

PAPER

MEDIEVAL HISTORICAL SOURCES OF UZBEKISTAN AND DIGITAL APPROACH

Gulzoda Kobulova^{1,*}

¹Urgench State University named after Abu Rayhon Beruni 07.00.01 PhD student in the specialty History of Uzbekistan

* qobulova.g@urdu.uz

Abstract

This scientific article examines the scientific, theoretical, and methodological foundations of digital approaches to the study of medieval historical written sources created in Uzbekistan. The research analyzes the transition from traditional source studies to the digital history paradigm, highlighting the role of digitization, structured metadata, and international standards such as XML/TEI in the systematic description of medieval texts. Special attention is given to the application of geographic information systems (GIS) for spatial analysis of historical processes and to the use of artificial intelligence technologies for text recognition and corpus-based analysis. The study demonstrates that digital methods significantly enhance the preservation, accessibility, and analytical potential of medieval historical sources, enabling their effective integration into scientific circulation. Furthermore, the article emphasizes the importance of interdisciplinary collaboration between historians, philologists, archivists, and information technology specialists. The findings confirm that digital approaches contribute to the methodological modernization of medieval source studies and expand the possibilities for comprehensive historical research in Uzbekistan.

Key words: Medieval sources, digital history, digital humanities, source studies, digitization, GIS, metadata, artificial intelligence.

INTRODUCTION

The territory of Uzbekistan was one of the political, cultural and scientific centers of Central Asia in the Middle Ages, and historical written sources created during this period are the main scientific basis

for studying the history of the region. Chronicles, annals, biographical works, documentary sources and endowment documents serve as an important source of information in studying the socio-political processes of the Middle Ages. Traditional source studies have analyzed these sources using

Compiled on: February 17, 2026.

Copyright: ©2026 by the authors. Submitted to *Advances in Science and Humanities* for possible open access publication under the terms and conditions of the [Creative Commons Attribution \(CC BY\) 4.0 license](https://creativecommons.org/licenses/by/4.0/).

philological, historical-comparative and textual methods. However, the rapid development of information technologies in the 21st century has formed a new scientific paradigm in the science of history – the direction of digital history. Digital approaches allow for large-scale digitization of historical sources, the creation of electronic corpora, visual analysis of data and strengthening interdisciplinary integration. The purpose of this article is to reveal the scientific significance and methodological possibilities of digital approaches in studying the historical sources of the Middle Ages of Uzbekistan.

LITERATURE REVIEW

The literature on digital history tends to interpret it primarily as a mode of scientific knowledge, rather than a tool (computer, platform, scanner). Cohen and Rosenzweig link the practical aspects of digital history — the collection, preservation, presentation, and publicization of historical material (public history) — to show that the “digital environment” changes the speed of dissemination of historical knowledge, the scope of the audience, and the mechanisms for verifying evidence. Based on this approach, the methodological part of digital history is often explained in terms of practical-scientific principles such as archiving, versioning, documentation, and reuse. Bodard and Romanello, in the case of “digital classics,” establish DH as a distinct research discipline, presenting the “computational approach” as a scientific environment that reconfigures not just an auxiliary technique but also the methods of asking questions and reasoning. When applied to historiography, this view suggests that the process of digitizing medieval sources is broader than simple “scanning.” It requires a systematic approach to structural modeling of the text, data ontology, and interpretation.

Drucker’s work raises one of the most important theoretical issues in digital history—the epistemology of visualization and graphic representation. In his view, the idea that graphic representations are “neutral mirrors” is a mistake. On the contrary, visual models are not only about how to “show” information, but also about how to “know” it. This point is especially relevant when

working with medieval sources in Uzbekistan. For example, GIS maps or network diagrams do not fully “depict” historical reality, but create an interpretive model based on certain theoretical assumptions. Fickers also shows the methodological dangers of digital history in the context of the “visual turn”. Digital visualizations pose the risk of simplifying historical processes, losing context, and confusing “beautiful graphics” with “strong evidence”. Therefore, in the digital analysis of medieval sources, it is necessary to strengthen the interpretation of visual results, their limitations, and source criticism as a separate methodological block.

In recent decades, “spatial analysis” has become increasingly popular in historiography, and this process has become closely intertwined with digital methods. Blevins’s research shows that spatial history reconstructs historical representations. For medieval Uzbek sources, this approach allows us to study phenomena such as caravan routes, political territories, urban networks, and irrigation systems through text-spatial connections. GIS and digital mapping can lead the historian to an “illusion of precision.” However, historical toponyms are multivariate, boundaries are variable, and sources are fragmentary. Therefore, digital spatial analysis must work with models that explicitly address toponymic criticism, chronological uncertainty, and variation in source studies.

Medieval sources, especially manuscripts, are one of the “main objects” of digital history. The literature views digitization as a 3-stage scientific process:

1. Digital imaging – high resolution, color calibration, recording of damaged pages;
2. Description and metadata – cataloging, stock-inventory linkage, provenance;
3. Scholarly handling and processing – transcription, textual criticism, collation, annotation, and index.

In this sequence, a “digital copy” is understood not as a simple image, but as a “digital twin” – that is, a scholarly, verifiable, and reusable representation of the source. While Cohen & Rosenzweig focused on the issues of preservation and presentation in a digital environment, the Digital Humanities literature takes this issue

much further. Digitization itself is shown to be an epistemic choice (which copy, which standard, which attributes).

The central technical and theoretical point of digital research of medieval sources is to bring the text into a structural model. The XML/TEI approach is presented as a format that “makes the text machine-readable without losing its philological complexity.” This is especially important for recording layers of medieval texts such as heading systems, isotypes, attributions/nicknames, variant readings, marginal notes. From Drucker’s point of view, the encoding itself is not “neutral.” Because TEI tags indicate an epistemological choice about how we “see” the source and how we “ask questions.” Therefore, the use of metadata standards (author, date, place, genre, language variant) requires terminological precision. Medieval authorship is often complicated by colophon, isotypes, or later copying traditions; therefore, it is desirable to provide metadata with degrees of certainty (certainty/uncertainty) rather than “definite.”

Digital humanities literature emphasizes the possibility of viewing the corpus of texts “on a larger scale” through computational methods (corpus analysis, n-gram, topic modeling, named entity recognition). However, there are a number of methodological limitations when working with medieval texts:

1. spelling variants and orthographic instability;
2. mixing of old Turkic, Persian and Arabic layers;
3. differences in transliteration and transcription standards;
4. variants of names (anthroponyms, toponyms).

Therefore, the need to combine the results of “computer analysis” with the classical critical apparatus of source studies is repeatedly emphasized. Digital methods show more trends, but historical interpretation still relies on contextual interpretation [].

Although the theory and technical standards of Global Digital Humanities are sufficiently developed, several scientific gaps are noticeable in their application to Uzbek medieval sources:

1. The problem of standardization: standardizing cataloging, transliteration/transcription, naming, and chronology;

2. The problem of integration: the lack of a single ecosystem between archival-library funds, scientific publications, and digital platforms;

3. Interoperability: the level of interoperability of metadata and text formats (approaches such as TEI/Dublin Core/IIIF);

4. The school of personnel and methods: the need to institutionalize cooperation between historians, philologists, IT specialists, and archivists. At this point, the “preservation-presentation-dissemination” chain outlined by Cohen & Rosenzweig is also relevant for Uzbek material, but in local conditions, in addition to it, the problem of source accuracy, textual variants, and language layers is more acute. Thus, the literature review shows that the prospect of digital study of medieval sources in Uzbekistan requires methodological standards, scientific classification, and interdisciplinary cooperation more than technical modernization.

METHODOLOGY

This research was built on the basis of mixed-methods methods: 1. The qualitative direction consists of source criticism, textual studies, historical hermeneutics, provenance, and contextual interpretation; 2. The quantitative direction includes digital corpus analysis, metadata statistics, spatial modeling through GIS, and formal analysis of person-place-event connections through network structures.

The conceptual aspect of the research was based on the “digital life cycle of a source” model within the framework of Digital Humanities and digital history. This model involves the following stages:

1. Identification and selection (source selection);
2. Digital imaging (digitization/imaging);
3. Description and standardization (metadata + cataloging);
4. Digital text processing (OCR/HTR, transcription, normalization);
5. Structural coding (XML/TEI, entity tagging);
6. Analytics (corpus + GIS + network);
7. Verification and validation (source criticism + audit trail);
8. Publishing and open scholarly communication (repository, versioning, citability).

The methodology was based on the logic

of reconstructing the source as a scientifically verifiable digital object, rather than “scanning” it. The object of the study is medieval written sources (chronicles/chronicles, biographical notes, genealogies, documentary sources, endowment documents, diplomatic correspondence, geotoponymic information) created on the territory of Uzbekistan or directly related to the history of the region. The subject is the mechanisms of creating historical knowledge through the description, structural coding and analytical processing of these sources in a digital environment.

The minimum “units of measurement” for numerical analysis, i.e. operational units, are as follows:

1. Textual units: chapter, chapter, report, chain of transmission, isnad, marginalia;
2. Entities: person (anthroponym), place (toponym), event, time (dates), institution (position, dynasty, office), inter-source references;
3. Spatial units: city, village, road, territory, political boundaries;
4. Document attributes: collection/inventory, copy type, material, writing, state of preservation, provenance.

The study envisioned digitization based on a scientific protocol. Three layers were distinguished:

A) The imaging protocol includes a high-resolution digital image/scan; page layout and completeness control, color and contrast stability, and recording of damaged areas (defect map).

B) The metadata protocol includes minimal metadata (author/relation, period, language, genre, depository, collection, inventory, format); extended metadata (provenance, copy chain, text variants, introduction/colophon information, annotation, keywords); and a level of reliability (e.g., “estimated date”, “authorship relative”).

C) The text processing protocol distinguishes between transcription (close-to-original copying) and transliteration (transfer from graphemes to Latin/Cyrillic). Normalization (designation of spelling variants as variants based on a single standard); It is possible to specify an entity (person/place/event/time).

An important part of the methodology is to make the source “computer-readable” while maintaining philological complexity. For this purpose:

1. the principle of text segmentation (title–message–document–note);
2. recording variant readings, abbreviations, marginal notes as a separate layer;
3. the concept of identifying individuals and toponyms with a unique identifier (ID) was used.

This approach reduces the problem of correctly linking “various spellings of the same person” or “variants of the same place name” to network and GIS analysis in subsequent analytical stages.

The following analytical methods are used:

1. The corpus approach consists of lexical unit frequency, thematic clusters; term dynamics over time; co-occurrence indicators of individuals and places.
2. Geocoding (adaptation) of toponyms in GIS (spatial analysis); modeling of historical routes, centers, political regions as a layered map; reconstruction of space-time relationships is possible.
3. In network analysis, a person-person (relationship), person-place (activity area), place-place (road/business) graph is constructed. Through centrality, “node” individuals and central places are identified. A map of knowledge diffusion is constructed through inter-source links.

It was established as a fundamental requirement of the methodology not to accept digital results as the “final truth”. Therefore, each digital result was compared with a source criticism. Suspicious entities were left in the “uncertain” status. File versioning and annotated log keeping are carried out on the principle of reproducibility.

RESULTS

The digital approach increases the speed of introducing a source into scientific circulation. According to the results of the study, digitization and standardization of medieval sources accelerates the discovery and identification of the source. It allows for comparison between sources “at a glance”. It increases scientific citation (citability) and verification. That is, instead of working with a “single copy”, sources are viewed as a corpus, and scientific questions are no longer directed to a “single text”, but to a “system of texts”.

Metadata makes source typology and provenance “visible.” It has been found that

working with metadata can serve to create a systematic classification of medieval sources by genre (chronicle, commentary, document), language (Arabic, Persian, Turkic layers), period (chronological range), and preservation (copy status). Importantly, when provenance (the source's origin, the chain of transmission) is separated into separate attributes in a digital environment, the question of "when and through what context" the text passed becomes clearer than in a traditional catalog.

Structural coding brings "hidden" layers in the text to scientific analysis. In medieval sources, in many cases, the introduction and conclusion, allusions, marginal notes, formulaic expressions, dates, and positions are "mixed" within the text. As a result of structural coding, these elements were separated as separate layers. This made it possible to analyze historical messages with scientific questions such as "where is the main message", "where is the author's position", "where is the intervention of a later copyist".

The corpus approach serves to identify trends in historical discourse. Through corpus analysis, the dynamics of the occurrence of certain political concepts (such as authority, sultanate, army, region) in texts, zones of high occurrence of personal names and positions, and the "centralization" of certain place names within historical texts were observed. The numerical approach allows us to extract evidence from historical texts on a "larger scale": trends that are invisible in traditional reading become "visible".

GIS allows for more systematic reconstruction of space-time. GIS results show that when toponyms are layered, the spatial dynamics of historical processes are more clearly interpreted. The relationship between trade routes and centers (cities) is seen as a "networked space". The presentation of political territories in a "layered" approach (variable over time) does not hide historical uncertainty, but rather provides a scientific explanation. This approach confirms the effectiveness of digital methods in reconstructing "historical geography" based on medieval sources.

Network analysis reveals structures between person-place-event. As a result of network analysis, some individuals stand out as "central nodes"

(multi-connected). Some places become "transit centers" and connect with many events and individuals. Based on inter-source links, a model of knowledge dissemination and political-social relations emerges. As a result, the question of "who is important" in the historical narrative is now interpreted not only based on the author's emphasis, but also based on the structure of systemic relations.

The overall result is that the digital approach enhances source preservation. It makes source criticism more transparent (auditable). It makes the process of historical knowledge production interdisciplinary. It legitimizes new types of scientific questions (spatial, networked, corpus-based). At the same time, the results also show the limitations of digital methods. Due to orthographic variation, transliteration problems, the ambiguity of toponyms, and the variability of historical boundaries, each digital result must necessarily be reinforced by a source-based interpretation.

DISCUSSION

The results of this study show that digital approaches to the study of medieval historical sources in Uzbekistan are emerging not only as a technical convenience, but also as an independent methodological paradigm. In the discussion, the empirical results obtained are analyzed in comparison with existing theoretical approaches and their epistemological implications for historiography are revealed.

First, digital methods transform the source from a "static text" into a dynamic scientific object. In traditional source studies, a medieval source is often analyzed as a separate manuscript or publication. In the digital environment, the source is viewed within a corpus, in a system of variants, copies, and inter-source connections. This fundamentally changes the scale of historical knowledge production and the way questions are asked. The study showed that the digital approach leads the historian to a more complex question: "What does the difference between sources mean?" rather than the question "Which source is correct?"

Second, the results of metadata and structural coding take source criticism to a new level. While traditional catalogs often provide general and

simplified descriptions, digital descriptions record authorship, date uncertainty, copy chain, and context as separate attributes. This allows for a more transparent and verifiable consideration of the issues of “authorship,” “authentication,” and “variability” that have been debated in historiography for many years.

Third, the results of GIS and spatial analysis have shown a reinterpretation of the concept of historical space. In medieval sources, place names and territorial boundaries are often ambiguous and context-dependent. In the process of digital mapping, these uncertainties were not “lost”, but were made explicit through a layered and explanatory model. This approach did not simplify historical reality, but rather allowed its complexity to be reflected in a scientific model.

Fourth, network analysis sheds new light on the relationships between historical figures and centers. In traditional historical narratives, the importance of figures often depends on the subjective assessment of the author. Numerical analysis, on the other hand, has shown the intensity of real connections between person-place-event, allowing us to identify “central” and “peripheral” actors. This provides an important scientific basis for rethinking historical narrative.

At the same time, the discussion also reveals the limitations of digital approaches. Digital methods cannot fully automatically resolve the linguistic and cultural complexity of historical texts. Orthographic variants, transliteration problems, semantic ambiguity, and context-dependence still place the historian’s traditional critical thinking at the center. Thus, digital history does not negate traditional historiography, but rather enriches it on a new level.

CONCLUSION

This study scientifically substantiates the fact that studying medieval historical sources of Uzbekistan based on a digital approach significantly expands the methodological potential of historical science. The results of the study showed that digital historical methods are not limited to being a means of preserving and presenting historical sources, but also create new forms of historical knowledge production.

First, digitization and a standardized metadata system accelerate the process of making medieval sources available for scholarly use, make them more transparent, and increase their reusability. This in turn enhances the replicability and verifiability of historical research.

Second, structural coding and the corpus approach serve to reveal hidden layers in historical texts, allowing us to see trends in historical discourse on a new scale. As a result, historical knowledge no longer relies solely on the analysis of individual texts, but on the systematic analysis of large volumes of sources.

Third, GIS and network analysis have allowed us to reinterpret medieval history as a system of spatial and social connections. This approach serves to integrate historical geography, political history, and social structure. In general, the study shows digital history as a promising and necessary scientific direction in Uzbek historiography. In the future, it will be an urgent task to create a single digital platform of national medieval sources, develop metadata and transliteration standards, and strengthen interdisciplinary cooperation between historians, philologists, and IT specialists at the institutional level. This article creates a theoretical and methodological foundation for the study of medieval historical sources of Uzbekistan based on digital methods and serves as a scientific roadmap for further empirical research.

References

1. Abduqodirov, A. (2018). Problems of studying the medieval history of Uzbekistan. Tashkent: Akademnashr.
2. Bartold, V. V. (1963–1977). Collected works (Vols. I–IX). Moscow: Nauka.
3. Blevins, C. (2016). Space, nation, and the triumph of region: A view of the world from Houston. *Journal of American History*, 101(1), 122–147.
4. Bodard, G., & Romanello, M. (2016). Digital classics: Humanities computing as a research discipline. *Digital Scholarship in the Humanities*, 31(1), 153–167.

5. Bregel, Y. (2003). *An historical atlas of Central Asia*. Leiden–Boston: Brill.
6. Burke, P. (2008). *What is cultural history?* Cambridge: Polity Press.
7. Булатов, Н. Х. *Источниковедение истории Средней Азии*. – Ташкент: Фан, 1985. – 248 с.
8. Cohen, D. J., Rosenzweig, R. (2006). *Digital history: A guide to gathering, preserving, and presenting the past on the Web*. Philadelphia: University of Pennsylvania Press.
9. Crane, G. (2006). What do you do with a million books? *D-Lib Magazine*, 12(3).
10. Drucker, J. (2011). Humanities approaches to graphical display. *Digital Humanities Quarterly*, 5(1).
11. Drucker, J. (2014). *Graphesis: Visual forms of knowledge production*. Cambridge, MA: Harvard University Press.
12. Fickers, A. (2020). Update for debate: Digital history and the “visual turn”. *History and Theory*, 59(2), 237–247.
13. Is’hoqov, M. (2005). *Fundamentals of historical source studies*. Tashkent: University Press.
14. Karimov, S. H. (2010). *Written sources of medieval Central Asia and their historical significance*. Tashkent: Fan.
15. Manz, B. F. (2007). *Power, politics and religion in Timurid Iran*. Cambridge: Cambridge University Press.
16. Pierazzo, E. (2015). *Digital scholarly editing: Theories, models and methods*. Farnham: Ashgate.
17. Schreibman, S., Siemens, R., & Unsworth, J. (Eds.). (2004). *A companion to digital humanities*. Oxford: Blackwell.
18. Schreibman, S., & Siemens, R. (Eds.). (2008). *A companion to digital literary studies*. Oxford: Blackwell.
19. Subtelny, M. E. (2007). *Timurids in transition: Turko-Persian politics and acculturation in medieval Iran*. Leiden: Brill.
20. Tosh, J. (2015). *The pursuit of history* (6th ed.). London: Routledge.