Historians' archival research looks quite different in the digital age
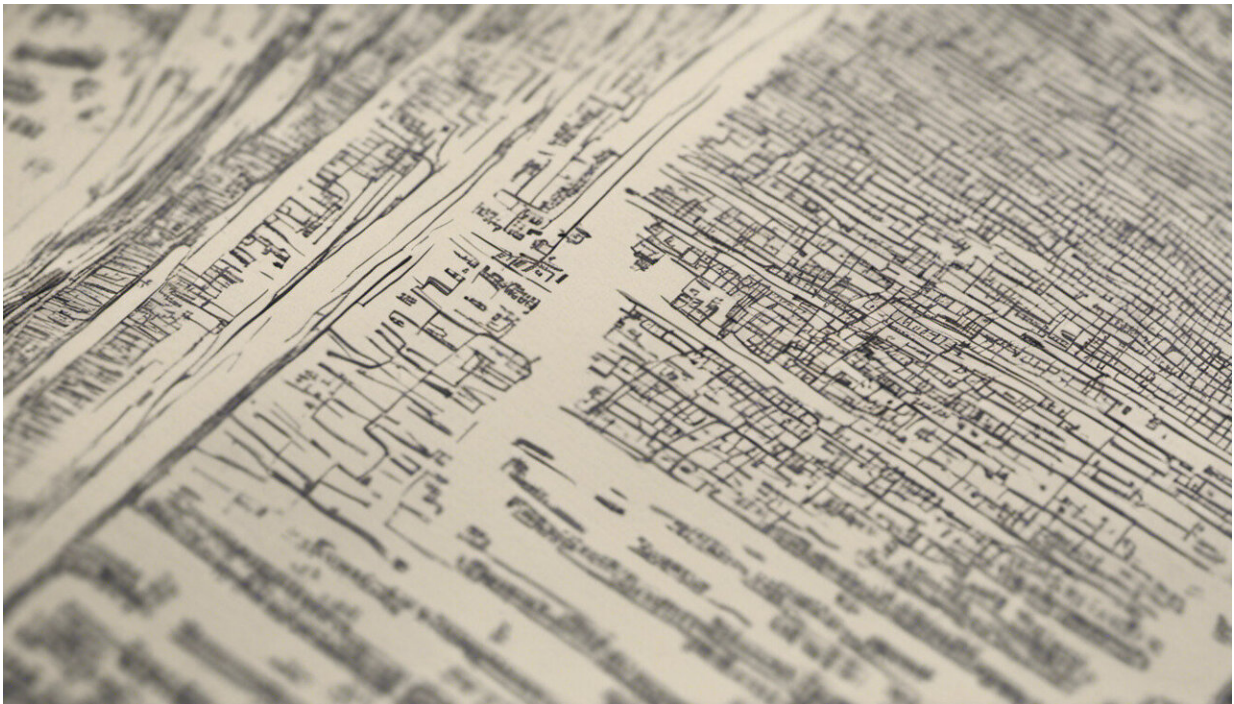
August 20 2019, by Ian Milligan

Credit: AI-generated image (disclaimer)

Our society's historical record is undergoing a dramatic transformation.

Think of all the information that you create today that will be part of the record for tomorrow. More than half of the world's population is online and may be doing at least some of the following: communicating by

email, sharing thoughts on Twitter or social media or publishing on the web.

Governments and institutions are no different. The American National Archives and Records Administration, responsible for American official records, "will no longer take records in paper form after December 31, 2022."

In Canada, under Library and Archives Canada's Digital by 2017 plan, records are now preserved in the format that they were created in: that means a Word document or email will be part of our historical record as a digital object.

Traditionally, exploring archives meant largely physically collecting, searching and reviewing paper records. Today, and into the future, consulting archival documents increasingly means reading them on a screen.

This brings with it opportunity—imagine being able to search for keywords across millions of documents, leading to radically faster search times—but also challenge, as the number of electronic documents increases exponentially.

As I've argued in my recent book *History in the Age of Abundance*, digitized sources present extraordinary opportunities as well as daunting challenges for historians. Universities will need to incorporate new approaches to how they train historians, either through historical programs or newly-emerging interdisciplinary programs in the digital humanities.

The ever-growing scale and scope of digital records suggests technical challenges: historians need new skills to plumb these for meaning, trends, voices and other currents, to piece together an understanding of

what happened in the past.

There are also ethical challenges, which, although not new in the field of history, now bear particular contemporary attention and scrutiny.

Historians have long relied on librarians and archivists to bring order to information. Part of their work has involved ethical choices about what to preserve, curate, catalog and display and how to do so. Today, many digital sources are now at our fingertips—albeit in raw, often uncatalogued, format. Historians are entering uncharted territory.

Digital abundance

Traditionally, as the late, great American historian Roy Rosenzweig of George Mason University argued, historians operated in a scarcity-based economy: we wished we had more information about the past. Today, hundreds of billions of websites preserved at the Internet Archive alone is more archival information than scholars have ever had access to. People who never before would have been included in archives are part of these collections.

Take web archiving, for example, which is the preservation of websites for future use. Since 2005, Library and Archives Canada's web archiving program has collected over 36 terabytes of information with over 800 million items.

Even historians who study the middle ages or the 19th centuries are being affected by this dramatic transformation. They're now frequently consulting records that began life as traditional parchment or paper, but were subsequently digitized.
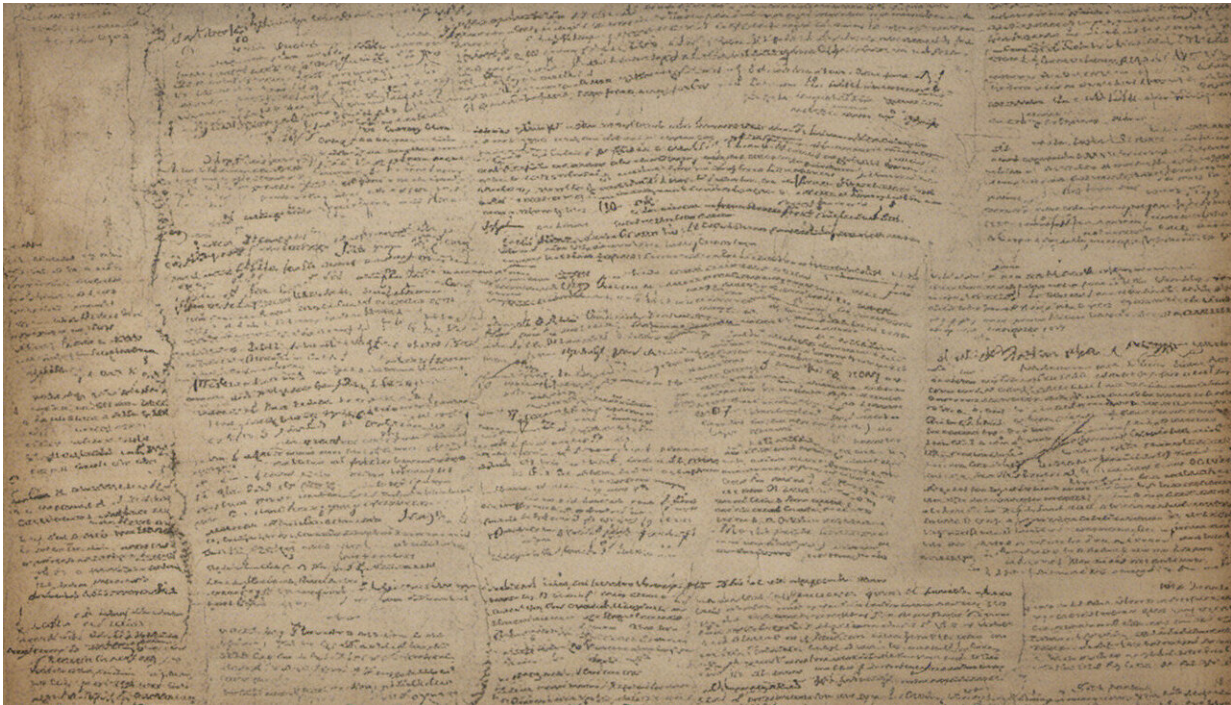
Credit: AI-generated image ([disclaimer](#))

Historians' digital literacy

Our research team at the University of Waterloo and York University, [collaborating on the Archives Unleashed Project](#), uses sources like the GeoCities.com web [archive](#). This is a collection of websites published by users between 1994 and 2009. We have some 186 million web pages to use, created by seven million users.

Our traditional approaches for examining historical sources simply won't work on the scale of hundreds of millions of documents created by one website alone. We can't read page by page nor can we simply count keywords or outsource our intellectual labor to a search engine like Google.

As historians examining these archives, we need a fundamental understanding of how records were produced, preserved and accessed. Such questions and modes of analysis are continuous with historians' traditional training: [Why were these records created](#)? Who created or preserved them? And, what *[wasn't preserved](#)*?

Second, historians who confront such voluminous data need to develop more contemporary skills to process it. Such skills can range from knowing how to take images of documents and make them searchable using [Optical Character Recognition](#), to the ability to not only count how often given terms appear, but also what contexts they appear in and how concepts begin to appear alongside other concepts.

You might be interested in finding the "Johnson" in "Boris Johnson," but not the "Johnson & Johnson Company." Just searching for "Johnson" is going to get a lot of misleading results: keyword searching won't get you there. Yet emergent research in the field of [natural language processing](#) might!

Historians need to develop basic algorithmic and data fluency. They don't need to be programmers, but they do need to think about how code and data operates, how digital objects are stored and created and humans' role at all stages.

Deep fake vs. history

As historical work is increasingly defined by digital records, historians can contribute to critical conversations around the role of algorithms and truth in the digital age. While both tech companies and some scholars have advanced the idea that technology and the internet will strengthen democratic participation, historical research can help uncover the [impact of socio-economic power throughout communications and media history](#). Historians can also [help amateurs parse the sea of historical](#)

[information and sources now on the Web](#).

One of the defining skills of a historian is an understanding of historical context. Historians instinctively read documents, whether they are newspaper columns, government reports or tweets, and contextualize them in terms of not only who wrote them, but their environment, culture and time period.

As societies lose their physical paper trails and increasingly rely on digital information, historians, and their grasp of context, will become more important than ever.

As deepfakes—[products of artificial intelligence that can alter images or video clips](#)—[increase in popularity online](#), both our media environment and our historical record will increasingly be full of misinformation.

Western societies' traditional archives —such as those held by [Library and Archives Canada](#) or the [National Archives and Records Administration](#) —contain (and have always contained) misinformation, misrepresentation and biased worldviews, among other flaws.

Historians are specialists in critically reading documents and then seeking to confirm them. They synthesize their findings with a [broad array of additional sources and voices](#). Historians tie together [big pictures and findings](#), which [helps us understand today's world](#).

The work of a historian might look a lot different in the 21st century —exploring databases, parsing data —but the application of their fundamental skills of seeking context and accumulating knowledge will serve both society and them well in the digital age.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).