Entrevista com Matthew Connelly[1]

Interview with Matthew Connelly

Entrevista con Matthew Connelly

Concedida a

Arbel Griner[I]

Celso Castro[II]

[1] Matthew Connelly is Professor of international and global history at Columbia University, New York.

[I] Arbel Griner is a Postdoctoral Research Associate in Princeton University's Global Health Program. New Jersey.

[II] Celso Castro is full professor and the Dean of CPDOC, the School of Social Sciences of Fundação Getulio Vargas (FGV) Rio de Janeiro (RJ), Brazil.

Entrevista concedida em 20 de maio de 2019, em Nova Iorque.

CC: We would like to ask you about your family and the place you were raised, and about your studies before the university.

MC: I come from a big family. I am the youngest of eight children. And all my family is originally from Ireland. My mom was raised in Ireland. So, growing up, I had a bit of that kind of second-generation immigrant experience, where I was often going back to Ireland with my mother and had more relatives in Ireland than I did in the U.S. I think that background did – in a way that I did not even realize at the time – influence the kind of History that I was interested in. I read lots and lots of History. I was not necessarily a very good student [laughter]. It was only when I got to only read History, as I started to do at university, that I started to do better. And, I think, more even than being at Columbia, as an undergrad here, it was going to Cambridge for a year in the U.K. that had a big impact. Whereas, growing up, as a second-generation immigrant family, if you wanted to be successful, you wanted to become a lawyer, like my father, or you would want to be a doctor or something. I did not know anybody who was a professor. So, for the first time doing those one-on-one tutorials, I finally felt that I got to know a historian, and I began to think maybe this is something I could do. So, it was only then, and this was my third year at university, that I started to think I might want to become an historian and not be a lawyer. Right after university, I worked for about a year and a half as a journalist here in New York. I worked at a very small publication on a very obscure subject, which was the health effects of electromagnetic radiation from things like powerlines, and cellphones, and so on. So, it was interesting to have to write on deadline. I learned how to write quickly when I had to, which is useful. And I learned a bit about the history of science and regulation and such. But what I really wanted to do was to get back into university, and I did that. I went to Yale. At that time, I was still not completely certain that I wanted to be an academic. And so, I thought that, like some professors, it might be possible both to become an historian but also, maybe, to have some role in foreign policy.

CC: Why is it you chose Yale, instead of staying at Columbia, for example?

MC: Because there is a professor at Yale named Paul Kennedy. And back when I was figuring out where to go to graduate school, he had just written a best-selling book called *Rise and Fall of the Great Powers*. It was striking, to me, how here was somebody who already had had a very impressive career as an historian, but he was also having an impact in how people were debating what was going to happen after the end of the Cold War. Would the United States be the world's only superpower, or were other powers already rising to challenge it? And so I thought that would be the kind of career I would love to have: to be an historian but also feel that I could have some impact in the world.

This was the 1990s. They were only then declassifying things from the 1950s. Whereas we now think of it as the height of the Cold War, it was already a time in which even cold

warriors like John Foster Dulles and Dwight Eisenhower were starting to think of the world in terms of North and South, especially when it came to a place like North Africa and France's colony in Algeria. It seemed that it did not really fit into this kind of Cold War framework. There was an insurgency there. One that was not clearly communist inspired. But it was one that was challenging America's Cold War alliance with France. And it seemed, at times, that the whole future of the Atlantic Alliance was at stake if France was to pull out of NATO because of the way the U.S. was not supporting them in North Africa. So, for me, I thought it would be an interesting project to do an international history of that war, as a way of understanding how, even in the 1950s, the Cold War was becoming more complex as people were thinking about North-South issues, like relations between poorer countries with rapidly growing populations and declining powers, former colonial powers, like France.

I worked for several years at Yale to finish that dissertation. And I was fortunate because I decided to do that right from the outset, and so I had a long time to think about it. By the end of it, I was thinking about the Algerian war as a way of understanding other things, bigger things like demographic trends, changes in new means of communications, and so on.

CC: At that time, were you already identifying yourself as a global historian? Or that label came later?

MC: Yes. I was very struck when I was first doing my reading for orals, an exam where you have to be able to talk about 200 different books and articles. I was reading up on the literature on the history of U.S. foreign relations. And there was a long-running debate about if you wanted to write about the history of American foreign policy, was it enough just to work in American archives? And I thought it was ridiculous. Even if you were only interested in what impact the U.S. has had in other parts of the world, you couldn't do that by definition if the only thing you knew about was what the Americans thought they were doing. Right? Because, as we all know, sometimes the impact is different than what you imagined. So I was persuaded by those like Charles Maier, Christopher Thorne, Sally Marks, who were arguing that the whole field of the history of American foreign relations had stagnated because it had become isolated. Not just from the work of scholars in other countries working in multiple languages, but also because, within the American academy, people thought that diplomatic history was a very small and parochial field, and a very unfashionable field. So early on – I think it was probably partly because of Kennedy's influence – I was committed to learning languages and doing multi-archival research. And to be honest, I thought why wouldn't I want to travel the world? And I thought as a student one of the great things is to be able to go on long research trips where you can live in a society for six months or a year and learn something about it firsthand.

So, yes, I started to identify the work I was doing as international history. And the book that ended up coming out of this, called *The Diplomatic Revolution*, was very much a work of

international history. But because of another project that I did with [Paul] Kennedy, where we imagined the future and we focused on the things that you think you can imagine when you are looking 20 or 30 years out, even that book dealt with issues to do with demography and changes in communications technology. And also changes in the way that politics is organized: when we think about, in the case of the Algerian war, the role of the Red Cross, the United Nations, and so on in the international media. I was very interested in the role of the so called nonstate actors.

That's what gave me the idea for my second book, a history of the population control movement. So it is global in the sense that the people who try to control world population, they were carrying out a project – actually, multiple projects that were meant to have a global impact, whether it was about controlling numbers, or it was about controlling the quality of populations, so called.

But I do not typically use the term global history because, when people hear it, they think you are trying to write the history of everything. And in these fields, one of the big challenges is to write history that allows you to grapple with things that move across borders and, in some cases, circle the world, but without having to do the history of everything. Because the history of everything would be the history of nothing. So, that [second] book, *Fatal Misconception*, was one where I thought I should focus on some issue or set of issues where, if I want to understand the role of international and nongovernmental organizations, I will pick something where I know that they have had a big role, an important one.

AG: We would be interested in knowing how do you make global history today. What is transnational history today and how do you go and where do you go, beyond the archives? I am thinking of *"The Freedom of Information Archive[1] "*, where you ask people to contribute with materials. So, what kind of materials are you trying to bring together in doing this?

MC: Well, the way I have described it is I wanted to write a book about international history where it was going to be about one time and one place [the war in Algeria in the Cold War context], but it was going to allow me to address larger things happening all over the world. And then I ended up writing this book about population control, which is, or at least tries to be, a kind of global history. People sometimes ask me, how do you do that? So even if I did not necessarily plan this out in advance, the way I understand how each of these projects is related, including the one I am working on now, *The Declassification Engine*, is that they are all about sovereignty. They are all about unaccountable power.

Government secrecy is, by definition, a kind of sovereign power because, as citizens, we do not know what we do not know – we are not allowed to know. This is the power that

governments have, right, to keep us – all of us – in the dark as to what they are doing. And, even in the United States – where we have a long tradition of thinking we know how to keep government accountable –, in fact, secrecy has always presented a conundrum. Even if you go back to the founding of the United States and the Constitutional Convention, it was not clear how it is you could keep an Executive accountable if the Executive was going to be able to withhold information from the public and even from Congress.

So, each of these projects is about sovereignty. It is about unaccountable power. The next one I want to write is going to be a history of forecasting and future scenarios, and so, how institutions use this claim to be able to see the future as a way of asserting their power in the here and now. So, when you ask that question: "how do you do this kind of work?" Well, to me, it helps if you have this idea in mind as to what it is you are doing. If you are not going to just kind of do one thing and then the next. And I think that, over time – I am hoping, anyway – I will have more of an impact. If I can, in fact, write the history of a concept, in this case, a concept of sovereignty and the ways that it works in different times and different places. But, in each case, each of these projects has required a different way of doing it. And so the first two books, I felt I went as far as I could using what is already a very traditional method of doing history, which is multi-archival and multilingual.

But I think there comes a time when there is no one person who can learn all of the languages and go to all the archives, and so on. Further, I am also struck by the way – especially because of the work I have been doing on secrecy – the conditions for doing archival research are really changing. And it is coming to the United States, I think, earlier than it has in other countries because the U.S. started using electronic records before other countries did. But, I think, eventually, historians all over the world are going to find that there won't be archives for contemporary history, at least not in the way we understand them. There will not be finding aids. There will not be archivists who will have curated these collections. Instead, we are going to have databases. And maybe we will not even have databases. Maybe, if we are thinking about using the archive of social media, we are going to be hard pressed to find any kind of usable collection to do our research. And we will not know what to read, because the volume is going to be so overwhelming, it is going to be difficult to know how do we apply the methods that we have been trained to apply in this world in which the internet is the archive. So, with these projects on secrecy, I am trying to develop a new set of methods to try to do what I think of as multi-archival research, the same way I have been doing all along, but doing it in a way that scales to the size of the archive, when the archive is millions, upon millions of electronic records. Once again, I am picking – in each of these projects, I try to pick something that I thought was doable but that was also going to be revelatory, it was going to be able to reveal other things, bigger things. So, in this case, I picked secrecy. Because, I think that, if you are an historian, you cannot not be interested in secrecy. This goes all the way back to Thucydides, the first historian. He is the first historian because he was the first

one to talk about his sources. How did he know what he knew? And he talked about how he talked to different people and tried to compare accounts, and so on. So, historians, I think – we are not historians if we are not critical of our sources, if we are not understanding where this information comes from. And if we do not, at some point, ask: what is missing? What is it that we do not know, because someone or some government, maybe every government, does not want us to know it? So, I thought that this would be a great question to use where you could use computational methods. It is really hard to do traditional archival research on secrecy because, by definition, what you really want to know is not there. And so, using algorithms, you can figure out what is missing. You can figure out what are the patterns or the anomalies in what is available and what is not available.

CC: Let me ask you about different labels that are in the market like "digital humanities", "data science", "big data", "social science computing"… Everything is related, in some way, to this change you were talking about in the work of the historian. So, how do you see these different labels? How do you describe your position?

 MC: Of the things that you mentioned in history, in my discipline, digital humanities is the one that people are most likely to recognize and understand, or at least they think they do. Because digital humanities is already a pretty familiar concept. It has been around, I would say, certainly, for more than 20 years. Going back to the early days of the Internet, people were already talking about it. And some of the earliest projects were digital archiving projects. Typically, what that meant was historians would do traditional archival research as they always had done. But beginning when you could start to scan documents, and then when you could start to take pictures of them, they would start to share their research. They would put these digitized images online. And, early on, digital humanities in many cases meant, basically, sharing your research and making it more accessible. There was a lot of hope and expectation as to how this was, in itself, going to be transformative. In addition, certainly, it has been transformative in teaching. So, for example, back when I was in graduate school, and I was teacher assisting classes, at best, maybe a teacher would distribute a course pack where you would have maybe a few dozen documents that they had copied for you and made part of your readings. And that was great, right? Because they were already giving you some kind of direct access to the archival record.

 With digital humanities, you can give your students access to hundreds of thousands of documents, and you could let them explore. Instead of kind of choosing the curated items that you think they should be interested in, you can give your students the ability to explore themselves. That is fantastic. But the field of digital humanities is still largely seen as a teaching field. And much of what it consists of, at least in terms of what people know about, is the same kind of thing. It is *more*, right? It is more digitized images, more websites, but featuring research that is still rather traditional. Because, for the most part, people doing

digital humanities are still doing archival research using traditional methods. It is just that they are finding new ways to share that research. Where there has been more of a change, it has been in the use of GIS – Geographic Information Systems. For certain kinds of projects, it is really quite revealing if you can use geo data to show, spatially, how things change and evolve, and so on. And there are many examples of that. In fact, many libraries now have librarians whose job is, mainly, to provide GIS data and to help people use it.

Nevertheless, the problem with the digital humanities field, generally, is that, with few exceptions, the people doing it, they try to do it all. Where they will both do the research and then try to find ways to share that research with others, which now oftentimes means building a website, sometimes building interactive tools that allow people to explore that research in new ways. And, of course, they are still expected to publish, and publish not just online but to publish in peer-reviewed publications and scholarly monographs. Now, when you think about it, that is a lot of different skills. It is all the same skills that scholars would traditionally be expected to have, and then, on top of that, you are expecting them, as well, to know how to build websites, maybe even build interactive tools. Then, if you wanted them to actually discover something through nontraditional means, you might even be expecting them to be data scientists, to learn how to use algorithms to discover things, or at least explore data in ways they could not do in the traditional manner.

That is one reason – I believe, the main reason, even – why it is that, as a field, digital humanities has not had a bigger impact on the rest of the discipline. Because it does not threaten anyone. I am half joking, but I feel like, if digital humanities was really having an impact, there would be more people attacking it. And that is not happening because digital humanists, for the most part, are not really challenging the way that other historians do their work. They are not necessarily making entirely new discoveries using entirely new methods. So, I think that one analogue, if you imagine how things could be different, would be how the field of genomics developed. Genomics is basically the union of data science and biology. And I was on a tenure review committee, some years ago, and we would look at a new case, like, every week or two. Over time, what I was struck by was how when we had a case of somebody working in genomics, they would talk about how this person is a really good data scientist, but they are just a competent biologist. Or they would say: "well, this person is a brilliant biologist but they are just kind of applying data science methods in a new domain". And they said, the people who knew better than me, that you do not typically find – in fact, it is rare, almost impossible – somebody who is both a brilliant biologist and a brilliant data scientist, and do innovative work in both of these different and very distinct kinds of disciplines. I then thought to myself: "that is after it is been billions of dollars invested and decades of development in the field of genomics". So, people who are interested in doing data science and who want to use data science to do history – I think it may be a long time, it may be forever, until we have somebody who is both a brilliant historian and a brilliant data scientist.

AG: Should we aspire to that?

MC: Well, I think if you think of history as a kind of data science, why not? I mean, we are working with data. And we are going to be working with more and more of it, whether we like it or not. Then, we should be using the best methods available to analyze that data, right? Now that, again, does not mean that we all have to become data scientists. Because, in fact, I think that we are more likely to make more rapid progress if historians are learning to work with data scientists and to apply data science methods to historical questions. So, that is the kind of work that I have been doing for the last five or so years. I have been working with statisticians and computer scientists and, in some cases, applying new computational methods to relatively old questions that come out of history or come out of other social sciences, especially political science. However, it is also making it possible to ask different questions than we normally could. Questions that might not have even occurred to us, otherwise.

CC: How was it to convince the students, colleagues, and funders for research of the relevance of this new kind of work?

MC: Students are very excited about it. I have not had any trouble getting students interested in this, if we are talking about undergraduates. There also has been interest among some PhD students, but many PhD students in history have no interest in it whatsoever. And it is striking to me because you would think that they would have the most to gain, over the course of a career, of mastering a new set of methods that are still relatively rare. But I think that PhD students, at least in the U.S. – and I think this is true in other countries, as well –, are often quite conservative, in fact. They may not think of themselves this way. They may take risks in terms of what they choose to study for their dissertations. And they are certainly taking a big risk by going into a PhD program, which is going to be at least five years, and probably longer, with no clear prospects of getting jobs as professors at the end. But I think it is for that reason that they can be quite conservative in terms of using new methods. Unfortunately, I think, a lot of PhD students who might benefit from using data science methods are not necessarily eager to do that because they feel, perhaps, that they are already taking enough chances in joining a PhD program and putting in the years to complete it.

Among other professors, I am finding that there is definitely interest in the field. In my department, there is about a half dozen of us now who, in some way, are using computational methods to do our work. And that did not happen according to any plan. I think it is just because more and more of us realize there are opportunities here, and there are also risks if we do not start rethinking our methods, and also if we do not start thinking about how we can train our students to do this kind of work.

AG: I was interested in an example, if you could give us one. You talk about identifying bursts or spikes. So, how do you construct an algorithm? What are you looking for and what are the questions that come from history? And what did your partners bring into this algorithm?

MC: In the past, I have sometimes organized workshops with other historians to see how they would look at the potential for using these kinds of methods. And one of the things I found was that they, like me, are particularly interested in some method or set of methods that allow them to open up the black box when they have a very large collection of documents, to know, in effect, what is in there. That could mean many different things, but one thing it can mean would be, for example, events. My field, diplomatic history, is sometimes criticized because we are often very interested in events. And if you go back all the way to Fernand Braudel, the Annales School would hold that what you should really care about are the kind of structural changes that are sometimes decades or centuries in development, and events are just like the froth that ride the backs of these waves of change. Be that as it may, I do think diplomatic historians are always going to be interested in events because some of these events are really important, like war and peace. It is important to know when wars start and why they happen, and also understand why they end and whether they might end sooner. So, one of the things that I think you would want to know is, if we are going to begin to face millions of electronic records and no finding aids, how do you identify events? How do you find things happening that we would want to look at more closely?

As I work in this field, I become aware that there are huge fields of research on, for example, traffic analysis. Traffic analysis is a technique that goes back more than 100 years. It starts with espionage. Actually, in many cases you would not be able to decipher the communications of another country, but you could at least measure the amount of communications, whether you are intercepting telegraph lines or what have you. Therefore, traffic analysis is all about seeing whether there are spikes where you might think that something is happening, just because you see an acceleration in the rate of communications.

So, working with statisticians, first of all, you can say: look, I have a lot of data. I have, in fact, really important data. Because you have, in this case, millions of State Department cables over many years. And you could say: is there some way that we could use traffic analysis to identify more important events? Not just the ones that we already know about, but potentially the ones that we did not know about that might be worth exploring more closely. But the key, I find, with finding a partner to work on an experiment like this, is you have to find something that is going to allow the data scientist to innovate. That is, if all they are doing is applying a known technique, then it is not very exciting for them because they are not really innovating. They are interested if you have a lot of interesting data. Sometimes these conversations start with "I have a lot of interesting data". In this case, we started working with statistics professor

Shawn Simpson. And to start with, we were just applying a known technique, which was actually derived from analysis of e-mail. E-mail and diplomatic communications are not so different. There is a "from" and a "to", and there is a date of transmission, and so on. So you can basically take the same – it is called the Kleinberg Model – modeling traffic among people exchanging e-mail. You can apply that to diplomatic traffic between different diplomatic posts. And sure enough, you can find events. Some of those events are things that are well known but it allows you to identify them more precisely, like when do we see this kind of beginning and when does it end, and what formation does it take, and how do we begin to develop a kind of taxonomy that allows you to distinguish between different kinds of events.

Soon enough, however, we found that we needed more and better models. So I worked with another professor who is at MIT, Rahul Mazumder, and he, together with two of his students, developed a new way and a better way of doing traffic analysis with a large volume of communications like this. We are hoping to get this published, shortly, in a journal of statistics. I find the way in which you can begin to do this kind of work is to start with, you could say, "I have a lot of really interesting data and, even if we are applying a known technique, we might have interesting conclusions". Then, sometimes, it progresses from there to finding that there are new ways you need to analyze that data to find even more interesting things. And then, sometimes, you can innovate in a different discipline.

CC: Dealing with the State Department cables, and the governmental records etc, you are entering in a sensitive and sometimes dangerous field, especially in times of WikiLeaks and episodes like Hillary Clinton's e-mails. Do you not fear that FBI or CIA will come to your door to ask you what are you doing?

MC: Well, I have had to learn a lot. Yeah, I have learned a lot about how it is that there actually are laws on the books that, if they were applied, would potentially be quite restrictive in terms of what kind of research can be done. For example, you take the Espionage Act. It is a World War I era law, but it is still the law that is used, in many cases, to prosecute people for leaking information, for instance. If the espionage law was applied in every possible case, there would be many journalists and professors who would be locked up right now. But, in fact, over the years the law has only been applied in a more restricted way and courts have established that if the work you are doing is meant to be in the public interest and if you are a journalist or you are a professor, that you are given some latitude. Because, long ago, judges recognized that if journalists and professors are too fearful to do research, then, really, we are all in trouble. So, it would be different if, for example, I was doing work that was seen as favoring a foreign power. But, in fact, all of the work that I have done has been very much about trying to better understand the nature of American foreign policy in a way that I hope will contribute to better policy. Or try to better understand what information the government

is not sharing with us, so that we can hold government accountable, to keep government transparent. Therefore, it is very much consistent, I think, and most people would agree, with the national interest, that people would be able to do this kind of research. And, to be honest, we are a bunch of professors. And if we are doing this kind of research, you can be sure that others are doing this work with more nefarious intentions. I sometimes think that what we are doing is – we are a kind of a red team, to see what can be done. If we ever did find anything that we thought was potentially going to create problems, then it would be our practice to share that with people who could tell us whether there might be problems or issues with it. In the same way that journalists do it sometimes when they discover things and they may think that there might be a concern with sharing those discoveries publicly.

One other thing I will say about it, though, is that we only use declassified documents. So, we have only ever done analysis on documents that have been duly released by the U.S. government. We are not using WikiLeaks data or other leaked documents. The reason is not so much the legal aspects, but that we cannot trust the provenance of leaked documents. It is hard to do research people will believe if you do not know where your data came from, and also if your data is very incomplete the way the WikiLeaks data is, for instance. They only have a relatively small, perhaps five percent, non-random sample of U.S. State Department cables. That is not a very useful data set to do research. Using declassified documents, therefore, allows us to look at a much bigger and much more representative selection of the historical record. And so we are able to ask questions and ask them more rigorously than we could if we were using leaked documents.

CC: How do you think more traditional historical research and computational methods can relate to each other? Are they converging, complementing, or relatively independent methodologies?

MC: When I work with data scientists on a project, I find that, first of all, data scientists do not necessarily know what kind of question to ask of this kind of data. And so we often have to talk with each other as to what would be interesting to find out. And, are these things, these questions, tractable with the kind of data and the kind of methods that we have available to us? So, my role in many cases is to identify what kinds of questions we want to ask. Then, the other role that I play – I think, the biggest role – is to interpret the results that we get. For example, the research I was talking about doing traffic analysis and identifying historical events, it generated more than 500 different candidate events. That consists, basically, of finding when you have bursts of communications. To understand what those events were about you have to read a lot of those communications. And so I end up spending a lot of time doing this research reading documents. Also, reading about those collections and how they came to be. Many times, when we find things that seem anomalous or strange, I have to do research on those anomalies.

When I started out, for instance, I was not an expert on the history of the 1970s. But because we have tons and tons of data from this period, when we began – the U.S. government begins to produce electronic records – I had to learn a lot about the 1970s. So, really, I had to combine the skills of doing algorithmic research, and the historian skills of close reading and contextualization, not only looking at the individual documents, but looking at the archive and understanding how it came to be. Further, also reading around once you find something interesting and find out what other people have written about it and what we knew about this before we made this new discovery. All those things require qualitative research of a very traditional kind. I find these things absolutely go together. They do not always go together, though. And, in fact, there is often – among data scientists, it is notorious that if – as an engineer, – if you think of yourself that way – if all you are interested in is improving the performance of your algorithm, in many cases, people do not read much further and they do not necessarily look at what results they are really getting in looking more closely at the data. That does not lead to very good science. And so, what we are doing is trying to do both. We are both doing data science, but we are also doing qualitative research on those results. In that way, I think, it leads both to better data science and to better history.

CC: Thinking about your students today, do you think that some specialized digital method would be mandatory for them? And what impact it should have in the teaching of history or social sciences, in general?

MC: Well, I think you would be hard pressed to find anybody doing any historical research who does not use databases, who never uses a search function, anybody who has ever used a database, or ever used Google Scholar, or ever tried to do keyword searching in a collection of documents. I think, if you are going to do that, you need to understand something about how databases work and how they are indexed. If you are going to use a search function that ranks your results, you need to understand how that system operates and why it is ranking the results in the way it is. If you are doing something that involves searching scanned documents, ones that were not born digital but things maybe like newspapers or documents that have been digitized, then you need to understand how text was extracted from those images. And you need to understand how, typically, a lot of text is lost. So I think anyone who is going to be doing any of these things really needs to understand a bit about how these systems are built and how they are limited, and how the results can be quite misleading.

For example, I have often heard students and professors say: "well, I do not find any references to human rights before 1968, and I have searched all the article databases from ProQuest Historical Newspapers". Well, you cannot prove that nobody was using that term in any of those newspapers just because your search did not yield any results. Because digitization is not perfect, and OCR results are often garbled. And, in fact, sometimes those

different newspapers might have been digitized at different times and they might have applied OCR at different times, so you may not even be able to compare one to another in the way that people like to do when they talk about numbers of hits that you have in one place or in one period versus another. These questions might not even occur to you, unless you've had some experience yourself in doing digitization and extracting text.

So I do think we are getting to the point where: yes, graduate students, at least, are going to need to learn something about how the systems that they use are built and how it is they work and sometimes do not work. That said, I think for a long time to come, certainly if you are working in the period before the era of electronic records, you are not necessarily going to have to learn how to do computational history. You can use these tools and understand how not to misuse them, but without having to learn how to build them for yourself. Or, even as I do, working on projects that are about tool building. But I think that students shouldn't be afraid of at least exploring what can be done. Because I have seen PhD students who have had very little experience in coding, who, in a period of a few months, have been able to do very sophisticated kinds of research like training machine learning algorithms to automatically classify documents. These are students, again, with very little background in coding. But what has happened is that, because more and more people are interested in using things like machine learning, a lot of these techniques have been made easier. Where people have built packages, as they call it, to let people, even novice programmers, use them with new data. And so, I would encourage especially students just starting out to consider what is possible, and to think about whether some of these methods might be useful for their own work.

CC: You have built, in the last years, a network of collaboration with other countries like U.K. and Brazil. Has this networking affected your work?

MC: Absolutely. My project started out as exclusively one in which we were collecting documents from the history of U.S. foreign relations. We are now building relationships in order to make this a more international collection. And what that typically means is finding partners in other countries to work collaboratively with us. When I first started doing this kind of work, I found that there really was not another project like the one I was trying to build at Columbia, in the U.S., where you could find data scientists working with social scientists in large digitized corpora of historical archives. Instead, I found that there was a project in Brazil, at FGV, and a project too in the U.K., in project Abaca. There are still not a lot of examples of people doing this kind of work, so it is all the more important, then, that we network amongst ourselves. And we have been successful, both in the U.S. and in Brazil, in raising funds to make those kinds of collaborations possible. But I think there is going to be more that we need to do. And a lot of times that means being able just to work side by side. So, some of the most rapid progress we have been able to make has been in those moments where we have been able to work in the same place.

CC: Could you tell us about the FOIA, the Freedom of Information Archive, to which you are a leader, as that was just launched by Columbia?

MC: This project started out where we were especially interested in understanding secrecy, and declassification. In doing that research, we had a grant from the MacArthur Foundation, and in the course of doing the work we had to aggregate many millions of documents and structure them in databases. Once we had done that, it was clear that this was a resource that others could use to do research, too. Therefore, we were able to get a grant from the National Science Foundation in order to further develop this collection and to collaborate with Columbia Libraries in order to make it accessible to anyone who wants to, for instance, use the Columbia Library search interface. So, whereas, at the outset the project was one where we were working on very fraught difficult kinds of policy and sometimes political questions – like, how do you do this kind of research on secrecy in an ethical way, and a way that is likely to have some impact on the public discussion and debate about secrecy –, more and more the project became one in which we were trying to help others do research, Not just on the history of secrecy, and different ways of doing policy on secrecy, better ways, but we were also interested in allowing people to do research on a whole range of subjects.

So, the Freedom of Information Archive is a resource that we are making available to everyone who gets online and wants to either go to the History Lab website or wants to explore that collection through the Columbia Library's website. Now we have an advisory board made up of people from multiple university libraries, and we want to put together a consortium where we can make this a collaborative project and that would serve multiple functions not just to create a resource for researchers, but also to ensure the preservation of these materials. So, in the United States there are many libraries now who see themselves as data refuges where, especially if you think about data like declassified documents, you can imagine, and it has sometimes happened in the past, that things can get reclassified. Things that might have been provided by the government previously could be withdrawn. So we want to serve as a data refuge for this kind of resource. And also, we want to continue innovating and showing new ways that you can develop techniques for exploring this kind of data and making it usable, not just for historians but for other researchers, as well.

AG: What is the role of the imagination in making history, still, with all these data?

MC: Yeah, the imagination. So, I feel like I have the best job in the world. I just love being able to get up every day. I sometimes say to students that if you are an academic, a scholar, and you are bored in your work, it is your own fault, because nobody is going to force you to do boring work. The work you do is as interesting as you want it to be. I have been very fortunate that I am at an institution that is very well resourced and we have not just a great history department but we have great computer scientists and data scientists. We have a great library. I have really been incredibly fortunate to have such amazing resources available

to me to do something really ambitious. For me, then, all the more reason then that I want to attempt really ambitious things, challenging things. I cannot remember now who said this, but somebody once said that you should try to work on the most important problems confronting your discipline, right? And I feel like there is no more important problem facing historians than the transformation of the historical record. The way that we have been trained to do our work over many decades, even centuries now, is going to become obsolete in our lifetime. At least for those of us who want to do contemporary history. So, I feel like, if not me, then who? I mean, people like me should be working on this problem and we should be trying to come up with ways of grappling with it that is going to make life easier for our students. Because if we are not training them and preparing them to deal with this new kind of born digital world, then there is nobody – I do not see anybody else doing it.

When you ask about the imagination, then, I try – as an historian, I think, even if we do not admit it, we do imagine what things are going to be like down the road. It takes us five or ten years to finish a book, so you have to imagine that people are still going to be interested in whatever you are doing some years from now. But when it comes to training our students, you have to imagine the work that they are going to be doing, not just 5 or 10 years from now but 20 or 30 years from now. When I imagine, then, that world, and my students, and the work they will be doing decades into the future, I cannot help but think that they are going to have to be using, if not the techniques that I am working on now, then techniques that may be new and improved and more advanced and perhaps even more imaginative versions of what I am doing. So, to me, that is incredibly exciting. Even if I fail utterly, I will feel like it was worthwhile just to make that effort. Because I do feel it is an urgent problem and it is becoming more urgent the more we have to worry about the expansion of secrecy and the way it is getting harder and harder, even to preserve the historical record for anyone to do research.

There is no more traditional or important question than asking: "how do we know what we know?" And source criticism is the hallmark of good history, and it always has been. We are now in a period where, I think, history is already half digital. Those historians – I think practically every one of them who is using the internet to do research is using search engines – are already doing digital history whether they know it or not. So the question is: are we going to do it well? And how do we do it well? My father, when I was growing up, was always telling me: "do not be narrow." I did not really know what he meant at the time. But over the years I have come to understand that – and I think it is completely true – one nice thing about history is you can think broadly. History is as capacious as you want it to be, and it can be many different things. One thing for sure it has to be is it has to be grounded in evidence, and we have to know what that evidence is and where it comes from. So I absolutely agree that we need to be imaginative. In the U.S., we have seen a now ten-year decline in the number of students who are majoring in history. So, what we are doing is not working. We have to find new ways to communicate the value of what we do to a new generation. And so I am hoping I can be a part of that.

'Notas de fim'

1On The Freedom of Information Archive, see: https://www.indiegogo.com/projects/the-declassification-engine-saving-history-from-official-secrecy#/

Referências:

CONNELLY, Matthew. A Diplomatic Revolution: Algeria's Fight for Independence and the Origins of the Post-Cold War Era. Oxford and New York: Oxford University Press, 2002.

CONNELLY, Matthew. Fatal Misconception: The Struggle to Control World Population. Cambridge e Londres: The Belknap Press of Harvard University, 2008.

CONNELLY, Matthew & KENNDY, Paul. "Must it be the Rest against the West?". The Atlantic, 1994.

KENNEDY, Paul. Rise and Fall of the Great Powers. New York: Vintage Books, , [1987] 1989.

WRIGHT MILLS, C. The Sociological Imagination. Oxford and New York: Oxford University Press, [1959] 2000.