ARCHIVES IN A DIGITAL SOCIETY

D^R LUCIANA DURANTI is Chair of Archival Studies at the School of Library, Archival and Information Studies of the University of British Columbia (UBC), and a Professor of archival theory, diplomatics, and the management of digital records in both its master's and doctoral archival programs. She is also Faculty Associate Member of the UBC College for Interdisciplinary Studies, Media and Graphics Interdisciplinary Centre, and Affiliate Full Professor at the University of Washington iSchool. Duranti is Director of the Centre for the International Study of Contemporary Records and Archives (CISCRA, www.ciscra.org) and of InterPARES, the largest and longest living publicly funded research project on the long-term preservation of authentic electronic records (1998-2018), the "Digital Records Forensics" Project, and the "Records in the Clouds" Project. She is co-Director of "The Law of Evidence in the Digital Environment" Project.

CORINNE ROGERS is adjunct professor (diplomatics and digital records forensics) and doctoral candidate (researching concepts of authenticity of digital records, documents, and data and authenticity metadata) at the University of British Columbia. She is Project Coordinator of InterPARES Trust – international multidisciplinary research into issues of trust in digital objects in online environments, and a researcher with the Law of Evidence in the Digital Environment Project (Faculty of Law, UBC).

INTRODUCTION

The focus of this paper is on the responsibilities of the Canadian archival system for the born-digital residue of societal activities and the strategies for its preservation. It will first outline the characteristics of the digital by-products of the activities of organizations and individuals and it will then proceed to describe the consequent responsibilities of the Canadian archival system and to envision the way in which it could fulfil the expectations of Canadian society.

THE CHARACTERISTICS OF TODAY'S BORN DIGITAL INFORMATION (DATA, DOCUMENTS, RECORDS)

Understanding the characteristics of today's born digital information is not possible without first analyzing the behaviour and expectations of its creators and users.

In the past few years a clear generational divide has emerged between those born before and after 1981, the latter group being defined as the millennial generation, or Generation Y. The tendency of the former group has been to bring to the digital environment the behaviour it had in the analogue one, based on the same values and understandings, particularly in relation to the fundamental concepts of access, privacy, intellectual rights, ownership, authorship, custody, responsibility, accountability,

recorded memory, and preservation. As a consequence, rigorous analysis and research of its digital output has resulted in widely accepted recordkeeping and records preservation practices, international standards that are scalable and implementable, and – as it regards Canada – in a generation of archivists confident in their knowledge and in their ability to acquire and maintain control of the existing output of our digital society, at least for as long as adequate technological and financial resources will be available. In other words, the digital documentary by-products of the activities of the pre-millennial generation, in both organizational and personal contexts, can be preserved and made accessible without the need for fundamental changes to the Canadian archival system.

This is not the case when we consider the digital behaviour and expectations of Generation Y, which is now reaching adulthood and has begun its working life in an environment characterised by media convergence and constant connectivity, often described as mobile and liquid. This is an environment in which old and new technologies merge in the same device so that cameras, phones, computers and the Internet are seamlessly integrated. Social networking and photo sharing sites encourage users to communicate online and give their contacts the permission to add comments, notes, tags, images, etc. to their materials, so that the boundaries between the producers of content and the users become blurred. The by-products of such interaction are no longer finite entities, but processes that are always changing, mediated by the technological platform that hosts them through metadata, protocols and codes. This phenomenon is well known to social science researchers, who refer to the participants in it, be they organizations or individuals, using the neologism *produsers*, a term that intentionally merges the roles of producer and user.

Thus, one can see that the relatively new practice of BYOD, or Bring Your Own Device, in organizations and businesses constitutes an issue not simply because corporate digital information is no longer captured or moved into a central recordkeeping system as records, but also because it is rarely complete, lacks fixity, is difficult to attribute to one author and/or owner – especially because of the collaborative way in which people act, and is not clearly linked to one business process, action or transaction. It is very difficult in this case to assign responsibility or demand accountability. The situation presents further complexity when employees, together with their devices, use their own cloud, or several different clouds in which their outputs are stored in a widespread way, mostly related to applications rather than to business functions (e.g. DropBox or Google Docs). "A recent survey indicates that more than half of Generation Y workers are [sic] prepared to contravene corporate bring your own device and cloud computing policies if it [sic] cramps their personal and professional computing and social networking activity."[1] The work will be conducted in the most efficient way, regardless of whether it leaves a recorded trace – data or records – usable as reference for further action or as evidence of actions and transactions.

Over time the digital information *produsedd* by these employees grows in volume, and, being spread across different devices and multiple online social networking platforms, becomes unmanageable, and its producers simply choose to let it accumulate, given the availability of large amounts of free cloud storage. Thus, the problem of absence of records usable as precedent or as evidence is compounded by uncontrolled abundance, an issue that is beginning to bother also the *produsers* themselves. But technology has already provided a solution and now offers automatic deletion after a certain amount of time – a time span certainly not based on a retention and disposition schedule approved by the employer, but on the intention of the communication generating the information, as identified by the person initiating it. Why would an employee not consider the possibility that the information in question may be needed later on or even in the long-term? Simply because Generation Y considers communication more important than memory and the material it generates is supposed to have an immediate impact and be consumed instantaneously. Thus, there is no expectation of preservation, only a desire to generate output throughout each and every event, continuously, as if nothing could possibly happen if it is not communicated but its occurrence were not deleted by the disappearance of its digital traces.

If for a moment we put on our archival hat, it is pretty clear that this behaviour in organizations undermines not only their long-term productivity and growth, as well as present and future compliance, responsibility, accountability, ability to protect themselves and, last but not least, the building of an organizational culture, but also any chance for archives of identifying records, series, or fonds, carrying out systematic appraisal, and acquiring the sources of societal memory. And the

situation is much worse if, again as archivists, we think of the fonds of voluntary associations, NGOs, and individuals. While in all these cases many of the activities are carried out in multiple devices and online environments, individuals add a level of complexity to their digital output by selecting communication languages that are outside established parameters, for example by increasingly resorting to images and abandoning words (e.g. sending the picture of a building and a Google Map instead of a message with the invitation to a place and address). Furthermore, for archivists, tracking down the components of a personal archives over time would involve having access to the creator's social networking sites, mobile phone accounts, email accounts, *Twitter* and photo-sharing accounts, etc. Even if the material still existed and could be located and accessed, it would still be difficult to identify the "personal fonds." [2]

The characteristics of the digital information *produsedby* Generation Y are thus quite clear: when it is traceable, and it has identifiable boundaries, it is a co-authored entity (by persons and platforms, software codes and protocols of the host applications), in continuous becoming, generated to serve multiple purposes in multiple contexts, composed of more than one medium, performed, rather than issued, and therefore ephemeral in nature, generally accessible by anyone who shares the same platform and/or belongs in the same circles, and unprotected by intellectual rights.

How does this blurring of boundaries between public and private, organizational and personal, ephemeral and permanent, complete and in a state of becoming, platform generated and user generated, owned and open, recorded and performed, etc. meet the expectations of those who regard themselves as records and archives users? Has their behaviour changed as well?

The users of records and archives still for the most part belong to the pre-1981 generation, and this means that they expect the digital environment to facilitate their established behaviours; however, they are also discovering new opportunities offered by complex technologies. Researchers who have always worked collaboratively now expect to do so in a virtual environment; users of material in archival holdings expect that these materials be made available through the Internet, possibly instantaneously and in the required and interoperable formats; citizens interested in the actions of government and corporations expect open access to the records of their actions through time; scientists from the world of data and statistics are interested in open data initiatives and expect traceability of such data; and an increasingly large number of users from the private sector as well as academia is employing diverse sources of big data to create data mashups for a variety of purposes. While the Canadian archival community has been aware for a long time of these users' behaviors and expectations, the most noticeable change has been a growing interest in data – as opposed to records, and in large aggregations of research data, observational data, and government-collected data for supporting decision making. The key consideration to make here though is that this interest in data comes with a strong interest in their accuracy, reliability and traceability to an authentic source. This requirement of trustworthiness, once implied in the preference given to primary sources, is becoming an explicit request of a documented chain of unbroken and responsible preservation: in other words, the trust traditionally given to the preserver now expects transparent grounds.

RESPONSIBILITIES OF THE CANADIAN ARCHIVAL SYSTEM – RECOGNIZING THE CHALLENGES

How can archival institutions, organizations, units and programs – in short, the Canadian archival system – address the opportunities and challenges posed by the changing ways in which Canadians are communicating and working? The focus of the archival community on 'memory' and 'evidence' encourages the popular (mis) perception of archives as being concerned only with records of the past (whatever the medium). This tends to marginalize archives and archivists, perpetuating the myth of the passive preserver (even though we know how wrong that myth is). We have known for a long time that our born-digital material will not survive in useable, trustworthy form if its preservation is not considered from the moment of its creation. While this is difficult enough for born digital records that mirror traditional counterparts, it creates many new challenges in records created collaboratively and online. Identifying issues of authorship, ownership, access, and intellectual rights would merely scratch the surface of the challenge in front of us. What are the responsibilities and strategies

of the Canadian archival system in supporting a controlled creation of data, documents, and records that can demonstrate transparent processes and a clear chain of custody or provenance, and whose accuracy, reliability, and authenticity can be assured over time and across technology platforms? Now more than ever archivists must position themselves at the side of the creators if they wish to be able to fulfill their organizational mandate: they need to identify records and data worthy of capture and long term preservation at creation, determine their capability to be preserved, and monitor their use and transformations through time. This means understanding the changing boundaries between public and private, organizational and personal, author and owner, creator and preserver, producer and user, and the attributes of the *produser*.

A broad regulatory framework needs to be developed and implemented to ensure and protect the creation and management of these new organizational records, addressing the new reality of BYOD policies, and the integration of personal and organizational, private and public information that inevitably results from its adoption. Within it, specific policies should require that retention and disposition schedules be in place for records created in personal devices as well as in organizational recordkeeping systems, and that employees do not apply deletion dates to records designated for long-term preservation. They should also consider that the connectivity of the network database mediates the user experience and could allow tech-savvy individuals access to information that must be protected under privacy legislation, and should ensure that this does not happen.

In the realm of personal archives, the Canadian archival system needs to understand and figure out how to identify a person's archives over the long term (e.g. how to track down the correspondence of one person with a set of friends, over time, through photosharing accounts, social networking sites, mobile phone accounts, email accounts, *Twitter*, etc.), to locate what still exists, to gain access to it (dealing with password protection), to retrieve it, and to make sense of an overabundance of ephemera, and then provide guidance to individuals and small communities on how to appreciate, maintain, and protect a documentary memory.

FULFILLMENT: EMBRACING THE FUTURE

The digital society is in a continual state of becoming. More than ever before, the Canadian archival system must take responsibility for the archival functions that are at its core (appraisal and acquisition, arrangement and description, retention and preservation, management and administration, reference and access), but it must do so across the entire cycle of creation, management, use, and re-use, across public and private sectors, in an integrated way.

The archival core functions can only be accomplished successfully when based on a solid foundation of interdisciplinary collaborative work, research, and education. Archivists are not computer scientists, nor should they aspire to be. We will not influence the direction of technological developments, but we cannot be by-standers. We must put ourselves in the position to predict such direction and those developments, and to be responsive to them if we wish to identify, capture and preserve tomorrow's heritage. This will not happen by sitting back and waiting for academia to study the problem and deliver solutions. All archivists can and must be among those who design solutions, working together in institutions, repositories, academic departments with colleagues in a variety of other fields and disciplines if we are to successfully answer the questions that were posed by the CAST Interim Report.[3]

A model for this type of interdisciplinary activity exists at the University of British Columbia. InterPARES Trust is an international research partnership involving archivists, records managers, librarians, as well as lawyers, computer scientists, digital forensics specialists, information security experts, journalists, ethicists, policy experts, and more, from industry, government, and academia. Experts of each discipline listen to and learn from those of the others, lending the strength of many disciplines to one purpose – how and what is required to trust data, documents, and records in online environments.

All the knowledge acquired by this and similar collaborations certainly allows all participating organizations to hit the ground running as new issues surface and are addressed, but would be wasted if not transmitted to the new generations of professionals.

This needs to be done by involving students in active collaborative research within the projects' test-beds, be they archives or records creators' environments. Interdisciplinary research needs to be embedded in archival functions at every layer of the archival system – we reach out by going out, learning to work together, conducting rigorous investigations into new developments, passing the knowledge on by involving students in research. Continuing education courses are important, but they convey acquired knowledge. We need stronger programs of archival education that embed research capability and interdisciplinarity into the field, and that support the delivery of archival knowledge and the acquisition of new knowledge at the conferences of a variety of industries. We do not learn what is needed to address the challenges outlined earlier by inviting a Google (or Microsoft, or Amazon) representative to speak to us, but by submitting proposals for papers at the conferences which Google-like industries sponsor, like the Global Cloud Security Alliance annual conference, the Digital Forensics annual conference, etc.

Canada's National Strategy for Critical Infrastructure[4] (2009) identifies 10 critical infrastructure sectors: energy & utilities; finance; food; transportation; government; information and communications technologies; health; water; safety; manufacturing. All of these depend on authentic, reliable, and accurate records that are preserved for as long as needed. "To be effective, the National Strategy must be implemented in partnership among all levels of government and critical infrastructure sectors… Governments bring value to the partnership through activities such as: providing owners and operators with timely, accurate, and useful information on risks and threats, …" (authors' bold).

A strong Canadian archival system is a key part of that infrastructure and should be proactive in determining its place in and contribution to the Canadian digital infrastructure ecosystem. The Leadership Council for Digital Infrastructure is certainly a forum in the context of which archivists should advance proposals for an integrated archival infrastructure.[5] Can we develop and support an archival community cloud? Records creating institutions and organizations are moving to the cloud in increasing numbers because this can simplify interactions with users by reducing information processing time, lowering cost of services and enhancing data security: it would be extremely difficult to acquire control of their materials selected for permanent preservation by moving them to an in house Trusted Digital Repository (TDR), especially considering their massive amounts. In addition, it would be financially impossible for every institution to have an in-house TDR, except one – limited in its functionalities and dimensions – for handling the preservation of data and records that must be strictly protected because of privacy laws and legal rights, and cannot be shared even across repositories. Certainly, before choosing the cloud option, the archival system would have to address standards for service levels, security, and interoperability/reversibility.[6]

A Canadian archival cloud would need to provide much more than storage, enabling all the archival functions in a framework of distributed responsibility. In fact, by centralizing the Canadian born digital archival holdings in cloud storage, we can eliminate the need for concentration of the same holdings under the control of a few major institutions. Born digital materials of agencies and organizations that have a designated preserver (e.g. public bodies, universities) would be generated and maintained in the archival cloud and remain in the control of their creators for as long as they need to be used by them, but would be monitored at the same time by the designated preserver. Afterwards, the legal and intellectual control of the materials identified for long term preservation would be transferred to the designated preserver.

Records creators – be they organizations or individuals – who do not have a designated preserver, can be directed to standard clauses to include in the contracts with their providers, which would contemplate the possibility of transfer of selected materials to the archival community cloud under the control of the archives wishing to acquire them, and facilitate it. In fact, collaborative work with the industry to design contract models that support archival acquisition has already begun in the context of a variety of research projects, and the Global Cloud Security Alliance is paying attention.

Of course much thought has to go into how the archival functions would be exercised, especially appraisal, given the gigantic proportion of ephemeral output. However, it is already clear that, as records creators place more emphasis on communication over memory, performing over recording, images over words, and on reuse, co-authoring, and crowdsourcing, the documentation

of context becomes vital to understanding and preserving meaning, providing a collective authentication of the materials within each fonds, and ensuring controlled and accurate access: archival description will return to be the core of the archival endeavor and will become increasing complex.

Indeed, the Canadian archival system is facing challenges that seem insurmountable, but the resilience, creativity, and dedication of Canada's archival community can rise to the challenge. This is a very exciting time for a community as knowledgeable and committed as that of Canadian archivists.

NOTES

1 www.itworldcanada.com/article/gen-y-workers-ok-with-flouting-cloud-byod-policies/86041?sub=67871&utm_source=67871&utm_medium=comminfra&utm_campaign=enews. For reports on additional surveys go to www.ashridge.org.uk/website/content.nsf/wFARCRED/Culture+shock+Gen+Y+and+their+managers+around+the+world and www.pwc.com/gx/en/managing-tomorrows-people/future-of-work/key-findings.jhtml.

2 For several of these insights on individual behavior we are grateful to UBC doctoral candidate Jessica Bushey, whose research has unearthed them in social sciences literature.

3 The CAST Interim Report poses the following questions: With the use of new communication technologies, what types of records are being created and how are decisions being documented? How is information being safeguarded for usefulness in the immediate to mid-term across technologies considering the major changes that are occurring? How are memory institutions addressing issues posed by new technologies regarding their traditional roles in assigning value, respecting rights, and assuring authenticity and reliability? How can memory institutions remain relevant as a trusted source of continuing information by taking advantage of the collaborative opportunities presented by new social media? See ACA Canadian Archival System Taskforce, *Interim Report*, June 2013: 11.

4 Critical infrastructure refers to processes, systems, facilities, technologies, networks, assets and services essential to the health, safety, security or economic well-being of Canadians and the effective functioning of government. (NSCI 2009).

5 See digitalleadership.ca.

6 Like Australia (National Archives), New Zealand (National Library) and the US (Federal Agencies), Europe is clearly heading in the direction of the cloud: www.futuregov.asia/articles/2013/nov/25/eu-cyber-security-agency-issues-cloud-deployment-g/. With the appropriate firewalls ensuring separate control by each organization responsible for each body of records at any given time, the Canadian archival system can have one cloud embracing it all rather than a separate government cloud for each level of government or even for each agency, institution or organization. This would be a full implementation of the total archives idea as originally expressed by Wilfred I. Smith. See. W.I. Smith, "'Total Archives': The Canadian Experience," In Tom Nesmith, ed., *Canadian Archival Studies and the Rediscovery of Provenance* (Metuchen, 1993): 133.