



Article

Digital Perspectives in History

Anna Siebold ^{1,2} and Matteo Valleriani ^{2,3,4,*} 

¹ Fak. IV, University of Oldenburg Haarentor Campus, 26129 Oldenburg, Germany; asiebold@mpiwg-berlin.mpg.de

² Department I, Max Planck Institute for the History of Science, 14195 Berlin, Germany

³ Fak. 1, Berlin Institute of Technology, 10623 Berlin, Germany

⁴ Faculty of Humanities, Tel Aviv University, Tel Aviv 6997801, Israel

* Correspondence: valleriani@mpiwg-berlin.mpg.de

Abstract: This article outlines the state of digital perspectives in historical research, some of the methods and tools in use by digital historians, and the possible or even necessary steps in the future development of the digital approach. We begin by describing three main computational approaches: digital databases and repositories, network analysis, and Machine Learning. We also address data models and ontologies in the larger context of the demand for sustainability and linked research data. The section is followed by a discussion of the (much needed) standards and policies concerning data quality and transparency. We conclude with a consideration of future scenarios and challenges for computational research.

Keywords: computational history; digital humanities; digital database; network analysis; machine learning; sustainable research data; transparency; digital hermeneutics



Citation: Siebold, Anna, and Matteo Valleriani. 2022. Digital Perspectives in History. *Histories* 2: 170–177. <https://doi.org/10.3390/histories2020013>

Academic Editors: Volker Remmert, Dania Achermann, Cécile Stephanie Stehrenberger and Fabian Link

Received: 31 March 2022

Accepted: 30 May 2022

Published: 4 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The state of the art of digital perspectives in historical scholarship can only be considered within the larger framework of the Digital Humanities (DH). Although digital humanists were engaged in definitional debates up until the early years of the 2010s, in recent years DH scholarship has both proliferated and matured (Gold 2012). Not only is it increasingly published in a number of established journals and represented institutionally by university chairs, but DH scholarship is also characterized by a deepening and narrowing of scholarly niches in the field. Digital or Computational History, encompassing the range of digital perspectives in historical research, may be considered such a niche. Its relevance, however, to DH as a whole lies not only in the historical content it produces, but also in the methodological, self-reflexive debates it prompts (Zaagsma 2013). Recent examples include the ongoing debates around digital source criticism, digital hermeneutics, the demand for the development of algorithm and tool criticism, as well as discussions about the potential effects of digital methods on the production of (historical) knowledge (Fickers and Zaagsma 2022).

The digitization of historical sources, however, gained momentum about twenty years ago and was driven primarily by two factors. First, the increasing availability of the Internet offered new possibilities for the exchange of ideas between scholars. Email, electronic mailing lists (distributed for instance by the online forum H-Net launched in 1993), electronic publishing (disseminated initially via CDs, DVDs, and later the Internet), and the creation of digital libraries and catalogs serve as early examples. From its very beginning, the spread of the Internet was thus accompanied by the dream that the online world would create a new form of global (academic) community. Second, memory institutions such as archives, libraries, and heritage institutions saw the digitization process—one not necessarily tied to Internet dissemination—as a new means of preserving historical sources, since a digitized manuscript does not need to be physically browsed.

The surge of research projects engaged in digitizing historical sources led to the creation of large digital collections, which have since been integrated into research practices and are now essential for the making of new historical knowledge. Against this background, highly structured and painstakingly curated collections of data have emerged, and at the same time, various digital tools and methods have been conceived, developed, and applied in order to analyze these collections. The question of whether these changes give rise to new research questions and approaches remains open, although recently it has been the focus of increased scholarly attention.

Given that Computational History itself has already undergone various developments, our aim in this article is to provide an introduction to a selection of computational tools and technologies for historians interested in digital approaches. Our contribution is not intended as an instructive introduction for practicing digital historians, but rather as an overview and consolidation of some of the computational approaches relevant to historical research developed over the last years, as well as some thoughts on how they might evolve in the future. We focus on three main areas—databases, network analysis, and Machine Learning (ML)—illustrated by examples from research projects in the history of science. The subsequent sections address the related issues of sustainability as a necessary condition for the outlined future development of Computational History, and the challenges of ensuring transparency and data quality.

2. Computational Approaches and Their Future

Computational approaches used for historical research purposes are manifold, and the three areas discussed here are not exhaustive. Their selection is informed by our scholarly environment and experience in the field. Comprehensive introductions to a wider range of technologies and tools can be found in what has become a large body of literature (Graham et al. 2015; Drucker 2021).

2.1. Digital Databases

As mentioned in the introduction, access to the Internet as well as the digitization of historical sources led to the creation of large digital collections. These collections, in turn, required repositories in which data could be stored, described, and managed. Now an essential part of humanities research, digital databases range from relatively simple digital catalogs containing numerical or textual information to more advanced systems managing a variety of multimedia content such as images, videos, audio files, and maps.

Initially, DH research projects used database technologies developed in the context of other disciplines and applications, mostly around the intersection of computer science and business. Among the most widely used technology is the relational database, which was developed in the early 1970s and released as a product in 1978. Compared to earlier database systems (which were based on hierarchical or network database models), the relational database simplified the management, processing, and querying of data. This is partly due to the introduction of the table as an organizing principle, which enables data to be organized into rows and columns that can be related to each other. For historical research, the relational database meant that historians could conduct qualitative research digitally (Kemman 2021). The first system designed specifically for historical research purposes, called CLIO and developed by historian Manfred Thaller in 1980, has in fact been regarded as the beginning of “history and computing” (Boonstra et al. 2004). Today, relational database systems in DH prevail, which is particularly evident from the fact that general handbook entries about databases often exclusively cover relational databases (Crompton et al. 2016).

More recently, however, against the background of the development of so-called semantic technologies, a new type of database model has emerged: the graph database. Graph-based structures are well suited for the modeling and storing of complex, highly connected datasets and thus also for historical data. In comparison to the relational data model, in the graph database, distinct entities (people, places, events, objects, or concepts)

are connected via relations that form a network. Such representations are also referred to as knowledge graphs, and essentially represent the historian's view of a particular research field. In addition to enabling the modeling of complex structures, graph databases also provide more flexibility with regard to the database schema (i.e., it is easier to make changes to the data and structure), facilitate the querying of relational structures, and support the visualization of networks. These qualities have made them increasingly promising for DH research.

Examples of large collections managed in databases for research purposes are collections management systems, i.e., cataloging databases used in libraries, museums, and archives. Most of them also offer access to digitized versions of (some of) their sources via additional database systems. The library of the Max Planck Institute for the History of Science, for instance, began digitizing its collection of rare sources in 2002 as part of the project ECHO—Cultural Heritage Online, which was recently replaced by Digital Libraries Connected ([Collections MPIWG—Digital Libraries Connected n.d.](#)). Today, its library holds more than 1500 volumes of digitized primary sources and texts, as well as maps, relevant to all aspects of the history of science. Scholars engaged in Computational History and beyond, however, do not only make use of already existing database collections. Some contribute to them or build their own, participating in the database design process and the numerous decisions that accompany it.

To conclude this brief overview, it is important to point out that while digital historical research initially integrated discipline-foreign database technology into its research practices, it is now in the process of developing its own technologies and standards to address the specific requirements of humanities research. Particularly, many efforts focus on how information is best modeled, stored, and exchanged, questions discussed at greater length in the section below-titled Sustainability: Data Models and Ontologies. Finally, it is important to add that databases have evolved from being used mainly as repositories for storing, manipulating, and searching data to serving as the backbone of research projects, the site from which further software is applied to analyze and visualize the collected data. Databases have thus become vehicles for further data exploration.

2.2. Network Analysis

To analyze data, humanities scholars, especially in the frame of Computational History, have increasingly turned to network analysis. Examples can be found in the publications, workshops, and conferences organized by the Historical Network Research Community (launched in 2009), its open-access journal launched in 2017, and the many introductions to historical network research ([Düring et al. 2016](#); [Kerschbaumer et al. 2020](#)). Network analyses (similar to graph databases) emphasize the relationships between people, places, events, objects, or concepts. They aim to describe the character of a network, its density or central orientation, the nature of relationships in the network, and who or what occupies a central role. They allow, in contrast to a single document or biography, for the description of complex behavior in a network of relationships over time ([König 2019](#)). Modeled networks that emerge in this context, however, are not to be understood as representations of real-world networks, but rather as a way to approach and view a particular field of research. Once a researcher has determined that the application of network analysis is beneficial to the overall research question, it involves several steps: the collection of data, its encoding, algorithmic computation, and subsequent visualization.

An illustrative example is the research project *The Sphere: Knowledge System Evolution and the Shared Scientific Identity in Europe* ([The Sphere n.d.](#)). It consists of several components: at its core is a rich collection of bibliographic data from a corpus of more than 350 different editions of textbooks used for teaching astronomy in European universities from the late fifteenth century to the mid-seventeenth century. This data is stored together with digital copies of the textbooks in a graph database. The corpus is centered around a specific text, namely Johannes de Sacrobosco's treatise *Tractatus de sphaera*, and consists of editions that either contain this text or are closely related to it. Originally compiled in the thirteenth

century at the University of Paris, the text represents a qualitative introduction to geocentric astronomy, which was read at nearly all European universities until ca. 1650. The text was repeatedly reprinted, annotated, modified, and, in the printed editions enriched by other texts that deepened particular aspects. Given such a long tradition, attention was directed to the question of how the embedded knowledge changed over time and, in particular, how it became increasingly homogenous. To investigate this question, different parts of each edition were identified and semantically related to one another, creating a semantic network on which network analysis, the second important building block of the project, could then be applied. Among other things, this approach enabled the computational identification of which “families of books” were successful, in as much as they (a) became models imitated by others over a long period of time and (b) introduced new knowledge that was included in subsequent editions (Kräutli and Valleriani 2017; Valleriani et al. 2019; Zamani et al. 2020).

It is important to point out that there is no all-encompassing universal network theory. What does exist, however, is a shared core of analytical concepts, such as density, centrality, and community building. Many of these have been transformed into indicators, implemented in software, and are presented in textbooks (Lemerancier 2015). Although initially scholars primarily invoked Social Network Analysis (SNA)—the approach that seemed most appropriate to their subject matters—recent developments reveal the limitations of SNA, which is why some historians are moving toward approaches that were originally developed in the frame of physics of complex systems. The reason for this shift lies in the nature of the datasets that the historical sources generate; they are usually somewhat smaller than those of the natural or life sciences. Big Data, especially in the context of a single research project, are still rare phenomena in the humanities, and using the term often reflects the intent of using a buzzword rather than describing the scale of the dataset. What matters, however, is that humanities datasets exhibit a degree of data heterogeneity and dimensionality with which statistics and physics have little experience. In other words, the widespread notion that the historian depends exclusively on the tools and approaches of computer science or mathematics is increasingly proving false. Instead, the humanities are both challenging these disciplines and encouraging fruitful collaborations—developments that may well continue into the future.

2.3. Machine Learning

Machine Learning (ML) is based on the idea that once trained with given examples, computers are able to recognize connections and patterns and thus ‘learn’ (Schöch 2017). ML can be performed with two kinds of rationale. On the one hand, it is used to provide new information about datasets and their structures by classifying, clustering, or inferring new relations between various data points using supervised and unsupervised approaches. On the other hand, it is possible to analyze which features or rules were crucial for a particular classification or clustering, potentially resulting in a better understanding of the object of study (Holzinger et al. 2022).

ML, and most recently Neural Networks, entered the toolbox of historians in the context of computer-vision applications, for instance as part of further developing Optical Character Recognition (OCR), the conversion of images of hand-written, typed, or printed text into machine-readable text (Lyu et al. 2021). The poor quality of OCR techniques, which persistently proved unsatisfactory when dealing with early modern historical sources or working with handwritten material, led to the integration of ML, with the hope of turning the existing corpus of manually cleaned texts and their corresponding images into a training set for neural networks. In addition to applying ML technology to create machine-readable texts—an achievement which, in turn, allows for the application of a plethora of text analysis approaches—ML is increasingly applied with the intention of atomizing (segmenting) historical sources in order to discover and capture specific elements, such as illustrations, tables, and diagrams, or in fact any other text or shape that reappears (Lee et al. 2020).

One example of a project that uses Deep Learning-based image recognition is *Digital Heraldry*. Its research goal is to trace coats of arms, the most common visual medium of the Middle Ages and Early Modern times, across time, space, and societal groups ([Digital Heraldry n.d.](#)). The first step involved collecting digitized historical sources that contain representations of coats of arms and building a database. A part of the dataset then served as the basis for training image recognition algorithms to successfully predict areas in images that depict coats of arms in additional historical sources ([Hiltmann et al. 2020](#)). In the future, these trained algorithms may be applied to the digital collections of other institutions, thus expanding the dataset. The project also aims to use ML and Semantic Web technologies to enable semi-automatic descriptions of coats of arms, thereby generating metadata that can be used for further analysis.

The possibility of generating this type of data represents a fundamental change in the epistemic status of ‘error’. In comparison, the process of manually analyzing a large corpus of historical sources, i.e., humans extracting data, is time-consuming and error-prone—incorrect metadata, for instance, may be entered by mistake, and predefined keywords in a database can be mismatched. Processes of data cleaning are therefore applied up to the point where humans (scholars) decide that the data is of sufficient quality. In the case of historians, this means confidence that the number of errors is small enough to not affect the final research results. The advantage of using ML to replace the manual work of humans previously responsible for data extraction is that it will become possible to move from medium-sized corpora of sources to large ones—possibly a step towards “Big,” or at least, bigger data. By applying ML, however, errors become unavoidable products of source analysis. This is because ML technologies function in terms of probabilities and predictions. Neural networks trained to identify illustrations in historical sources, for instance, are unlikely to find all the illustrations in a given corpus. It can be argued, however, that the resulting incompleteness of the dataset is compensated by the increased size of the dataset and thus by the law of large numbers.

The development and proliferation of such algorithms in the frame of Computational History has only just begun, but it is expected that their implementation will soon generate larger datasets than previously possible. Not only will these datasets be available to humanities scholars, creating new challenges and opportunities for the formation of historical knowledge, but given the size of such datasets, ML will be required once again to analyze them adequately, for example, to search for genealogies of influence between sources. To achieve this, a new philology is being developed based on (graph-based) principles of sameness and similarity. It will enable, for instance, to cluster corpora of illustrations according to sameness, similarity, or the recurrence of specific elements among them.

3. Sustainability: Data Models and Ontologies

A close look at digital databases, network analysis, and ML shows that a very diverse and highly specialized set of new research practices has evolved, many of which are still being further developed. However, it must also be noted that the current situation is marked by a certain heterogeneity as to *how* these new practices are carried out. During the time that a project is active, the digital tools applied may, for instance, be based on a particular data model, tied to certain software, or be bounded by (limited) access rights. The need for standards and open or, even better, linked data policies, therefore, takes on an important role. Clear standards and policies are urgently needed, particularly because the heterogeneity described above not only prevents the sharing and linking of datasets, but worse, causes them to disappear altogether. Once a project is completed, publications may remain, but the datasets that were originally generated, together with their external web presentations and possibly even the electronic copies of the analyzed material are often no longer available online. This phenomenon is due to the lack of a sustained perspective: Because projects are usually funded for a limited period of time, it is highly uncommon for the institution that hosted the project to take over its maintenance and preservation ([Kräutli et al. 2021](#)). A record of lost datasets has so far not been compiled,

but the problem is widely recognized. The discipline has responded to this situation by developing DH-specific approaches with regard to data modeling and formal ontologies. The idea behind these endeavors is to standardize the structure of data to the extent that it becomes independent of the platform on which it was originally hosted, and that it can circulate, i.e., be hosted by different platforms and integrated with different datasets. This allows data to survive beyond the lifetime of the project in which it was created.

An increasingly widespread data model is the Resource Description Framework (RDF), which provides a syntax for representing data and resources on the Web. RDF breaks statements about resources into expressions of the form subject–predicate–object, also known as triples. The subject determines the resource, and the predicate determines traits or aspects of the resource, which can include relationships between the subject and another object. In this way, data stored and linked according to RDF form a graph structure. Each element of the triple can be expressed using re-usable Uniform Resource Identifiers (URI), which are compact sequences of characters that identify abstract or physical resources. A linked-data context thus means evolving from a document-based web to a web of linked data, allowing data to be linked at a whole different level.

Although the use of a common data model is an important step toward standardizing how data is structured, formal ontologies provide the possibility of expressing concepts in the same way, thus integrating different datasets in coherent semantic systems. Particularly relevant in this realm is the CIDOC Conceptual Reference Model (CRM), which is both a theoretical and conceptual tool for information integration in the field of cultural heritage. The use of a specific conceptual reference model presupposes a shared understanding of cultural heritage information and can, if applied across research projects, enable their integration in a semantic environment that stores and connects their data. The implementation of a shared data model and ontology thus allows for the realization of a vision that has always been inherent to the DH, namely being able to share and link data as well as enable inter-operability across projects and institutions.

4. Data Quality and Transparency

Another aspect that concerns the discipline as a whole is the need for new research principles that are agreed upon and followed by the entire community. The first concerns the transparency of data. More than often, research results are presented and published without accounting for the above-mentioned decisions. Which data model was applied? Which formal ontology forms the basis of the project's ontology? How was the data collected? Other factors might include more practical questions, such as how and to what extent was the collection of data limited (for example due to image rights limitations or uncooperative institutions)? Not only should these conditions, which ultimately shape the research endeavor and thus its results, be made transparent and considered, but the actual data should also be made available through so-called API access for others to use. In combination with the demands formulated here, this practice would support interoperability between projects and institutions, and hence fuel creative research endeavors. Just as importantly, it would enable datasets to be verified by other researchers at any time, thereby channeling expertise, making mutual reviewing more common, and finally guaranteeing a higher level of data quality.

Another pressing matter is that of crediting; working with and on data needs to be accounted for in academic crediting systems, since it is not only a time-consuming task, but also one that takes knowledge and skill. A broader debate is needed to find solutions for integrating crediting systems in the humanities that adequately reflect the nature of this type of work. Should this matter remain unresolved, the motivation to follow research questions that contain extensive data labor will diminish. This is especially the case for young scholars who are highly dependent on credits at the beginning of their careers. Another overdue discussion related to these issues concerns how to deal with the fact that research groups are increasingly interdisciplinary and, correspondingly, their publications multi-authored. Similar to what has been done in the natural sciences, it is necessary to

introduce standards that reflect the degrees of responsibility within the group so that they are transparent to other researchers and crediting systems.

5. Outlook

Despite the fact that not all historical sources are available in digital form and any form of digitization process involves the act of selecting, the availability of an unprecedented number of sources in digital form is making it both increasingly appealing and necessary for historians to turn to computational approaches. Although traditional approaches, especially with regard to case studies and in-depth analyses of specific sources, remain available to scholars and are equally essential for historical study, Computational History can be regarded as an extension of the historical disciplines, allowing for a broadening of research approaches and methods. In particular, it allows for *longue durée* investigations based on the close analysis of a large number of historical sources, which enables the formulation of new kinds of research questions. As mentioned in the introduction, we believe that one of the great potentials of Computational History for historical scholarship, but also for the humanities as a whole, lies in the fact that it addresses its own methodology. In recognizing the potential of Computational History to make tacit, hidden, or unconscious assumptions explicit (Krämer 2018), it becomes possible to reflect on how digital historical perspectives were produced in the first place. Which historical entities and relationships were modeled and thus taken into account, and which were not? What does a knowledge graph include and where do its limits lie? What tools or digital research approaches were applied, and on what grounds? Do they have a different effect on the making of historical knowledge than traditional approaches? The history of science in particular, with its long tradition of investigating epistemological changes, can make relevant contributions to answering these questions. Many of them will—if aptly considered and discussed—help to further formulate a hermeneutics of the digital, necessary not only in order to achieve digital literacy in academia but throughout the entire education system.

Author Contributions: Conceptualization and Writing: A.S. and M.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Max Planck Institute for the History of Science, by the Carl von Ossietzky Universität Oldenburg, and by the Berlin Institute for the Foundations of Learning and Data (BIFOLD): ref. 01IS18037A.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Boonstra, Onno, Leen Breure, and Peter Doorn. 2004. Past, Present and Future of Historical Information Science. *Historical Social Research* 29: 4–132. [CrossRef]
- Collections MPIWG—Digital Libraries Connected. n.d. Available online: <https://dlc.mpg.de/partner/mpiwg/> (accessed on 24 January 2022).
- Crompton, Constance, Richard Lane, and Ray Siemens. 2016. *Doing Digital Humanities: Practice, Training, Research*. New York: Taylor & Francis, ISBN 978-1-317-48113-3.
- Digital Heraldry. Cooperation between Historians and Computer Scientists. n.d. Available online: <https://digital-heraldry.github.io/> (accessed on 22 January 2022).
- Drucker, Johanna. 2021. *The Digital Humanities Coursebook: An Introduction to Digital Methods for Research and Scholarship*. London: Routledge, ISBN 978-1-00-310653-1.
- Düring, Marten, Ulrich Eumann, Martin Stark, and Linda von Keyserlingk, eds. 2016. *Handbuch Historische Netzwerkforschung: Grundlagen und Anwendungen*. Schriften des Kulturwissenschaftlichen Instituts Essen (KWI) zur Methodenforschung. Berlin: LIT, vol. 1, ISBN 978-3-643-11705-2.
- Fickers, Andreas, and Gerben Zaagsma. 2022. Digital Hermeneutics: The Reflexive Turn in Digital Public History? In *Handbook of Digital Public History*. Berlin: De Gruyter, pp. 139–48.
- Gold, Matthew K. 2012. Introduction: The Digital Humanities Moment. In *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press.
- Graham, Shawn, Ian Milligan, and Scott B. Weingart. 2015. *Exploring Big Historical Data: The Historian's Macroscope*. Singapore: World Scientific Publishing Company, ISBN 978-1-78326-611-1.

- Hiltmann, Torsten, Sebastian Thiele, and Benjamin Risse. 2020. Friends with Benefits: Wie Deep-Learning Basierte Bildanalyse Und Kulturhistorische Heraldik Voneinander Profitieren. Paper presented at DHd 2020 Spielräume: Digital Humanities zwischen Modellierung und Interpretation. 7. Tagung des Verbands "Digital Humanities im Deutschsprachigen Raum" (DHd 2020), Paderborn, Germany, March 2–6.
- Holzinger, Andreas, Anna Saranti, Christoph Molnar, Przemyslaw Biecek, and Wojciech Samek. 2022. Explainable AI Methods—A Brief Overview. In *xxAI—Beyond Explainable AI: International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers*. Edited by Andreas Holzinger, Randy Goebel, Ruth Fong, Taesup Moon, Klaus-Robert Müller and Wojciech Samek. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 13–38. ISBN 978-3-031-04083-2.
- Kemman, Max. 2021. *Trading Zones of Digital History*. Berlin: De Gruyter, ISBN 978-3-11-068225-0.
- Kerschbaumer, Florian, Linda von Keyserlingk-Rehbein, Martin Stark, and Marten Düring. 2020. *The Power of Networks: Prospects of Historical Network Research*. London and New York: Routledge, ISBN 978-1-315-18906-2.
- König, Mareike. 2019. Digitale Methoden in der Geschichtswissenschaft. Definitionen, Anwendungen, Herausforderungen. *BIOS—Zeitschrift für Biographieforschung, Oral History und Lebensverlaufsanalysen* 30: 7–21. [CrossRef]
- Krämer, Sybille. 2018. Der, Stachel des Digitalen—ein Anreiz zur Selbstreflexion in den Geisteswissenschaften? Ein philosophischer Kommentar zu den Digital Humanities in neun Thesen. *dco* 4: 5–11. [CrossRef]
- Kräutli, Florian, and Matteo Valleriani. 2017. CorpusTracer: A Cidoc Database for Tracing Knowledge Networks. *Digital Scholarship in the Humanities* 33: 336–46. [CrossRef]
- Kräutli, Florian, Esther Chen, and Matteo Valleriani. 2021. Linked Data Strategies for Conserving Digital Research Outputs: The Shelf Life of Digital Humanities. In *Information and Knowledge Organisation in Digital Humanities*. London: Routledge, ISBN 978-1-00-313181-6.
- Lee, Benjamin Charles Germain, Jaime Mears, Eileen Jakeway, Meghan Ferriter, Chris Adams, Nathan Yarasavage, Deborah Thomas, Kate Zwaard, and Daniel S. Weld. 2020. The Newspaper Navigator Dataset: Extracting Headlines and Visual Content from 16 Million Historic Newspaper Pages in Chronicling America. Paper presented at 29th ACM International Conference on Information & Knowledge Management; Association for Computing Machinery, New York, NY, USA, October 19; pp. 3055–62.
- Lemercier, Claire. 2015. Formal Network Methods in History: Why and How? In *Social Networks, Political Institutions, and Rural Societies*. Edited by Georg Fertig. Turnhout: Brepols Publishers, vol. 11, pp. 281–310.
- Lyu, Lijun, Maria Koutraki, Martin Krickl, and Besnik Fetahu. 2021. Neural OCR Post-Hoc Correction of Historical Corpora. *Transactions of the Association for Computational Linguistics* 9: 479–93. [CrossRef]
- Schöch, Christoph. 2017. Quantitative Analysen. In *Digital Humanities. Eine Einführung*. Edited by Fotis Jannidis, Hubertus Kohle and Malte Rehbein. Stuttgart: Metzler, pp. 279–98.
- The Sphere. n.d. Available online: <https://sphaera.mpiwg-berlin.mpg.de/> (accessed on 24 January 2022).
- Valleriani, Matteo, Florian Kräutli, Maryam Zamani, Alejandro Tejedor, Christoph Sander, Malte Vogl, Sabine Bertram, Gesa Funke, and Holger Kantz. 2019. The Emergence of Epistemic Communities in the 'Sphaera' Corpus: Mechanisms of Knowledge Evolution. *Journal of Historical Network Research* 3: 50–91. [CrossRef]
- Zaagsma, Gerben. 2013. On Digital History. *BMGN-LCHR* 128: 3. [CrossRef]
- Zamani, Maryam, Alejandro Tejedor, Malte Vogl, Florian Kräutli, Matteo Valleriani, and Holger Kantz. 2020. Evolution and Transformation of Early Modern Cosmological Knowledge: A Network Study. *Sci. Rep.* 10: 19822. [CrossRef] [PubMed]