Introduction

Dorit Raines
Università Ca' Foscari Venezia, Italia

Over the past two centuries, the increasing adoption of scientific methodologies in historical studies has prompted debates among historians regarding the most suitable unit of measurement for analysing historical phenomena and structuring them coherently to uncover underlying processes. Central to this discourse is the distinction between 'facts' and 'events.' While a fact recorded in historical sources requires contextualization to attain historical significance, an event represents a network of interconnected facts that, despite being open to multiple interpretations, provides the necessary framework for historical analysis.

Recent technological advancements have shifted this debate toward data-driven methodologies. Historians now conceptualize historical data as discrete informational units categorized into typologies such as persons, places, objects, and institutions. Although these units, when stored in databases, often lose their original context, they can be aggregated and analysed statistically to reveal latent historical structures. Nevertheless, their transformation into historical facts remains contingent upon contextualization.

The emergence of relational databases – such as entity-relationship models and linked data – has further raised questions regarding the extent to which historical records can be accurately represented through structured data extraction. From a historiographical perspective, reconstructing historical contexts necessitates the development of complex data architectures capable of expressing the intricate relationships between data units. This approach reflects the evolving nature of historical methodology, integrating traditional historiographical concerns with contemporary digital humanities techniques.

Edizioni Ca'Foscari

Studi di archivistica, bibliografia, paleografia 8
e-ISSN 2610-9093 | ISSN 2610-9875
ISBN [ebook] 978-88-6969-919-1 | ISBN [print] 978-88-6969-920-7

Peer review | Open access
Submitted 2025-03-21 | Published 2025-05-22

DOI 10.30687/978-88-6969-919-1/000

Tim Hitchcock argued back in 2013 that academic historians had largely failed to respond effectively to challenges such as algorithm-driven discovery, misleading forms of search, poor OCR, and biased selection of sources. He suggested that while historians had preserved the form of scholarly good practice such as a critical use of evidence, they had ignored some important underlying principles that the 'keyword searching' culture imposed upon them (Hitchcock 2013, 9, 12, 14). The complexity of historical narration or the multidimensional reality of the past (Dedieu 2016, 1; Diaz-Ordoñez, Rodríguez Baena, Yun-Casalilla 2023, 1016) was challenged again and again in several projects striving to create 'structured historical narrations' by the limited binary capacity of data processors. It became clear that a database, even a relational one, had difficulties coping with a nonlinear narration expressed in prose and with sequential logic (Bodenhamer 2008, 224).

In 2020, the nationally funded project *Nuncio's Secret Archives: Papal Diplomacy and European Multidenominational Societies Before the Thirty Years' War* (2017JMPYTA) was launched with the aim of enhancing a stratified private political archive through the creation of an open-access research portal. The central objective was to develop a relational database representing the contents of the correspondence of two envoys of the Holy See to Eastern Europe in the second half of the sixteenth century, utilizing a 'keyword-searching' approach. However, as the team grappled with designing the database and defining the entities to be extracted, they realized that the well-established historical methodologies they had relied on clashed with the binary structure of the digital framework. This realization prompted a creative approach to overcoming the challenges encountered along the way.

In essence, the project raised the question of whether historians should critically reassess their scholarly practices and engage in broader methodological reflection – specifically, whether what Hitchcock refers to as 'keyword searching' can be challenged or at least refined (Hitchcock 2013, 14). The starting point was the recognition that traditional practices based on source criticism and classification must not only coexist with but also inform the evolving digital research methodologies. Accordingly, the team began by identifying the specific challenges posed by this integration and exploring potential solutions.

This intellectual inquiry culminated in an international conference held at Ca' Foscari University of Venice on 27-28 October 2022, aimed at examining the present and future relationships between historical research and relational databases. The conference specifically addressed the following topics:

- modelization of historical serial records for data extraction;
- models of historical data architecture in relational databases;

- different uses of historical structured data in relational databases;
- the future of relational databases for historical research.

The present publication is based on part of the papers presented at the conference and on other essays specifically requested to cover several topics emerging from the conference but not specifically addressed in it. The outcome provides multidimensional insights into the evolving landscape of historiography. Through meticulous examination, the essays illustrate the transformative power of digital tools in reshaping the methodologies of historical inquiry, augmenting traditional practices with innovative approaches.

The section *Perspectives* delves into the theoretical frameworks that underpin the integration of digital methodologies in historical research, offering a comprehensive analysis of the long-standing interplay between narrative contextualization and source authentication. It highlights the shift towards entity classification and the restructuring of data, ultimately facilitating the remodulation of historical narratives.

In *Experiences*, five historians share their firsthand encounters with digital technologies, detailing their initial expectations, the integration of sources within digital platforms, the challenges faced, and the solutions devised. This section underscores the practical implications of digital tools in historical research, revealing the dynamic relationship between historians and the digital realm.

Lastly, *Challenges* presents a case study unfolding the creation of the Graziani archives portal, and illustrating the complexities involved in constructing a digital repository. This section provides a step-by-step account from four distinct perspectives, encompassing data architecture, handling disambiguation and uncertainty, leveraging relational databases for enhanced historical evidence, and describing material culture within a relational database framework.

What emerges from the essays included in the present publication are several key problems and their corresponding solutions.

According to the essays included in the present publication the problems the historian encounters when using relational databases are:

- complexity of historical context – historical databases must maintain archival context and relational integrity, which is often neglected in favour of specific inquiries;
- metadata standards – existing metadata standards and relational databases do not fully capture the complexity of historical narratives;
- fragmented research process – numerous digital tools exist for spatio-temporal inquiries, but they often operate in isolation, leading to a fragmented research process.

- disambiguation of entities –the process of disambiguating entities can complicate documentation and standardization, impacting historical research. In fact, it is essential to understand the process that leads historians to identify or disambiguate an event, a person or a place;
- complexity of private archives – historians struggle with organizing unstructured data, particularly from fragmented private collections. These archives often have complexities such as lack of explicit attribution, missing information, and unknown dates or locations of composition;
- absence of long-term preservation infrastructures for relational databases portals – the lack of support and general services stems from the absence of shared common vision regarding the use and management of research data.

Several solutions were proposed by the scholars, based on their experience with relational databases:

- simulations for metadata extraction – employing simulations to extract metadata from archives results in quicker and more accurate identification compared to manual methods;
- structured data – the transition from unstructured to structured data through preprocessing techniques like content management, topic modelling, and social network analysis forms the basis for effective historical research;
- text mining techniques – integrating text mining with metadata about authors provides a robust method for investigating historical phenomena;
- critical approach to data – emphasizing the difference between 'data' (objective observations) and 'capta' (selected, interpreted information) helps maintain a critical approach to knowledge production.

In conclusion, while the challenges in historical data organization and analysis are manifold, the adoption of digital tools, structured approaches, and advanced methodologies offer promising solutions for more accurate and efficient historical research. Undoubtedly, non-digital historiography can benefit from algorithm-based research, focusing on structured data extraction and analysis. By addressing these issues, researchers can better navigate the intricacies of historical narratives and contribute to a deeper understanding of the past. Projects like the Graziani Archives facilitate digital rearrangement and analysis of data, addressing the inherent complexities and fostering easier identification and classification.

Bibliography

Bodenhamer, D.J. (2008). "History and Gis: Implications for the Discipline". Knowles, A.K. (ed.), *Placing history. How Maps, Spatial Data, and GIS Are Changing Historical Scholarship*. Bedlands (CA): ESRI Press, 219-33.

Dedieu, J.P. (2016). "Designing Databases for Historical Research with Special Reference to Fichoz". *Réseaux et Histoire*, 8 July.
https://reshist.hypotheses.org/files/2016/07/Dedieu_Historical_Databases.pdf

Diaz-Ordoñez, M.; Rodríguez Baena, D.S.; Yun-Casalilla, B. (2023). "A New Approach for the Construction of Historical Databases – NoSQL Document-Oriented Databases: The Example of *AtlantoCracies*". *Digital Scholarship in the Humanities*, 38(3), 1014-32.
https://doi.org/10.1093/llc/fqad033

Hitchcock, T. (2013). "Confronting the Digital or How Academic History Lost the Plot". *Cultural and Social History,* 10(1), 9-23
https://doi.org/10.2752/147800413X13515292098070