

This is the author-created version of the following work:

**Langtree, Tanya, Birks, Melanie, and Biedermann, Narelle (2020) *Risky business? Addressing the challenges of historical methods in the 'digital age'*.
Collegian, 27 (6) pp. 589-594.**

Access to this file is available from:

<https://researchonline.jcu.edu.au/64395/>

© 2020 Australian College of Nursing Ltd. Published by Elsevier Ltd.

Please refer to the original source for the final version of this work:

<https://doi.org/10.1016/j.colegn.2020.06.009>

-Risky business? Addressing the challenges of historical methods in the 'digital age'

ABSTRACT

Background

The 'digital age' has led to a renaissance in historical methods. The way in which nurse historians can search, collate and analyse sources has changed exponentially over the past two decades. The mass digitisation of books, newspapers and other documents has resulted in the removal of many long-standing barriers to performing historical research, such as budgetary and access restrictions. Despite these expanded opportunities, the nurse historian now faces new challenges when performing historical research.

Aim

This paper aims to stimulate discussion on the risky business of conducting nursing historical research in the 'digital age'. In this paper, we examine the technology-born challenges encountered by nurse historians with the objective of proffering potential solutions to address such issues.

Discussion

Three contemporary challenges faced by nurse historians are: not knowing how to contain and articulate online searching; being unable to reduce the number of optical character recognition inaccuracies with digitised archaic sources; and being unsure of how to safely incorporate technological tools into historical analysis.

Conclusion

Used correctly, new technologies can augment and strengthen traditional historical methods. Nurse historians need to be mindful that the way in which technologies are used is controlled by the *user*, rather than the technology *itself*.

Keywords: Digital research; digitisation; historiography; history of nursing; nursing research.

SUMMARY OF RELEVANCE

Problem/issue

Nurse historians remain ill-equipped to navigate the use of digital technologies in their research. Used inappropriately, these technologies can dilute the rigour of their research outputs.

What is already known

The 'digital turn' has created new challenges and controversies in how to conduct historical research in the 21st century.

What the paper adds

This paper contributes to the scholarship of historical methods in the 'digital age' by presenting a commentary of the challenges encountered by nurse historians when using digital technologies.

1. INTRODUCTION

The 'digital age' has led to a renaissance in historical methods. The methods by which nursing historians collect and analyse sources are profoundly different from the practices employed two decades ago (Nicholson, 2016). Mass digitisation of books, newspapers and other primary sources has resulted in the historian being able to access 'big data' - volumes of text corpora - with just one click of the mouse. Being able to complete large (if not all) amounts of research online has eliminated many source access barriers once faced by the historian, such as securing funding to access those frequently hard-to-find collections in distant archives (Toms & O'Brien, 2008). In turn, historians are more open to investigating a wider range of research topics.

Despite the expanding possibilities for historical research, the emergence of 'big data' and digital research have created their own challenges and controversies (Cristianini, Lansdall-Welfare, & Dato,

2018; Drouin, 2014; Grossman, 2012). Concerns have been raised about: the legitimacy and quality of digital scholarship (Hitchcock, 2013; Huistra & Mellink, 2016; Knoblauch & Tomes 2014); the ethics of digitisation (Moravec, 2017); the preservation and safeguarding of digitised sources (Gailey, 2012); the impact of digitisation on library and archive budgets (Moravec, 2017); the risk of decontextualising the analysis (Hitchcock, 2013; Putnam, 2016); the usability and reliability of analytical tools (Cristianini et al., 2018); and, whether this ‘digital turn’ is just a passing trend (Knoblauch & Tomes, 2014).

While many of these concerns are valid, the ‘digital age’ is not repealing. Rather than shying away from digital technologies, nurse historians – nurses who have formally studied historical method – need to be at the forefront of conversations about their use and impact on historical methods (Grossman, 2012; Hitchcock, 2013; Huistra & Mellink, 2016; Putnam, 2016). Used correctly, such technologies can be an opportunity to address areas of methodological weakness (Knoblauch & Tomes, 2014). Used poorly, such technologies can weaken the rigour and quality of historical research. In this article, we examine several strategies that can aid the novice historian to navigate the risky business of conducting nursing historical research in the ‘digital age’.

2. BACKGROUND

The advances in information technology over the past two decades has revolutionised how evidence is searched, collated and analysed in historical methods. A ‘digital turn’ has occurred (Nicholson, 2016) – the methodology is fundamentally changed. The time-honoured traditions of the past no longer match how the majority of historians, including nurse historians, conduct and use historical research, yet we are at an impasse for knowing how to suitably advance the methodology (Hitchcock, 2013; Hoekstra & Koolen, 2019).

Not since the ‘linguistic turn’ have historians been confronted with so many confounding methodological uncertainties. The role of digital tools in historical scholarship remains under-theorised (Underwood, 2014) and under-evaluated (Koolen, van Gorp, & van Ossenbruggen, 2019). Standard

source criticism questions have not been updated to reflect how nurse historians should assess digitised materials (Koolen et al., 2019). The use of analytical tools is a divisive topic (Knoblauch & Tomes, 2014). Reports about the use of digitised sources and digital tools in narratives is at best extemporaneous (Hitchcock, 2013). And ultimately, it remains unclear if the new methods arising from the ‘digital turn’ have strengthened or diluted the rigour of research outputs (Underwood, 2014).

Twenty-first century nurse historians need to probe the efficacy of the ‘digital turn’ (Putnam, 2016) and develop new skills to negotiate these uncharted waters (Hoekstra & Koolen, 2019). During this era of risky business, we need to reflect not only on the benefits that technology can bring to the methodology, but also critique its limitations. For example, there is the risk that an overreliance on technology may impede the quality of historical research in areas such as data collection (e.g. endangering the intuitive process of discovery that normally occurs while searching for sources; Putnam, 2016) and data analysis (e.g. potentially negating the contextual and temporal semantics of language use found within the source; Drouin, 2014). In order for nurse historians to make informed decisions about the future direction of historical methods in nursing, it is therefore imperative that we begin to experiment with technology to gain a better understanding of its application and its inherent challenges. By having the courage to acknowledge and openly discuss the difficulties encountered while using such tools, nurse historians have the opportunity to innovate current and future practices. The subsequent sections of this paper provide a commentary on three contemporary challenges: online searching; optical character recognition (OCR); and technology-assisted analysis, before proffering potential solutions.

3. DISCUSSION

3.1 Challenges in online searching

Mass digitisation has increased the speed of discovery (Putnam, 2016). Thanks to sophisticated algorithms, nurse historians are now able to search and locate sources without having to first know where

to look (Putnam, 2016). In the past, this activity was a time-consuming and often serendipitous process of discovery where the researcher searched for the ‘unknowns’ by browsing the footnotes and references of secondary sources for potential leads (Given & Willson, 2018). This discovery is now a much more rapid affair accomplished by inputting keywords into a database or search engine (Hitchcock, 2013).

With this ease of discovery, however, comes new challenges that nurse historians must learn to navigate. The substantial volume of digital materials that are now available is seemingly endless as a result of the progressive digitisation of analogue sources (e.g., newspapers, books, government records) and the proliferation of ‘digitally-born’ materials including emails, e-books, e-magazines, blogs, webpages, and social media outputs (Huistra & Mellink, 2016). The proliferation of primary and secondary sources that are now readily accessed makes it easier for the historian to stumble into rabbit-holes of unrelated or inconsequential information.

These rabbit-holes often result from the messy process of online searching. A process that is further compromised by the lack of clear guidelines for how to conduct a thorough search. This lack of clarity can lead to deeper burrowing, leaving the historian questioning whether their search has been effective in capturing enough pertinent primary and secondary sources to adequately undertake their research (Huistra & Mellink, 2016). This phenomenon described as “digital dumpster diving” (due to the sheer volume of potential data available to the researcher) can severely compromise one’s ability to articulate and justify their search strategy (Gailey, 2012, p. 341).

The messiness of online searching helps to explain why many historians remain reluctant to cite the retrieval details of digitised sources in their footnotes or referencing (Huistra & Mellink, 2016; Koolen et al., 2019). Prevailing stigma about the use of some online searching and reference materials (e.g., Wikipedia) as being non-research, non-academic methods may also contribute to this reluctance (Wolff, 2013). Nonetheless, the protraction of a ‘don’t ask, don’t tell’ mentality whereby nurse historians fail to disclose how digitised sources were found and used in their research is detrimental to the dependability, trustworthiness and legitimacy of historical research in nursing (Langtree, Birks & Biedermann, 2019).

3.1.1 Minimising distractions: Structuring the search

In order to address the messiness of online searching, historians need a search strategy that has structure, but is also malleable to supporting the processes of ‘browsing’ (Toms & O’Brien, 2008) and ‘sideways glancing’ (Putnam, 2016). The use of a scoping review can meet these requirements as its framework provides an overt yet iterative structure to data searching and reporting (Arksey & O’Malley, 2005). While the majority of scoping reviews performed to date have examined health-related topics, their use in health-related historical research remains under-explored.

Scoping reviews allow the researcher to quickly identify what is already known about the research topic (Arksey & O’Malley, 2005). In historical research, scoping reviews can help to navigate the searching of online databases, repositories and libraries by providing a structured approach to the searching, screening and mapping of sources. Such structure gives the historian a sense of security in knowing that they have conducted a comprehensive search, yet it saves the historian time and energy as the opportunity to stumble into rabbit-holes that contain irrelevant materials is reduced.

Numerous frameworks exist for undertaking scoping reviews, however Arksey and O’Malley’s (2005) work remains the foundation for most of these approaches. Arksey and O’Malley’s (2005) framework outlines five mandatory stages for conducting a scoping review: identifying the research question, identifying relevant studies, study selection, charting the data, and collating, summarising and reporting results. Consultation is an optional sixth stage where the researcher seeks advice from experts in the field (Arksey & O’Malley, 2005). For historical research, the stages ‘identifying relevant studies’ and ‘study selection’ can be replaced with ‘identifying relevant *sources*’ and ‘*source* selection’ to reflect the type of materials used in these investigations. Table 1 provides a summary of these stages.

Investing time to develop a scoping review protocol and then executing it can accelerate the data collection phase of historical research. This acceleration results from being more likely to detect flaws in the research question or study design during the protocol development stage. Potential issues that may be detected during this stage include poorly-constructed search terms, source access problems, the need to translate sources that are written in foreign languages, and OCR difficulties. By detecting these issues early on, nurse historians can then implement pre-emptive strategies to address them. The application of

the pre-determined eligibility criteria also speeds up the screening of sources as document triaging - the process where the historian screens all query-related materials in order to locate and save relevant sources for later analysis - becomes more efficient (Given & Willson, 2018). The use of a scoping review can also hasten the revision of search strategies and outcomes as each stage of the search process is clearly documented. Not only does this accelerate the historian's ability to review how a source was located and used in the study, it also improves the dependability of the study as an audit trail is produced (Langtree et al., 2019).

3.1.2 Maximising returns: Refining the search query

The rabbit-holes encountered during online searching can be limited by taking the time to carefully craft and refine search terms prior to their use. When developing the search terms, nurse historians need to first identify the key concepts of the area of study. From here, they can list alternative spellings for each identified term, such as the subtle differences in spelling between British and American English (e.g., haemorrhage vs. hemorrhage) and the somewhat ad-hoc application of hyphenisation in English (e.g., bloodletting vs. blood-letting vs. blood letting). Synonyms that can be used as an alternative to the search term should also be added to the search string to improve the representativeness of results (Huistra & Mellink, 2016). Using the previous example, bloodletting is also known as venesection or phlebotomy. It is also worthwhile to consider any closely-related applications of the search term. In this example, related procedures such as leeching (also referred to as leech application and leech therapy) and wet cupping (or wet-cupping) should be added to the search string.

Once the list of alternative terms is articulated, the nurse historian can build a search query using the Boolean operator of 'OR' (Huistra & Mellink, 2016). Using the example described above, the search query would be: 'bloodletting' OR 'blood letting' OR 'blood-letting' OR 'venesection' OR 'phlebotomy' OR 'leeching' OR 'leech application' OR 'leech therapy' OR 'wet cupping' OR 'wet-cupping'. This process can then be repeated for each remaining key concept found in the research question (e.g., the target group, time period and event being studied). Following the construction of each search string, the

sets can be pooled with the Boolean operator of ‘AND’ to create a combined search query (Huistra & Mellink, 2016).

When undertaking online searching, semantic nuances in language use must be considered. This includes recognising the intranational, transnational or cultural differences in word use and meaning that exist within a common language. Nurse historians must also be cognisant of the temporal nuances of language use (Thompson et al., 2016). One example of how a word’s meaning has changed over time is the term apoplexy. Apoplexy is derived from the ancient Greek term apoplexia meaning ‘to strike suddenly’ or ‘be struck down violently’ (Engelhardt, 2017). Until the Modern era, apoplexy was an umbrella term used to describe a number of neurological disorders where a sudden loss of consciousness was experienced. As autopsies became more prevalent in the 17th century, apoplexy was successively used to solely describe a stroke and is now rarely used in contemporary medicine (Engelhardt, 2017).

3.2 Challenges with using OCR on digitised sources

Digitisation has transformed how historical sources are read and analysed. The majority of these changes are due to the advent and use of OCR software. OCR allows analogue text to be converted into machine-readable data. This process facilitates many features of online searching, and the ability to perform keyword searching on full text documents (Hitchcock, 2013). It is also used in more sophisticated analytical tools such as text mining (TM) (Thompson et al., 2016) and network analysis (Anderson et al., 2017). OCR converts a scanned image of printed pages into alphanumeric characters by detecting the pixel patterns within the image (Blanke, Bryant, & Hedges, 2012).

Problems with the OCR of historical sources occur for numerous reasons. OCR tools are challenged by source degradation as a result of age and poor preservation practices, the presence of artefacts (e.g., dirt, the typeface of the back sheet bleeding through) and the non-standard layouts of sources (e.g., differing kerning - the spacing between letters; variable leading - the distances between lines) (Blanke et al., 2012). OCR tools also have difficulty recognising non-uniform or highly stylised characters. Therefore, sources that are handwritten or are printed using Early Modern typefaces such as

Gothic scripts are particularly prone to OCR inaccuracies (Hitchcock, 2013). The presence of archaic allographs can also confuse OCR. Examples of allographs include the letter ‘v’ that was used to represent both ‘u’ and ‘v’ and the long ‘s’ - f - being mistaken for the letter ‘f’) (Fig. 1).

3.2.1 Minimising misinterpretation: Improving OCR

While several pre-processing steps (e.g. binarisation, denoising, de-skewing) can be taken to improve the accuracy of OCR and readability of historical documents, most historians lack the knowledge and skillset to manually ‘clean’ a source (Blanke et al., 2012). Hence, when a source is plagued with OCR inaccuracies, a historian is often left with no other option but to return to hand-searching for key evidence. The employment of several workarounds to improve the accuracy of OCR with archaic sources can reduce the need for hand-searching.

The first strategy is enhancing the scanned image by adjusting the document properties within the PDF. Some programs such as Adobe Acrobat Pro DC feature an ‘Enhance Scan’ function allowing the historian to select a range of automated ‘cleaning’ tools such as de-skewing, text sharpening and background removal (used to remove the page pigmentation and contrast to improve text readability) (Fig. 2). Another useful feature that is available in this program is the ‘Recognise Text’ function. This feature enables the identification and correction of any OCR suspects - areas of text that are unreadable or questioned by the OCR tool.

In cases where the OCR tool does not have an in-built image optimisation, nurse historians can try to manually adjust the settings of the PDF to improve OCR. Manual adjustments that improve the OCR of a source include checking the image resolution of the PDF is set to 300 dots per inch (dpi) (Blanke et al., 2012; Cristianini et al., 2018); or for smaller fonts of less than 10 points, the resolution should be adjusted to between 400 and 600 dpi (Abbyy Technology Portal, 2017). Where possible, the colour scheme of the source should be adjusted to mimic binarisation (i.e., black text with a white background) so that the text becomes easier to recognise (Fig. 3) (Blanke et al., 2012). Before running

OCR on a foreign language source, the historian should also check and adjust the default language setting of the tool to ensure OCR is optimised (Cristianini et al., 2018).

In cases where OCR remains suboptimal, full text searching within the document can be accomplished through using a ‘stemming’ approach. ‘Stemming’ is a search method where only the stable stem of a word is searched rather than the entire word (Cristianini et al., 2018). For instance, the stem ‘nurs’ would locate words such as nursing, nurse and nurses. This approach reduces the orthographic variation of words and is particularly useful when spelling is non-standardised across countries (e.g., the differences between British and American English) or when the language is highly inflected (e.g., Italian and Spanish) (Reffle & Ringlstetter, 2013). While not ideal, ‘stemming’ enhances the historian’s ability to promptly review and triage potential sources.

3.3 Challenges with technology-assisted data analysis

The advent of macroscopic analytical tools such as TM and network analysis has transformed the way in which we *can* conduct historical research. Nonetheless, the use of such tools also creates a new set of challenges for the nurse historian who is inexperienced with using such software. The pressure to appear technologically-competent in the ‘digital age’ can result in some nurse historians being tempted to use an off-the-shelf analytical tool. For example, they could use an open-access tool such as Voyant Tools or Bookworm to identify key concepts within a corpus by reviewing word frequencies and usage patterns (Anderson et al., 2017). While mainstream tools are often user-friendly, they have the potential to produce unreliable results because of the way in which their artificial intelligence is generated (Cristianini et al., 2018). Many of these tools learn how to recognise different patterns and characteristics within multiple corpora using only contemporary materials (Thompson et al., 2016). Therefore, they may be unable to discern the temporal and semantic nuances of archaic language use, resulting in a misguided analysis.

Analytical mistakes can also occur if a tool applies the contemporary meaning of some words to older documents as the modern interpretation may be completely different to its original intent (Thompson et al., 2016). The interpretation of the word *crumpet* is a classic example: in contemporary lexicon,

crumpet refers to an air-filled pastry, whereas this word was once used to denote a person's head. Analysis gaps may also result as some content is literally lost in translation. This predicament can occur if the tool is unable to identify a word (e.g., the word *aliment* - comparable to food - has essentially disappeared from today's lexicon); cannot account for changes in word spelling (e.g., the archaic spelling of *show* was *shew*); or the word's meaning is the complete opposite of its archaic intention (e.g., *bully* used to mean 'sweetheart'). The risk of producing an inaccurate or indecipherable analysis means mainstream tools should be used cautiously with older sources.

3.3.1 Maximising context: Using CAQDAS to aid analysis

Adopting a hybrid approach to analysis can assist nurse historians to avoid the potential pitfalls of using macroscopic analysis tools while also enhancing the rigour of the study and its findings. This blending of traditional and nascent methods can be readily achieved through the use of a computer-assisted qualitative data analysis software (CAQDAS) program such as NVivo or ATLAS.ti. CAQDAS programs enable a blended reading approach to historical analysis as close reading (i.e., micro analysis) is combined with more advanced analytical tools (i.e., macro analysis) such as network analytical modelling. Despite CAQDAS being widely-used in other areas of qualitative research, their use in historical methods has received little attention.

Importantly, CAQDAS support the continuation of traditional historical methods such as close reading and source criticism but improves the auditability of these methods. For example, during close reading stage, annotations can be made directly onto the PDF of the digitised source or they can be added as a separate memo. The memo can then be directly linked to the source or any other file in the project using the program's linking feature (e.g., in NVivo, the 'See Also Link' or 'Memo Link' functions could be used). The use of such features serve as an effective way to jog one's memory at a later date as the created link will take the user straight back to the supplemental information. The use of such features can improve the efficiency and rigour of the analysis as the cross-checking of information is more straightforward and transparent than the handwritten prompts that are commonly used in traditional historical analysis.

The nurse historian's understanding of the research topic can also be improved through using the coding feature of the CAQDAS (Tummons, 2014). Coding, the process of assigning labels to different sections of text, enables the exploration of textual information by determining patterns within a corpus (Saldana, 2016). The use of CAQDAS allows nurse historians to concurrently code while performing close reading. Examples of how codes can be used with historical sources include denoting: a contextual aspect of the source (e.g., epoch), a bibliographical aspect of a source (e.g., author), or a key figure, concept, or other area of interest that is evident in the source (Saldana, 2016). This type of coding permits a range of different phenomena to be subsequently explored across the corpus (Tummons, 2014). These codes can then be reviewed, revised or collapsed into larger themes or categories as the analysis progresses, reflecting the iterative process that normally occurs during traditional historical analysis. Through coding the corpus, a systematic method to the historical analysis becomes evident – it creates a log of discovery that the nurse historian and others can use to follow the logic behind the analysis (Hoekstra & Koolen, 2019).

Most CAQDAS programs also have the ability to perform a number of distant reading analytics such as generating cluster analysis diagrams or concept maps. Normally, the researcher's coding is used to generate these visual representations rather than text recognition, therefore the accuracy of these visualisations remains uninfluenced by poor OCR or digitisation practices. A CAQDAS-generated visual model can highlight trends within the analysis that either confirm or refute the historian's preliminary inferences (Hoekstra & Koolen, 2019). Alternatively, viewing the coded data as a model may reveal gaps in the analysis or areas that need further investigation such as revealing an unexpected connection between two key figures. Such visual representations therefore act as a mediator between the corpus content and the nurse historian – helping them to develop a rich mental model and consider phenomena from multiple perspectives. This mental modelling is an essential component in historical contextualisation (Baron, 2016).

While the use of CAQDAS in historical methods is nascent, the potential benefits of their use – improving the transparency of analysis by documenting the historian's thought processes; enabling the prompt cross-checking of information; permitting the testing of inferences; and reducing the risk of

decontextualisation – offer valid reasons for why CAQDAS programs can have a place in contemporary historical methods.

4. CONCLUSION

The ‘digital age’ has transformed how nursing historical research is now conducted. Now, the nurse historian is able complete large (if not all) amounts of their research online – resulting in a broadened scope of potential sources and research topics. With these expanded opportunities, nurse historians face new challenges when performing historical research. As the end-users of such technologies, it is imperative that nurse historians be at the forefront for determining how these tools are adopted and used in the future. In the meantime, strategies such as conducting a scoping review, refining search queries, improving OCR or using a CAQDAS program can assist in navigating the present challenges faced in historical methods. Through implementing such strategies, nurse historians are best positioned to navigate the risky business of digital historical research by demonstrating the way in which technologies are used is controlled by the *user*, rather than the technology *itself*.

Disclosures

The authors declare they have no competing interests.

Ethical statement

An ethical statement is not applicable for this manuscript.

Conflict of interest

None.

References

- Abby Technology Portal. (2017). OCR - Optical image resolution. Retrieved from [https://abbyy.technology/en:kb:images_resolution_size_ocr?s\[\]=400&s\[\]=600&s\[\]=dpi](https://abbyy.technology/en:kb:images_resolution_size_ocr?s[]=400&s[]=600&s[]=dpi).
- Anderson, C., Ceserani, G., Donaldson, C., Gregory, I. N., Hall, M., Rosenbaum, A. T., & Taylor, J. E. (2017). Digital humanities and tourism history. *Journal of Tourism History*, 9(2-3), 246-269. doi:10.1080/1755182X.2017.1419455
- Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19-32. doi:10.1080/1364557032000119616
- Baron, C. (2016). Using embedded visual coding to support contextualization of historical texts. *American Educational Research Journal*, 53(3), 516-40. doi: 10.3102/0002831216637347
- Blanke, T., Bryant, M., & Hedges, M. (2012). Ocropodium: Open source OCR for small-scale historical archives. *Journal of Information Science*, 38(1), 76-86. doi: 10.1177/0165551511429418
- Cristianini, N., Lansdall-Welfare, T., & Dato, G. (2018). Large-scale content analysis of historical newspapers in the town of Gorizia 1873–1914. *Historical Methods*, 51(3), 139-164. doi: 10.1080/01615440.2018.1443862
- Drouin, J. (2014). Close- and distant-reading modernism: Network analysis, text mining, and teaching 'The Little Review'. *The Journal of Modern Periodical Studies*, 5(1), 110-135. doi: 10.5325/jmodeperistud.5.1.0110
- Engelhardt, E. (2017). Apoplexy, cerebrovascular disease, and stroke: Historical evolution of terms and definitions. *Dementia and Neuropsychologia*, 11(4), 449-453. doi: 10.1590/1980-57642016dn11-040016
- Gailey, A. (2012). Editing in the age of automation. *Texas Studies in Literature and Language*, 54(3), 340-356. doi: 10.1353/tsl.2012.0017
- Given, L. M., & Willson, R. (2018). Information technology and the humanities scholar: Documenting digital research practices. *Journal of the Association for Information Science and Technology*, 69(6), 807-819. doi: 10.1002/asi.24008
- Grossman, J. (2012). "Big data": An opportunity for historians. *Perspectives on History*, 50(3). Retrieved from <https://www.historians.org/publications-and-directories/perspectives-on-history/march-2012/big-data-an-opportunity-for-historians>
- Hitchcock, T. (2013). Confronting the digital: Or how academic history writing lost the plot. *Cultural and Social History*, 10(1), 9-23. doi: 10.2752/147800413X13515292098070
- Hoekstra, R., & Marijn K. (2019). Data scopes for digital history research. *Historical Methods*, 52(2), 79-94. doi: 10.1080/01615440.2018.1484676
- Huistra, H., & Mellink, B. (2016). Phrasing history: Selecting sources in digital repositories. *Historical Methods*, 4(4), 220-229. doi: 10.1080/01615440.2016.1205964
- Knoblauch, H., & Tomes, N. (2014). The history of medicine in the digital age. *Bulletin of the History of Medicine*, 88(4), 730-733. doi: 10.1353/bhm.2014.0087
- Koolen, M., van Gorp, J. & van Ossenbruggen, J. (2019). Toward a model for digital tool criticism: Reflection as integrative practice. *Digital Scholarship in the Humanities*, 34(2), 368-385. doi: 10.1093/llc/fqy048.
- Langtree, T., Birks, M., & Biedermann, N. (2019). Separating "fact" from fiction: Strategies to improve rigour in historical research. *Forum: Qualitative Social Research*, 20(2). doi: 10.17169/fqs-20.2.3196
- Moravec, M. (2017). Feminist research practices and digital archives. *Australian Feminist Studies*, 32(91-92), 186-201. doi: 10.1080/08164649.2017.1357006
- Nicholson, B. (2016). Digital research. In S. Gunn & L. Faire (Eds.). *Research Methods for History* (pp. 170-190). Edinburgh: Edinburgh University Press.
- Putnam, L. (2016). The transnational and the text-searchable: Digitized sources and the shadows they cast. *American Historical Review*, 121(2), 377-402. doi: 10.1093/ahr/121.2.377

- Reffle, U., & Ringlstetter, C. (2013). Unsupervised profiling of OCRed historical documents. *Pattern Recognition*, 46(5), 1346-1357. doi:10.1016/j.patcog.2012.10.002
- Saldana, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). London: Sage.
- Thompson, P., Batista-Navarro, R. T., Kontonatsios, G., Carter, J., Toon, E., McNaught, J., . . . Ananiadou, S. (2016). Text mining the history of medicine. *PLoS ONE*, 11(1), e0144717. doi:10.1371/journal.pone.0144717
- Toms, E. G., & O'Brien, H. L. (2008). Understanding the information and communication technology needs of the e-humanist. *Journal of Documentation*, 64(1), 102-130. doi:10.1108/00220410810844178
- Tummons, J. (2014). Using software for qualitative data analysis: Research outside paradigmatic boundaries. In M. Hand & S. Hillyard (Eds.), *Big Data? Qualitative Approaches to Digital Research* (pp. 155-177). doi: 10.1108/S1042-319220140000013010
- Underwood, T. (2014). Theorizing research practices we forgot to theorize twenty years ago. *Representations*, 127(1), 64-72. doi: 10.1525/rep.2014.127.5.64
- Wolff, R. S. (2013). The historian's craft, popular memory, and Wikipedia. In J. Dougherty & K. Nawrotzki (Eds.), *Writing history in the digital age* (pp. 64-74). Ann Arbor, MI: The University of Michigan Press. doi: 10.3998/dh.12230987.0001.001

Table 1

Summary of stages for conducting a scoping review for historical research

Stage	Features and Characteristics
<i>Identifying the research question</i>	<ul style="list-style-type: none">• Guides how the search and analysis will be conducted• Needs to be broad enough to summarise the breadth of evidence (e.g., What is known about...?)• Consideration points (e.g., place, event, time period, individual/group being studied)• Develop working definitions for each element of research question• Further refinements are made once historian gains a sense of the volume and contents of the text corpus
<i>Identifying relevant sources</i>	<ul style="list-style-type: none">• Development of a broad search strategy including planning for how sources will be managed and screened• Articulates which digital platforms will be used for searching (databases, online repositories/collections and webpages)• Development of search terms (e.g., key words, synonyms and variants; spelling variations; application of Boolean operators)• Development of eligibility criteria that will be used in future screening:<ul style="list-style-type: none">○ Types of sources included in study (e.g., Will discursive primary sources such as letters or journal entries be included?; How will secondary sources be used in the analysis?)○ Temporo-spatial considerations (e.g., applying filters to capture time- or geographic-specific materials)○ Language restrictions (e.g., Will foreign language sources be included?)
<i>Source selection</i>	<ul style="list-style-type: none">• Sources are screened using pre-determined eligibility criteria• Hand-searching of secondary sources to locate additional primary sources• Source criticism undertaken• Digitised copies of primary and secondary sources are stored in a reference management system (e.g., EndNote or RefWorks)• Consider using an independent reviewer for sources where there are queries about its eligibility or authenticity
<i>Charting the data</i>	<ul style="list-style-type: none">• Data is presented as a spreadsheet or table including:

- Bibliographical details (metadata) from eligible sources (title, source type, author, publication details - year, location, publisher, edition; language; archival details - repository and retrieval URL)
- Purpose and target audience
- Optional details include the physical features (dependent on the scope of study):
 - Front matter inclusions (dedications, acknowledgements, endorsements, prologue, table of contents)
 - Organisational layout (number of pages, indexing, appendices, addendums)
 - Typographical aspects (style – handwritten or typeset, use of drop caps, orthographic features, presence of typographical errors)
 - Other features (e.g., material, illustrations, handwritten annotations, source condition)

***Collating,
summarising and
reporting***

- Equates to analysis, interpretation and dissemination

***Consultation
(optional)***

- Seeking help to locate a source; translating a source; or clarifying an interpretation of the source

Adapted from Arksey and O'Malley (2005).