang Göderle and Malte Rehbein; funding acquisition, Wolfgang Göderle and Malte Rehbein. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Austrian Science Fund (grant numbers 10.55776/PAT1763723 and 10.55776/FG3100) as well as MALTE? MARKUS?

**Acknowledgments:** The principal project contributors for Sections 4 and 5 are Jörg Baumann and David Fleischhacker; significant conceptual contributions were further provided by Roman Kern. Figures 1, 2, 4, and 5 were produced by Jörg Baumann and David Fleischhacker.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

OCR	Optical Character Recognition
HTR	Handwritten Text Recognition
AI	Artificial Intelligence
CNN	Convolutional Neural Network
GIS	Geographic Information System/Science
DPI	Dots Per Inch
DIKW	Data-Information-Knowledge-Wisdom
IoU	Intersection over Union
R-CNN	Region-based Convolutional Neural Network
PDF	Portable Document Format
LiDAR	Light Detection and Ranging
STS	Science and Technology Studies

## References

- Antenhofer, Christina, Christoph Kühberger, and Arno Strohmeyer (Eds.). 2024. *Digital Humanities in den Geschichtswissenschaften*, Volume 6116 of *UTB*. Wien: UTB/Böhlau. <https://doi.org/10.36198/9783838561165>.
- Bauer, Volker. 2002. Territoriale Amtskalender und Amtshandbücher im Alten Reich. Bilanz und Perspektiven der Forschung. *Daphnis* 31(1–2), 217–252.
- Baumann, Jörg, David Fleischhacker, Wolfgang Goederle, Nicole Kamp, and Roman Kern. 2025. Automated segmentation of historical cadastral maps: A machine learning approach to the Franziszeischer Kataster. In *AGIT Conference*.
- Berry, David M. and Anders Fagerjord. 2017. *Digital Humanities: Knowledge and Critique in a Digital Age*. Cambridge: Polity Press.
- BEV – Bundesamt für Eich- und Vermessungswesen (Ed.). 2017. *200 Jahre Kataster. Österreichisches Kulturgut 1817–2017*. Wien: BEV – Bundesamt für Eich- und Vermessungswesen.
- Bloch, Marc. 1953. *The Historian's Craft*. New York: Knopf.
- Bloor, David. 1999. Anti-latour. *Studies in History and Philosophy of Science Part A* 30(1), 81–112. [https://doi.org/10.1016/S0039-3681\(98\)00038-7](https://doi.org/10.1016/S0039-3681(98)00038-7).
- Boonstra, Onno, Leen Breure, and Peter Doorn. 2004. *Past, Present and Future of Historical Information Science*. Amsterdam: NIWI-KNAW.
- Boshof, Egon, Kurt Düwell, and Hans Kloft. 1973. *Grundlagen des Studiums der Geschichte: Eine Einführung* (1 ed.). Cologne: Böhlau.
- Bowker, Geoffrey C. and Susan Leigh Star. 2000. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.
- Brandt, Ahasver von. 2012. *Werkzeug des Historikers: Eine Einführung in die Historischen Hilfswissenschaften* (18 ed.). Stuttgart: Kohlhammer.
- Capurro, Rafael. 2010. Digital hermeneutics: An outline. *AI & Society* 35(1), 35–42. <https://doi.org/10.1007/s00146-009-0255-9>.
- Collingwood, R. G. 1946. *The Idea of History*. Oxford: Oxford University Press.
- Daston, Lorraine and Peter Galison. 2007. *Objectivity*. New York: Zone Books.
- Droysen, Johann Gustav. 1868. *Historik: Vorlesungen über Enzyklopädie und Methodologie der Geschichte*. Munich: Oldenbourg.
- Drucker, Johanna. 2011. Humanities approaches to graphical display. *Digital Humanities Quarterly* 5(1).
- Döring, Karoline Dominika, Stefan Haas, Mareike König, and Jörg Wettlaufer (Eds.). 2022. *Digital History: Konzepte, Methoden und Kritiken Digitaler Geschichtswissenschaft*. Berlin: De Gruyter Oldenbourg. <https://doi.org/10.1515/9783110757101>.
- Eberle, Oliver, Jochen Büttner, Hassan El-Hajj, Grégoire Montavon, Klaus-Robert Müller, and Matteo Valleriani. 2024. Historical insights at scale: A corpus-wide machine learning analysis of early modern astronomic tables. *Science Advances* 10(43), ead1719. <https://doi.org/10.1126/sciadv.ad1719>.

- Ernst, Marlene, Sebastian Gassner, Markus Gerstmeier, and Malte Rehbein. 2023. Categorising legal records – deductive, pragmatic, and computational strategies. *Digital Humanities Quarterly* 17(3).
- Feucht, Rainer. 2008, 3. Flächenangaben im österreichischen Kataster. Diplomarbeit, Technische Universität Wien. Institut für Geoinformation und Kartografie.
- Fickers, Andreas. 2012. Towards a new digital historicism? doing history in the age of abundance. *Journal of European Television History and Culture* 1(1), 19–26. <https://doi.org/10.18146/2213-0969.2012.jethc004>.
- Fickers, Andreas, Juliane Tatarinov, and Tim van der Heijden. 2022. Digital history and hermeneutics – between theory and practice: An introduction. In A. Fickers and J. Tatarinov (Eds.), *Digital History and Hermeneutics: Between Theory and Practice*, Volume 2 of *Studies in Digital History and Hermeneutics*, pp. 1–20. Berlin: De Gruyter Oldenbourg. <https://doi.org/10.1515/9783110723991-001>.
- Fickers, Andreas and Gerben Zaagsma. 2022. "; towards a digital hermeneutics?". In G. Balbi, N. Ribeiro, V. Schafer, and C. Schwarzenegger (Eds.), *Digital Roots: Historicizing Media and Communication Concepts of the Digital Age*. Berlin: De Gruyter. <https://doi.org/10.1515/9783110740202>.
- Finzsch, Norbert. 2004. Latour und die historische forschung. *Historische Anthropologie* 12(2), 285–303.
- Fleischhacker, David, Roman Kern, and Wolfgang Göderle. 2025. Enhancing OCR in historical documents with complex layouts through machine learning. *International Journal on Digital Libraries* 26(1). <https://doi.org/10.1007/s00799-024-00428-0>.
- Fogel, Robert William and Stanley L. Engerman. 1974. *Time on the Cross: The Economics of American Negro Slavery*. Boston: Little, Brown.
- Frické, Martin. 2009. The knowledge pyramid: A critique of the DIKW hierarchy. *Journal of Information Science* 35(2), 131–142. <https://doi.org/10.1177/0165551508094050>.
- Generaldirektion der Staatlichen Archive Bayerns (Ed.). 1975–1977. *Widerstand und Verfolgung in Bayern 1933–1945. Hilfsmittel. Archivinventare, Bd. 3: Sondergericht München, 7 Teile*. Munich / Regensburg.
- Gerstmeier, Markus, Simon Donig, Sebastian Gassner, and Malte Rehbein. 2022. Die Archivinventare zum Sondergericht München (1933–1945) digital. Quellenwert – Verdatung – Erkenntnisperspektiven. *Archivalische Zeitschrift* 99(1), 215–251.
- Gitelman, Lisa (Ed.). 2013. *Raw Data Is an Oxymoron*. Cambridge, MA: MIT Press.
- Goederle, Wolfgang. 2015. Die räumliche Matrix des modernen Staates: Die Volkszählung des Jahres 1869 im Habsburgerreich im Lichte von Latours zirkulierender Referenz. *Schweizerische Zeitschrift für Geschichte – Revue Suisse d'histoire* 65(3), 414–427.
- Graham, Shawn, Ian Milligan, Scott B. Weingart, and Kim Martin. 2022. *Exploring Big Historical Data: The Historian's Macroscope* (2 ed.). New Jersey: Imperial College Press.
- Guldi, Jo and David Armitage. 2014. The history manifesto: A critique. *Perspectives on History*.
- Hiltmann, Torsten. 2024. (epistemologische) Grundlagen der Anwendung digitaler Methoden in den Geschichtswissenschaften. In C. Antenhofer, C. Kühberger, and A. Strohmeier (Eds.), *Digital Humanities in den Geschichtswissenschaften*, Volume 6116 of UTB, pp. 43–59. Wien: UTB/Böhlau. <https://doi.org/10.36198/9783838561165-44-61>.
- Hitchcock, Tim. 2013. Confronting the digital: Or how academic history writing lost the plot. *Cultural and Social History* 10(1), 9–23. <https://doi.org/10.2752/147800413X13515292098070>.
- Hitchcock, Tim. 2014. Confronting the digital: Or how academic history writing lost the plot. *Cultural and Social History* 10(1), 9–23. <https://doi.org/10.2752/147800413X13515292098070>.
- Howell, Martha and Walter Prevenier. 2001. *From Reliable Sources: An Introduction to Historical Methods*. Ithaca, NY: Cornell University Press.
- Hudson, Pat. 2000. *History by Numbers: An Introduction to Quantitative Approaches*. London: Arnold.
- Jannidis, Fotis, Hubertus Kohle, and Malte Rehbein (Eds.). 2017. *Digital Humanities: Eine Einführung*. Stuttgart: J.B. Metzler. <https://doi.org/10.1007/978-3-476-05446-3>.
- Journet, Nicholas, Muriel Visani, Boris Mansencal, Kieu Van-Phan, and Antoine Billy. 2017. DocCreator: A new software for creating synthetic ground-truthed document images. *Journal of Imaging* 3(4), 62. <https://doi.org/10.3390/jimaging3040062>.
- Kain, Roger J. P. and Elizabeth Baigent. 1992. *The Cadastral Map in the Service of the State: A History of Property Mapping*. Chicago: University of Chicago Press.
- Kiessling, Benjamin. 2019. Kraken - an universal text recognizer for the humanities. In *Digital Humanities Conference 2019*, Utrecht.
- Knowles, Anne Kelly (Ed.). 2008. *Placing History: How Maps, Spatial Data, and GIS Are Changing Historical Scholarship*. Redlands, CA: ESRI Press.

- Langlois, Charles-Victor and Charles Seignobos. 1898. *Introduction aux études historiques*. Paris: Hachette.
- Latour, Bruno. 1986. Visualization and cognition: Drawing things together. In H. Kuklick (Ed.), *Knowledge and Society: Studies in the Sociology of Culture Past and Present*, Volume 6, pp. 1–40. Greenwich, CT: JAI Press.
- Latour, Bruno. 1999a. Circulating reference: Sampling the soil in the Amazon forest. In *Pandora's Hope: Essays on the Reality of Science Studies*, Chapter 2, pp. 24–79. Cambridge, MA: Harvard University Press.
- Latour, Bruno. 1999b. *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press.
- Manovich, Lev. 2020. *Cultural Analytics*. Cambridge, MA: MIT Press.
- Moretti, Franco. 2013. *Distant Reading*. London: Verso.
- Muehlberger, Guenter et al. 2019. Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study. *Journal of Documentation* 75(5), 954–976. <https://doi.org/10.1108/JD-07-2018-0114>.
- Noflatscher, Heinz. 2004. "Ordonnances de l'hôtel", Hofstaatsverzeichnisse, Hof- und Staatskalender. In J. Pauser, M. Scheutz, and T. Winkelbauer (Eds.), *Quellenkunde der Habsburgermonarchie (16.–18. Jahrhundert)*. Ein exemplarisches Handbuch, pp. 59–75. Wien and München: R. Oldenbourg Verlag.
- Österreichische Nationalbibliothek. 2011. ALEX – Historische Rechts- und Gesetzestexte: Staatshandbuch. Digitale Sammlung. Digitalisierte Ausgaben des Hof- und Staatsschematismus (1702–1918).
- Putnam, Lara. 2016. The transnational and the text-searchable: Digitized sources and the shadows they cast. *American Historical Review* 121(2), 377–402. <https://doi.org/10.1093/ahr/121.2.377>.
- Rehbein, Malte. 2021. A responsible knowledge society within a colourful world. *Philosophisches Jahrbuch* 128(1), 156–168. On epistemological responsibility in digital humanities and knowledge production.
- Rehbein, Malte. 2024. Künstliche Intelligenz und Datenmobilisierung zwischen Geschichtswissenschaft und Archiv. über Forschung in der digitalen Transformation und die Notwendigkeit des Umdenkens. *Archivalische Zeitschrift* 101, 11–27. <https://doi.org/10.17613/5m74-sk48>.
- Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. 2017. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39(6), 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>.
- Rheinberger, Hans-Jörg. 1997. *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube*. Stanford, CA: Stanford University Press.
- Rowley, Jennifer. 2007. The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science* 33(2), 163–180. <https://doi.org/10.1177/0165551506070706>.
- Scharf, Kurt. 2017. Kataster und Grundbuch im Kaisertum österreich. Ausgangssituation und Entwicklung bis 1866. In BEV – Bundesamt für Eich- und Vermessungswesen (Ed.), *200 Jahre Kataster. Österreichisches Kulturgut 1817–2017*, pp. 37–51. Wien: BEV – Bundesamt für Eich- und Vermessungswesen.
- Schlögl, Michaela. 2008, 3. Ein digitales Abbild Kakaniens. *Die Presse*.
- Schöch, Christof. 2013. Big? smart? clean? messy? data in the humanities. *Journal of Digital Humanities* 2(3), 2–13.
- Siebold, Anna and Matteo Valleriani. 2022. Digital perspectives in history. *Histories* 2(2), 170–177. <https://doi.org/10.3390/histories2020013>.
- Spiegel, Gabrielle M. 2005. Revising the past/revisiting the present: How change happens in historiography. *History and Theory* 44(4), 1–19. <https://doi.org/10.1111/j.1468-2303.2005.00338.x>.
- Stulpe, Alexander and Matthias Lemke. 2016. Blended reading. Theoretische und praktische Dimensionen der Analyse von Text und sozialer Wirklichkeit im Zeitalter der Digitalisierung. In M. Lemke and G. Wiedemann (Eds.), *Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse*, pp. 17–61. Wiesbaden: Springer VS. [https://doi.org/10.1007/978-3-658-07224-7\\_2](https://doi.org/10.1007/978-3-658-07224-7_2).
- Thaller, Manfred. 1993. Historical information science: Is there such a thing? new comments on an old idea. *Historical Social Research / Historische Sozialforschung* 18(3), 5–19.
- Tosh, John. 2021. *The Pursuit of History: Aims, Methods and New Directions in the Study of History* (7 ed.). London: Routledge.
- Zhou, Zongwei, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. 2018. UNet++: A nested U-Net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 3–11. Springer. [https://doi.org/10.1007/978-3-030-00889-5\\_1](https://doi.org/10.1007/978-3-030-00889-5_1).

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s)

disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.