Chapter 7

Digital History: The Globally Unequal Promise of Digital Tools for History: UK and Colombia Case Study

Adam Crymble

Maria José Afanador Llach

Abstract: This chapter examines the opportunities presented by new 'digital' history tools. It argues that data wrangling and data crunching allows historians to ask new questions that we have never been able to ask before, because we can now make sense of data that was previously inaccessible. The promise of big data and new digital tools has been one of the key conversations of the past decade. However, it's not a globally equal promise. It's come on the back of nearly two decades of mass digitization that privileged the historical sources of the West, and of Anglophone countries in particular. Those data were created predominantly through the cultural perspectives and technical infrastructures of those same privileged nations. The original records themselves were produced as the result of an advanced bureaucracy designed to administer a global empire. Thus, many of these new techniques are inaccessible to scholars studying or living in countries that today some people call the 'Global South'. Not only are the data not digitized, but sometimes the sources simply do not exist because of the historic nature of the archive: what was recorded, and what has survived. These deep digital divides are made worse by a lack of technical infrastructures in many countries. Even in the wealthy countries of the West, social barriers and the expectations of humanities students to be able to avoid mathematics and computers raise difficulties for educators attempting to help students make the most of the digital age. This chapter therefore takes a West/South look at the promises of the 'digital' age for history, through the different challenges faced in the UK and Colombia respectively, to show not only what the future might hold, but also where we still need to focus our energy in pursuit of a more equitable learning environment of the future.

6.1 Introduction

Every few years a new tech buzz phrase bursts onto the scene in historical studies, promising (for five to seven years) to revolutionize scholarship and teaching. Recently 'Big Data' and 'Machine Learning' have added to generations of technical promises made by early-adopting historians eager to play with some new toys with their students. Both of these new phrases are a collective shift of attention onwards from the buzzwords of yesteryear: Web 2.0, the Internet, CD-ROMs, new media, databases, word processors, statistical packages, calculators, overhead

projectors. Historians invented none of these, but as each appeared, scholars of the past found productive (and counterproductive) ways to apply them to the classroom and the research landscape. This was Silicon Valley influencing the most traditional of lecture theatres. Simultaneously and often more substantially, those theatres were also being influenced by other buzz phrases, including those coming from educational theorists: student centred learning, decolonizing the curriculum, and blended learning, for example (Purvis, 1983; King, 1993; Beichner et al., 2007; Risam, 2018). The history classroom was constantly evolving as a result.

The current promise is one of new possibilities for historical insights through digitally-enabled analysis of large sets of historical sources, and of students as researchers rather than consumers of other people's ideas. By wrangling data and crunching numbers, our students can find never-before-seen patterns in the historical record and generate new knowledge. Some may resist a view of historical studies as a predictive science, Peter Turchin among them (Turchin, 2015), but we hope readers may be more open to what Dominique Vinck calls a 'useful detour for qualitative approaches' that sees quantitative approaches as a helpful complement to the more literary forms of historical method (Vinck, 2018). Writing from experience, this detour can be productively travelled with the right data, the right mentorship, and the right students.

6.2 The example of *Alumni Oxonienses*

In 2017 at the University of Hertfordshire in England, some of Crymble's undergraduate digital history students came up with some outstanding insights into early modern migration by studying the origins of University of Oxford students in the sixteenth and seventeenth centuries (Chowcat, 2016). Using data management and geospatial analysis skills learned in the class, and armed with a reasonably clean set of historical data provided by their teacher, a number of students were able to identify clear changes in the places of origin of Oxford students in the years before and after the English Civil War, 1642-1651 (Crymble, 2015; Colson, 2017). These patterns have never been identified previously by historians. The most creative essays even managed to put forth compelling arguments for why those changes occurred and how we might interpret the different patterns in various parts of the country. From a teaching perspective, it was heartening to see such depth to the student's work, and it was interesting to read such unique insights in undergraduate essays.

This small group of British students of history were up to something quite remarkable. Yet, it is no accident that this learning opportunity took place in Britain and not Colombia, for

example. Not for reasons that British policy makers might like to suggest. It is not because Britain is more advanced, or that it has a better education system, or even greater computer literacy. The quality of the work coming out of Colombia challenges all of those assertions (Borja, n.d.). Some of the wealthier Colombian universities have facilities that would make many British institutions look impoverished by comparison. Instead, it is because history education around the world is still overwhelmingly nationalist in its approach, meaning that British history students still study a lot of British history. This is significant because it means British students study the relics of British culture, and British culture, as part of a Western Christian tradition, has for two millennia been obsessed with writing things down and keeping the results of that work. They are, as the Islamic scholars say, the people of the book. For centuries, they were frequently also the authors of those books.

The digital or digitized resources that form the basis of 'digital' analyses in the twenty-first century are a product of that written tradition of the previous half-millennium in the west. The huge global museum and archival collections that exist in Britain today at institutions such as the British Library or the National Archives, are the product of a culture of collecting and organizing in the nineteenth century, which evolved into microfilming, microfiching, and digitizing over the course of the twentieth century. Without those earlier stages, the latter ones may not have followed, and we may not today be talking about 'digital analyses' in history.

Those data that Crymble's students analyzed are a perfect example of this centuries-long process. The records had been written down hundreds of years ago by an army of porters fulfilling their daily duties at the front desks of more than a dozen historic Oxford colleges and halls. As good clerks, they patiently recorded what are now mini-biographies of more than 60,000 students who came past those desks on their way to a higher education. Five hundred years later, we are fortunate to still have the fruits of their labour, and the little nuggets of information that range from the student's name, to where he came from, who his father was, to when and where he studied, and sometimes even to what he went on to become. For example, famous philosopher and author of *Leviathan,* Thomas Hobbes' entry reads thus (expanded for readability):

> Son of Thomas, cleric; born at Westport, Wiltshire, 5 April, 1588; Bachelor of Arts from Magdalen Hall 5 February, 1607-8, incorporated at Cambridge 1608, "the famous philosopher and mathematician of Malmsbury"; died at Hardwick, county Derby, 4 December, 1679, aged 91('Hieron-Horridge', in *Alumni Oxonienses,* pp. 706-747.).

From 60,000 entries like this, students were able to create maps that showed changes in the place of origin of Oxford students over time which went far beyond anything we knew about early modern migration in Britain. But the number of things that had to happen for that to be possible is, quite frankly, remarkable. Not only did the porters need to be part of a tradition whereby their internal management structures required them to keep such written records, but we are fortunate enough that they worked at an institution that, for whatever reason, thought it would be a good idea to keep and preserve their work for five hundred years, with the costs associated with that preservation absorbed despite no obvious economic benefit of doing so either then or at any time in the future.

Then, in the nineteenth century, a man named Joseph Foster had to devote years of his life, again with little obvious economic benefit on the horizon, to verify, collate, alphabetize, transcribe, and publish those scribbles in the porter's ledgers into a six volume reference work known as the *Alumni Oxonienses*. It is Foster's work that gave us the mini-biography of Thomas Hobbes above. Foster knew with hindsight that Hobbes would become a 'famous philosopher and mathematician'; the porter likely had no such inkling (Foster, 1891).

More than a century after that, just shortly after the new millennium, the team behind the newly created *British History Online* digital library identified the *Alumni Oxonienses* as an important enough work to raise money to include it in their collection. As a new digital memory institution, *British History Online's* team played a crucial role in the process that led to converting these written records into machine readable data for historians to begin to explore computationally. They spent considerable time retyping the six volumes out in full (twice to ensure accuracy), rather than rely upon mechanised processes that they worried would not produce as good a result. They then made those records available online as machine readable text, and gave it away for free. This particular digitization project was funded by the American Friends of the Institute of Historical Research (where *British History Online* is based), raised by donations. The digital library itself was funded principally by the Andrew W. Mellon Foundation, an American philanthropic organization committed to the arts and humanities, and it was one of dozens of similar digitization projects supported by the Foundation in the late 1990s and early 2000s. The decisions about what to digitize were made largely at the whim of these and similar funding bodies, with their representatives sitting in cities such as New York or London, many of whom had Western upbringings that fed their sense of purpose or urgency when deciding whether to fund one project or another. The decisions they made around those board tables have sculpted historical studies a generation later by deciding what would be available as digital data, and what would not.

Thus, for the students to conduct their research on Oxford students in the sixteenth century, various people had to create, preserve, reformat, and reformat the records again. A hidden collaboration over hundreds of years, at each stage costing the better part of a small fortune to enable. This same process happened again and again in countries like Britain, preserving billions of bits of paper with ideas, which centuries later thanks to a generation of work devoted to converting that paper into machine-readable text, gave students an opportunity to data mine in their history classrooms. For those of us who live in these Western countries that benefited from this extraordinary series of unlikely events leading us to the present situation, it seems an entirely normal progression of a learned society. But this is not normal; it is utterly unusual - the product of a very particular type of culture that was not replicated frequently around the globe. Instead, this is a culture that honed its record keeping practices during centuries of building and administering a global empire. The paper trail of that empire enabled a small group of people in European capitals such as London to exert control around the world. It has left hundreds of kilometres of shelves across Europe that are filled with that administrative paper trail, many of which are now ripe for data mining.

6.3 Digital Resources not in Britain

But what of the ruled rather than the rulers? What if you do not own your people's written record? Or, how do you mine ideas that were not kept on paper at all? Not part of the literary tradition? It is easy to forget, but this was most of the world - even parts of it we now revere as learned. The ancient Greeks had writing but used spoken verse as a means of storing and transmitting epic poetry. The Inca of South America, and the Aborigines in Australia were part of oral cultures, as were countless other civilizations. How do you data mine these traditions? So far, you can't - at least not at scale. The promise of data mining and machine learning have been over-egged because they are based on an assumption that the entire world collected and continues to collect information in a similar way. This over-egging is merely a repeat of the hyperbolic technological promises that came before them. Data mining and machine learning will make a big impact on aspects of historical teaching and scholarship, but not one that will touch all parts of the globe equally.

6.4 Colombia

Even in places with long histories of written culture such as Colombia, which was first visited by the Spanish at the dawn of the 16th century, they still find themselves in many ways hindered in the pursuit of big data analysis by decisions made centuries earlier. The country's colonial past is part of that hindrance. The territories that comprise Colombia today were, for over 300 years, colonial territories claimed by the Spanish monarchy. The Iberian empire built a complex bureaucratic machine made of papers produced on both sides of the Atlantic. The making of empire created one of the largest colonial archives ever written, which recorded information about former colonies across Latin America and the Philippines. Today, most of that archive sits in Spain at the Archivo de Indias de Sevilla (AIS). This puts legal title of some of Colombia's most important historical documents with the Spanish state. Even if Colombians did want to digitize and make these resources available for data mining, they do not have the legal means to force Spain's hand. This raises questions about who owns digital cultural heritage. A global legal framework designed to protect intellectual property, developed in the latter early modern period, has yet to come up with satisfactory solutions to shared cultural heritage in the digital age. A handful of transnational projects, such as *Aluka* (Rüther, 2003), an online digital library of African heritage, are leading the way in seeking such solutions, but national and international laws remain a barrier for many seeking access to their heritage.

This remains a crucial problem to solve, according to the postcolonial view of historical studies in the digital age. As Roopika Risam has suggested, 'Within colonized and formerly colonized nations and for people outside of dominant cultures, access to means of digital knowledge production is essential to reshaping the dynamics of cultural power and claiming the humanity that has been denied by the history of colonialism' (Risam, 2018). Colombians wanting to use many of their own colonial era materials are thus at the mercy of Spanish policies and decisions about digitization, many of which were made in the late 1980s and early 1990s by people who did not foresee a future when people would expect to be able to download and manipulate historical data (Sánchez Mora, 2017). While a huge number of documents are available online via AIS, the infrastructure of the archive cannot at present facilitate complex data analysis, and there is nothing Colombians can do about it.

The case in regional Colombian archives is even less promising. Many of those archives hold valuable colonial documents, but their situation is precarious. Low budgets, poor enforcement of archival legislation, conservation problems, and the lack of catalogues has endangered the preservation of many historical archives across the country. Here, organisations such as the British Library play a vital role in preservation through their Endangered Archives Programme, making available resources to protect and digitize materials that are at risk of

destruction, but even they do not have enough money to solve this global problem (British Library, 2004-2019). Some successful locally-led digitization projects, such as Colombia's *Neogranadina*, stand out as shining examples of home-grown initiatives that provide access to important colonial Colombian materials; nevertheless, there is a long way to go before Colombians have a similar level of digital access to their history (Fundación Histórica Neogranadina, 2019).

Colombians also sit on the outside of a global culture that remains English-dominated in a number of ways, and has been for a long time. English viewpoints persist in certain corners of the Colombian archive, both through what was originally created, and what has been made available in digital form. For example, there are English-drawn maps of early Colombia: a 1772 map of Cartagena by I. Andrews, as well as a series of English-language maps of the whole Colombia country by John Cary (1813), Aaron Arrowsmith (1814), John Finlayson (1822), and Sidney Hall (1828). As far as we are aware there are no early Colombian-drawn maps of England. Thus we have English eyes looking at Colombia, and nothing looking the other way, and the histories we write are sculpted in that image. We might call this the informal empire of Britain, extending intellectually even into territories that were never directly under their political control.

The English language also rears its head in many of the programming languages that scholars need to learn in order to work with big data. The 'reserved words' of programming languages that must be typed to perform various actions and commands, are all in English. The Python programming language, popular with many digital historians, is a great example of the English barrier to entry. Its commands include 'and, as, assert, break, class, continue, else, except, False, finally, for, from, global, if, import, in, is, None, not, or, pass, raise, return, True, try, while, with, yield.'

An Anglophone who has been introduced to programming concepts may be able to puzzle out what these commands might do, or can at least use their English language skills to remember them in meaningful ways. A Spanish speaker has no such advantage. There are no ways to substitute Spanish language commands. An 'if statement' in Spanish is still 'if statement' rather than something more vernacular such as 'si la declaración'.

What was digitized also sculpts those histories we tell today. The early priorities of commercial digitization companies based in the Global North had an impact on the development of digital skills in the Global South (Kenney et al, 2000; Putnam, 2016). These companies built many of the digital archives of primary sources in the early new millennium, access to which universities procured through expensive subscription fees. These projects

tended to focus on English-language materials, which could more easily be marketed to the larger American higher education industry. While these companies have begun to diversify their offering and to include materials in different languages or created in different parts of the world, countries like Colombia were for years faced with a choice: subscribe to English-language archives, or do without the cutting edge resources. This put fluency in English (itself a sign of privilege in many countries) as a key barrier in front of Latin American historians looking to work with many digital materials, and means any insights made using new methods such as data mining, will tell us only about the English speaking world. This is part of a larger history of colonization that reflect some biased historiographical tropes and issues of digital divides and open access, among others (Fiormonte, 2016).

Cultures of academic publishing in different countries also impact the type of work scholars engage with. Colombia, like many countries, tries to encourage scholars to up their game by providing a whitelist of journals in which they encourage researchers to publish, and reward them if they do so. These so-called prestigious journals are used as a measure of Colombian scholarship's march towards progress on the international stage - perhaps itself a hangover of colonial mentalities. But there are unintended consequences of journal whitelists. As Rafols et al. have noted, interdisciplinary research is often less well suited to these more prestigious and more traditional journals. That means that research employing newer methods such as machine learning and data mining may not find natural homes amongst the approved journals and scholars are thus dissuaded from engaging in that type of work (Rafols, 2012). Rafols pointed out in particular that countries such as Colombia, which may benefit most from interdisciplinary approaches to the problems they face, are particularly hit hard by this anti-interdisciplinary policy.

As a result of these and other factors in Colombia and other Latin American countries, the type of 'digital' work being done in the region differs from the priorities of the Global North. Colombia is developing a digital approach that's best suited to Colombia and the wider societal and professional pressures in the region. Sometimes that aligned with what was going on elsewhere, and sometimes it did not. Tech buzz phrases may get attention in the press or on social media, but priorities for cultural memory organisations in countries such as Colombia are not always aligned with the needs or interests of researchers and teachers hoping to work at the cutting edge of technology and history. That is not a sign of a deficiency in the archivists and librarians, but a different awareness of where scarce time and resources should be focused to ensure the greatest overall benefit - and an approach which mirrors that of most cultural memory organizations in the Global North. As the devastating 2018 fire at Rio de Janeiro's

National Museum or the 2019 fire at Notre Dame in Paris showed, without robust preservation strategies, all access is lost to both people and machines.

Thus, to date much of the focus of Latin American cultural memory organisations has been on accessibility and conservation rather than data mining, and Colombia is no different. This has certainly been the strategy of the Biblioteca Nacional of Colombia (National Library), which has created a 'Biblioteca Digital' website, as well as the Archivo General de la Nación (National Archives) and the Biblioteca Virtual del Banco de la República (Red de Bibliotecas del Banco de la República, 1997). All of these memory institutions have digitized a number of important historical collections. These include manuscripts, newspapers, engravings, literary works, photographic archives, and maps among others. All three of these collections have been produced with human readers rather than machine readers in mind. And why not? More Colombians can read text than write computer programmes, and many types of research and teaching are only possible with human-readable source material. The preservation and access approach is therefore the democratic and inclusive choice, even if it means few Colombians are able to harvest the benefits of robust data mining in historical studies the way their cousins in the North can.

Colombia's pathway into its own brand of digitally enabled historical work often relied upon the hard work of early adopting individuals working in near isolation. Despite the challenges, there are pockets of activity involving scholars seeking to build robust digital data about the past, which may yet find their way into more Colombian history classrooms. As is often the case in the Global North, these pockets develop where someone shows initiative and enthusiasm. Historian Jaime Borja at the Universidad de los Andes in Bogota is one of those enthusiastic pioneers of robust digital data. His database, *Arte Colonial Americano* (ARCA), contains more than 20,000 images and accompanying enriched metadata of American colonial paintings created between 1530 and 1830 (Borja, n.d.). Like many digital archives, digital copies of the paintings are accessible for users to look at. The metadata makes it possible to analyze the relationship between body and gesture in colonial paintings, for example. However, ARCA has gone further, building a series of 'big data tools' into the site to make analysis even easier. This includes a dashboard that allows users to explore patterns in geography, theme, artists, or a combination of several factors. Also available is a network graphing tool designed to highlight connections between bits of information in the database. The findings of that type of query might not be obvious to a student taking a traditional approach to looking at the paintings, meaning ARCA provides opportunities for educators and students to build new knowledge about the Latin American past.

Teaching too has been at the impetus of the adventurous. Early in the new millennium, Stefania Gallini at Colombia's National University pioneered some of the country's first forays into digital work in history classrooms. Gallini found an open and flexible history department that let her create and teach digitally-focused courses as part of the methodology offering in the history program. It started with a proposal to teach a course called 'Internet and History' back in 2002. Drawing her educational experience from abroad in the United Kingdom and Italy, her offering was distinctly in line with courses offered in the Global North at the same time. From Michael John Gorman's 'The Wired Historian' (Gorman, 2002) at Stanford to Ian Archer's 'Internet Data Course' (Archer, 2002) at the Institute of Historical Research in London the same year, Gallini's teaching shows that Colombia was never behind when it came to teaching technology and history together, and that there were always people looking beyond the seas for inspiration on how to change Colombian history classrooms. Her dedication to this new interdisciplinary approach to teaching expanded in 2013-2015, with the creation of *Laboratorio de Cartografía Histórica e Historia Digital* CaHID (Lab of Historical Cartography and Digital History), a physical space with computers, scanners, and software that students could use for experimenting with digital technology. What she created in Colombia was in many ways similar to digital history labs elsewhere, including William J. Turkel's 'Lab for Humanistic Fabrication' at Western University in Canada, Mike Cosgrave's digital learning environment at University College Cork in Ireland, or Stanford's 'Center for Spatial and Textual Analysis' in the United States - all operational at the same time as Gallini's own space. Gallini's lab was quickly appropriated by students who collectively created the rules for using the space and began experimenting with digital history projects. Very soon some students began graduating with theses that involve using digital media such as blogs or digital cartography (Díaz Ángel, 2007-2019). The emphasis on 'space' is perhaps key to understanding digital methods in Colombian historical studies. Faced with a relative lack of 'big data' in the archive, the land provided one of the best measurable sources of new data. As David Bushnell's *The Making of Modern Colombia* starts: 'In the beginning there were mountains, plains, and rivers, but especially mountains; no one geographic feature has so moulded the history of Colombia as the Andes' (Bushnell, 1993).

Drawn by the power and influence of the landscape, many key Colombian works have turned to spatial analysis for inspiration. For instance, Natalia Jaramillo, a Colombian graduate student with an interdisciplinary background in history and geography, is using drones to generate a topographic map of the Media Luna (Half Moon) island in Antarctica. Much like Google Earth, she uses software to stitch together hundreds of photographs, to produce the first

new map of the island in half a century (Wright, 2019; Jaramillo, 2016-2019). Afanador's own work on Colombia's transition from colony to republic in northern South America also takes a geographical approach, seeking to structure geospatial and economic data in small and medium size datasets from primary sources such as maps, economic texts, and geographic descriptions produced by bureaucrats, military men, clerics, and naturalists. The aim of the project is to experiment with data modeling and visualize the spatial dimensions and conflicting visions of territory and political economy, thus contribute to the understanding of conflicts that arise from such historical tensions. Big data may not be available for these Colombian scholars, but innovative approaches to what is available is certainly not lacking.

A lack of access to robust machine-readable data in textual format was probably a key reason why spatial humanities became so popular amongst the technologically inclined in Colombia. Data mining and machine learning typically depend on large sets of clean and consistent data, often in textual form (although algorithms are changing and this may not always be the case). Colombians have not traditionally had many home-grown datasets of historical materials that would facilitate this type of work. Infrastructural, economic, historiographic, and institutional challenges needed to be overcome before the conditions were right for building that data from the surviving archival materials. The process of creating reliable data also takes time and money, and practitioners must navigate local cultures of collecting that have been embedded in libraries and archives in the region far longer than the term 'digital' has been on the lips of those in the Global North.

Far more common than 'big data' for Colombian historians, has been the productive pursuit of public history. In 2016, the *Third International Public History Conference* was held in Bogotá, Colombia, bringing together national and international scholars (Universidad de los Andes, 2016b). Shortly after that, the project *Historias Para Lo Que Viene* was born after the majority of Colombia voted not to support the implementation of the peace with the guerrillas in 2016. The project was an invitation to create open discussions about Colombian history as a strategy to support the understanding of the current situation of the country and the need to tell stories of conflict after years of violence (Universidad de los Andes, 2016a).

The conflict has inspired other historically-themed projects. Sometimes interest in the land came together with public history when historians turned their minds to restitution and reconciliation. Conversations about land use and agency in Colombia have been at the heart of the country's attempts to overcome the fifty-years of armed conflict. Although the lack of an inventory of the country's land is a pressing issue, some advances have been made in building databases that document the armed conflict. This line of research has been carried out by the

National Center of Historical Memory in their Observatory of Memory and Conflict, the most information-intensive collection about the Colombian armed conflict. The project's creators have integrated more than 10,000 databases and documents containing information about the time and place of violent acts, as well as details of both victims and perpetrators of violence (Centro Nacional de Memoria Histórica, 2019). Thousands of data points about the armed conflict are now available in open access forms thanks to this initiative, opening up possibilities for historians by creating new sources put together for the express purpose of understanding Colombia's recent past (Universidad Nacional de Colombia, 2013-2019).

Along with these initiatives, historians have likewise been collaborating with mainstream media to create digital public history projects. At the same time, some institutional spaces have started to favour interdisciplinarity, despite the barriers of journal whitelists already mentioned (Villamizar et al., n.d.). When Universidad de los Andes decided to develop a master's degree in Digital Humanities in 2016, it chose to sit the programme at the intersection of several departments where students could learn from a range of different perspectives. Historians came together with musicians, journalists, artists, art historians, literature scholars, designers, social scientists, education experts, and computer scientists. The emphasis was on the power of interdisciplinarity itself, rather than an assumption that a few tech buzz phrases would provide Colombians with the answers it needs for the future. Perhaps tellingly, the history students who have enrolled on the programme have tended to favour digital public history projects over data-intensive ones (Facultad de Artes y Humanidades, Universidad de los Andes, 2016).

6.5 Conclusion

This leads us to the question: what does Colombian historical studies need to make the most of a digital future suited to its own needs and goals? We firmly believe that Colombians are best placed to tell us about their needs, and that building an understanding of the types of skills Colombians are currently trying to learn provides us with a clear window into the types of support they will need in the near future. One way to measure that need is through web traffic reports of the *Programming Historian* project, a peer reviewed digital learning resource published in English, Spanish, and French. The project's more than 150 tutorials are aimed at humanists interested in learning technical skills. It gets more than one million visitors per year around the world, with a growing number of readers in Latin America.

An analysis of the traffic logs shows that Latin American-based readers are seeking substantially different digital skills from readers of the English publication, or from readers based in Spain. In the twelve months following May 2018, English readers were more likely to access lessons that taught analysis skills that could be used to produce quantitatively-based historical research findings. In particular, lessons that teach skills that we might describe as 'data mining' or analysing large volumes of materials, were much more popular with English readers than those from other language groups. This included lessons on getting started with topic modeling, corpus linguistics basics, and downloading large numbers of webpages semi-automatically - for example to download a large database one page at a time for offline analysis (Graham *et al.,* 2012; Milligan, 2012; Froehlich, 2015). Given the buzz around data mining in the English press, this shows an alignment between English buzz and English skills seeking.

By comparison, readers in Latin America were more likely to be reading basic programming and computing skills, such as how to manipulate text or files with the Python programming language (Turkel, 2017a; Turkel, 2017b). Also popular in Latin America were lessons on digital mapping - again reinforcing the connection to studying space and place rather than text (Clifford et al., 2018). Geography proved important for understanding Latin American needs, because the lessons that were popular in South America differed from those widely accessible in Spain. For example, one of the most popular lessons in Spain was a tutorial on Linked Open Data - one of the buzz phrases of the decade that promises to enable you to access well-structured information from library, museum, and archival collections, for use in web-based tools, apps, or analyses. It's the same type of technology that lets app developers create a tool that mashes together bus and train timetables with satellite data to tell you the quickest route between two places in real time. But in order to be useful, you need access to lots of different well-structured datasets, and at the time of writing, those are much more common in Europe than in South America. The *Programming Historian* lesson in question uses the data from a European collection (Lincoln, 2017). Nothing remotely resembling the same scale or quality exists in South America, and thus it is not surprising that Spanish readers have different digital skills and interests than their cousins to the south. Self-learners are voting with their clicks, and providing an insight not only into what they want to learn, but how educators can best support the different regional technical needs in different parts of the world. English speaks in a loud voice, but we must remember to listen carefully to the diverse and global needs of learners seeking to make the most of the digital age.

This chapter has explained why this conversation about data analysis and history is happening in places like the UK, but not in Colombia. We have argued that the reasons are

nothing to do with computers, and everything to do with history. Big data and machine learning are therefore poised to have a big impact on historical studies. But it will be a measured impact, and one that does not affect the whole world equally. We as historians must thus take care to ensure that too much faith is not put into it, and that we collectively look at a globally relevant set of new skills for students of the past. To peg data mining and machine learning as the only next big thing is to ignore the contexts facing countries of the Global South, and to suggest that their problems are less advanced, or not important. Instead, we believe that digital technology can and will continue to transform the way we teach and learn history in the Global South, but that it will do so in a way that is most appropriate for the educational and archival context in which those countries operate. It's a different path, and one we must all support as it develops.

As long as historians keep that in mind moving forward, we are confident that an even greater proportion of the next generation of history students will be poised to tackle the buzz phrases of tomorrow. We think that the next buzz phrase should have something in it about globally equitable approaches to understanding and solving the challenges facing the discipline. Instead, we have a feeling it will have something to do with robots. We hope we're wrong.

References

Foster, J. (Ed.), (1891). *Alumni Oxonienses 1500-1714*. (1891). University of Oxford. *British History Online*, http://www.british-history.ac.uk/alumni-oxon/1500-1714.

Andrews, (cartographer). (1772). *A plan of the city of Cartagena* [map]. Biblioteca Luis Ángel Arango, Bogotá, Colombia, Sala de Libros Raros y Manuscritos, signatura: PH0062.

Anon. (n.d.). *Bicentennial of independence special*. https://www.eltiempo.com/colombia/otras-ciudades/especial-sobre-bicentenario-de-colombia-2019-368714.

Archer, I. (2002). Internet Data Course. *Institute of Historical Research*.

Arrowsmith, A. (cartographer). (1814). Outlines of the physical and political divisions of South America: Delineated by A. Arrowsmith partly from scarce and original documents, published before the year 1806 but principally from maps & surveys made between the years 1771 and 1806, corrected from accurate astronomical observations to 1810 [map]. Archivo General de la Nación, Bogotá, Colombia, Sección Mapas y Planos, *Mapoteca 4,* Ref. x 68.

Beichner, R. J., Saul, J.M., Abbott, D.S., Morse, J.J., Deardosrff, D., Allain, R.J., Bonham, S.W., Dancy, M.H. & Risley, J.S. (2007). The student-centred activities for large enrollment undergraduate programs (SCALE-UP) project. *Research-Based Reform of University Physics* 1, 2-39.

Borja, J., Guerrero, K., Villaveces, S. & Murcia, F.J. (n.d.) *ARCA (arte colonial americano)*. http://artecolonialamericano.az.uniandes.edu.co:8080.

British Library, (2004-2019). *Endangered archives programme*. British Library. https://eap.bl.uk/project/EAP334.

Bushnell, D. (1993). *The making of modern Colombia.* University of California.

Cary J. (Cartographer). (1813). *Viceroyalty of New Granada, and government of Caracas* [map]. Biblioteca Luis Ángel Arango. Bogotá, Colombia, Sala de Libros Raros y Manuscritos, signatura: H220.

Centro Nacional de Memoria Histórica. (2019). *Observatorio de memoria y conflicto (OMC).* http://centrodememoriahistorica.gov.co/observatorio/.

Chowcat, I., Colbron, K., & Crymble, A. (2016) Digital in the undergraduate history curriculum. *Jisc: Content and Digitisation.* https://digitisation.jiscinvolve.org/wp/2017/02/21/digital-in-the-undergraduate-history-curriculum-a-spotlight-on-the-digital-case-study/

Clifford, J., MacFadyen, J., & Macfarlane, D. (2018). Introducción a Google Maps y Google Earth. Traducido por Andrés Gattinoni. *The Programming Historian en español* 2.

Colson, J. (2017). Geocoding historical data using QGIS. *Programming Historian* 6.

Crymble, A. (2015). Using gazetteers to extract sets of keywords from free-flowing texts. *Programming Historian* 4

Díaz Ángel, S. et al. (2007-2019). *Network of the history of geographies and cartographies of Colombia*. www.razoncarografica.com.

Facultad de Artes y Humanidades, Universidad de los Andes. (2016). *MA in digital humanities*, Universidad de los Andes. https://hd.uniandes.edu.co.

Finlayson, J. (Cartographer). (1822). Geographica, statistical and historical, map of Colombia [map]. Archivo General de la Nación, Bogotá, Colombia, Sección Mapas y Planos, *Mapoteca 4,* Ref. x 68.

Fiormonte, D. (2016). Toward a cultural critique of digital humanities. *Debates in the Digital Humanities.* University of Minnesota.

Foster J. (Ed.). (1891). *Alumni Oxonienses 1500-1714*. University of Oxford. In *British History Online*. http://www.british-history.ac.uk/alumni-oxon/1500-1714.

Foster, J. (Ed.). (1891). 'Hieron-Horridge'. *Alumni Oxonienses 1500-1714* (Oxford), 706-747. In *British History Online.* http://www.british-history.ac.uk/alumni-oxon/1500-1714/pp706-747.

Froehlich, H. (2015). Corpus analysis with AntConc. *Programming Historian* 4.

Fundación Histórica Neogranadina. (2019). *Neogranadina, catalogación colaborativa*. www.neogranadina.org.

Gorman, M. J. (2002). The wired historian. Stanford University.

Graham, S., Weingart, S. & Milligan, I. (2012). Getting started with topic modelling and MALLET. *Programming Historian* 1.

Hall, S. (Cartographer). (1828). Colombia [map]. Biblioteca Luis Ángel Arango. Bogotá, Colombia, Sala de Libros Raros y Manuscritos, signatura: H4.

Foster, J. (Ed.). (1891). Hieron-Horridge, in *Alumni Oxonienses 1500-1714*, pp. 706-747. *British History Online*http://www.british-history.ac.uk/alumni-oxon/1500-1714/pp706-747

Jaramillo, N. M. (2016-2019). *Maptartic drone project.* https://maptarcticproject.wordpress.com/.

Kenney, A.R., Speiss, K.P., Crew, S.R., Smith, A., & Reilly, B. (2000). Collections, content, and the web. *Council on Library and Information Resources.*

King, A. (1993). From sage on the stage to guide on the side. *CollegeTeaching,* 41, 30-35.

Laboratorio de Cartografía Histórica e Historia Digital. (2019). *Archivo digital*. https://www.archedi.co/hechos.

Lincoln, M. (2017). Uso de SPARQL para acceder a datos abiertos enlazados. Traducido por Nuria Rodríguez Ortega. *The Programming Historian en español* 1.

Milligan, I. (2012). Automated downloading with WGET. *Programming Historian* 1.

Purvis, K. (1983). The teacher as moderator: A technique for interactional learning. *ELT Journal,* 37(3), 221-228.
Putnam, L. (2016). The transnational and the text searchable: Digitized sources and the shadows they cast. *American Historical Review,* 121(2), 377-402.
Rafols, I., Leydesdorff, L., O'Hare, A., Nightingale, P. & Stirling, A. (2012). How journal rankings can suppress interdisciplinary research. *Research Policy* 41(7), 1262-1282.
Red de Bibliotecas del Banco de la República. (1997). *Biblioteca Virtual Banco de la República*. http://www.banrepcultural.org/biblioteca-virtual.
Risam, R. (2018). *New digital worlds*: *Postcolonial digital humanities in theory, practice, and pedagogy.* Northwestern University.
Rüther, H. (2003). *Aluka*. https://www.aluka.org/.
Sánchez M., A. (2017). Digitize documents in the Archivo General de Indias: Technical Advantage to Preserve a Historical Legacy.
Sánchez, S. (2018). Mujeres de prensa. http://pasadoimpreso.uniandes.edu.co/.
Turchin, P. (2015). Cliodynamics: History as science. http://peterturchin.com/cliodynamics/
Turkel, W. J. & Crymble, A. (2017a). Manipular cadenas de caracteres en Python. Traducido por Víctor Gayol, *The Programming Historian en español* 1.
Turkel, W. J. & Crymble, A. (2017b). Trabajar con archivos de texto en Python. Traducido por Víctor Gayol, *The Programming Historian en español* 1.
Universidad de los Andes. (2016a). *Historias para lo que* viene. https://histparaloqueviene.wixsite.com/histparaloqueviene.
Universidad de los Andes. (2016b). *International Federation for Public History*. from: https://ifph2016.uniandes.edu.co/.
Universidad Nacional de Colombia. (2013-2019). *Laboratorio de Cartografía Histórica e Historia Digital*. http://www.humanas.unal.edu.co/cahid/.
Villamizar, J. C., Gallini, S., Murcia G., Ramírez, C. & Bernal, A. et al. (n.d.). La historia presidencial. https://www.eltiempo.com/datos/historia-de-todos-los-presidentes-de-colombia-251902.
Villamizar, J. C., Murcia G. & Ramírez, C., Bernal, A. et al.(n.d.).
https://www.eltiempo.com/colombia/otras-ciudades/especial-sobre-bicentenario-de-colombia-2019-368714)
Vinck, D. (2018). *Humanidades Digitales. La cultura frente a las nuevas tecnologías.* Gedisa Editorial.
Wright, A. (2019, September). This Colombian aimed to be an airline pilot, now she's mapping Antarctica. *Forbes.* https://www.forbes.com/sites/andrewwight/2019/09/08/this-colombian-aimed-to-be-an-airline-pilot-now-shes-mapping-antarctica.