ARCHIVES AND MANUSCRIPTS, 2017
VOL. 45, NO. 1, 18–30
http://dx.doi.org/10.1080/01576895.2016.1270838

Routledge
Taylor & Francis Group

Appraising email (using digital forensics): techniques and challenges

William P. Vinh-Doyle

Digital Preservation Archivist, Provincial Archives of New Brunswick, Fredericton, New Brunswick, Canada

ABSTRACT
With the growth of digital records, particularly email, archives must adapt their methods of acquiring, appraising and providing access to records. As one possible solution, the Provincial Archives of New Brunswick advocates the use of digital forensics. This article moves beyond simply advocating a technical solution, however, to include a fuller understanding of the challenges archivists may encounter when appraising email, such as the discovery of personal information, personally identifiable information and other information that is not reflected in traditional correspondence.

KEYWORDS
Email; digital forensics; EnCase

Archivists have long recognised the importance of preserving correspondence as part of archival collections. Manuscript collections or record series will undoubtedly include ‘correspondence’ or ‘letters’ in the fond- or series-level description. The William Lyon Mackenzie King Papers (MG26), for example, held at Library and Archives Canada, contain ‘letters on a wide range of political subjects with many different individuals. They include letters from cabinet ministers regarding government policy, letters regarding heads of state, as well as unsolicited congratulatory telegrams’.[1] Even with this brief snippet, it is clear that these letters hold archival value and would be of interest to researchers studying the internal workings and decision-making processes of the Liberal Government or external affairs.

Correspondence as a type of record has evolved from traditional analogue formats, such as letters, to include more complex digital formats, such as word-processed documents, social media and email.[2] While most archives organisations have invested resources in an effort to collect traditional analogue correspondence, the same attention to modern forms of communication, such as email, has not emerged.[3]

The volume of email messages an archival institution may receive as part of records acquisitions has also created a problem as they struggle to balance a backlog of existing analogue records and the emergence of digital records.[4] Recent statistics, for example, show that in the course of one day email users send and receive a total of 205 billion email messages.[5] Of these email messages, business users send and receive 112 billion. It is further projected that on average a single business user will receive 88 email messages per day. Of those received, 76 are legitimate email messages relating to work, while 12 are expected to be non-records (for example, spam). It is also expected the average business user will send

CONTACT William P. Vinh-Doyle ✉ William.Vinh-Doyle@gnb.ca

34 email messages per day.[6] Based on these numbers, it is estimated that in the course of one year a business user will receive up to 19,152 legitimate email messages, and send close to 8568 email messages[7] These numbers are expected to rise by 3% per year over the next four years.[8] Without an adequate email archiving solution, archivists are faced with the difficulty of appraising and preserving an exponential growth of unmanaged email.

As one possible solution, this article advocates the adoption of digital forensics to acquire, appraise, select (weed) and provide access to digital records, including email. This article moves beyond simply advocating a technical solution, however, to include a fuller understanding of the challenges archivists may encounter when appraising government email.[9] As this case study will illustrate, we were concerned about the accidental release of personally identifiable information. We were also faced with several ethical dilemmas and challenges, such as concepts of balancing the privacy of the individual versus the potential informational value of the records. While these challenges are not unique to email messages, the sheer volume of information and the preconceived idea that government email messages would remain private led the senders and receivers to participate in a more open dialogue than normally would have occurred with traditional correspondence, such as writing a formal letter.

Why an information management policy is not enough

To manage email, archivists and information management professionals have adopted policies encouraging (or requiring) users to file relevant email messages with other business records and delete transitory and non-records. The Corporate Information Management unit within the Provincial Archives of New Brunswick (PANB), for example, has published 'Managing E-mail – What to Keep and What to Delete'. It specifically declares that all email must be managed (inbox, sent items, deleted items, drafts and any other folders) 'in the same way you would manage records in other forms, such as paper'.[10] While some public bodies have insisted on adopting a strategy of printing email and filing the paper version, experience so far suggests that only a small volume of email messages are printed and filed or filed electronically. Although we have not undertaken an extensive study to determine why organisations are not filing their email, informal discussions with government employees reinforce what we have learned from other studies.

Early studies suggest that email management practices are not working as users continue to manage (or mismanage) their email in an ad hoc fashion.[11] Steve Whittaker and Candace Sidner raised the mismanagement of email twenty years ago. According to them, 'email was originally designed as a communications application, [but] … is now being used for additional functions, that it was not designed for, such as task management and personal archiving'.[12] As users' behaviours changed and fewer users managed their email and used it for different purposes, some users experienced an email overload. In an investigation of email practices, Whittaker and Sidner found that only 28% of users frequently filed their email. Another 39% filed once their mailboxes got too large (spring cleaners), and the other 33% did not file their email. Whittaker and Sidner listed a number of reasons why users were not managing their inboxes, including a concern about failing to remember where an email was filed, being unsure of its value as a record and not being able to remember the title of the folder they created. As a result, some users kept everything and used the full-text search to find individual messages.[13]

Ten years later, Danyel Fisher, AJ Brush, Eric Gleave and Marc Smith worked to reproduce the study by Whittaker and Sidner. They found the total mailbox sizes had increased from an average of 2482 messages in 1996, to 28,660 in 2006. The number of email messages users received also increased from 49 to 87 per day. Folders created by users had also increased from 47 to 133. Overall, their study confirmed the findings of Whittaker and Sidner's, namely that email users could be categorised into one of three groups: no filers (8–32%), spring cleaners (41–64%) and frequent filers (21–27%).[14]

Based on the above studies, Mark Brogan questioned why recordkeeping professionals would insist on email management practices. As he noted,

> … evidence on filing behaviour shows not only the limitations of policy and procedure type approaches to compliance and email archiving, but also its naïveté. If users are not systematically filing emails, and filing is adding to stress and lost productivity, why do recordkeeping professionals continue to insist that emails be filed to corporate stores, folders or anything else? Turning users into filing clerks is at odds with what is known about user responses to the problem of overloaded inboxes. Further, if overload is increasing and the association between filing behaviour and mail volume suggested by Whittaker and Sidner is true, then the existing tension between user behaviour and assumptions made by recordkeeping professionals can only grow.[15]

He further questioned the advantages of deploying an electronic document and records management system to assist with email management as such a system only 'works for frequent filers'.[16]

With the failure of email management to gain ground in the last 20 years, archivists can expect to receive an exponential growth of unmanaged email from users, including government employees. Digital forensics is one possible method archivists can utilise to manage these email messages as they are acquired from government.

Why adopt digital forensics?

In 2010, owing to the growing demands associated with digital preservation, PANB formally adopted a digital preservation program to complement the work already achieved in preserving sound and moving images. Since then, it has worked to develop its internal policies, standards and guidelines to meet the requirements of the Open Archival Information System, while actively acquiring digital records from public bodies and private donors.

The growth in the number of digital records received from government and private donors during this time has continued to grow. Among the noticeable growth in government records has been the acquisition of email. Since 2010, PANB has acquired approximately 2 million email records from various senior officials. It has also experienced an increase in the number of requests for information relating to these email messages from researchers and legal firms acting on behalf of the province.

One of the greatest challenges we faced when the digital preservation program was first being implemented was our inability to view the records owing to software obsolescence. If we could not view the record in its original format, we attempted to migrate or normalise the record to a readable format to make an appraisal decision. While migration or normalisation is a normal step in the process of preservation, we felt this led to wasted time and effort, particularly in cases where the records were later selected as a non-record, based on the archives' appraisal and selection policies.

Email required some additional considerations as it was transferred to the archives in a .pst file. While we had the option to open, search and view the records in Microsoft Outlook, there was concern about the possibility of altering the record. There were additional concerns about maintaining the files in a .pst container owing to the known risks associated with this file format. Since a .pst file contains possibly thousands of email messages, the loss of a single .pst file could result in a catastrophic loss. As an early solution we purchased an application called Aid4Mail, which was used to unpack the .pst file to multiple .msg files.[17] This allowed us to search the email utilising Windows Explorer. This method to search and retrieve records, however, proved to be slow and took hours to complete. In at least one case a slip of the hand resulted in messages being moved to another location, eventually requiring the archives to retrieve the email from the government backups to ensure no information was lost.

These challenges led us to conduct a more thorough review of the processes other archival institutions had implemented to acquire and select their records. Unfortunately, few provincial archives in Canada have implemented a strategy to systematically select their digital records. The archival literature on the selection of digital records however noted the benefits of using digital forensics as a solution to select and arrange digital records, leading PANB to consider its role in the archival process.

Understanding digital forensics: a few notable articles

The concept of digital forensics first emerged with the advent and growth of the micro-computer during the 1970s and 1980s. It did not emerge, however, as a formal discipline until the 1990s. Its use was largely confined to police investigations or other types of investigations such as inappropriate use of workplace computers. While it is unclear when archives first started to adopt digital forensics, Jeremy Leighton John from the British Library discussed his organisation's adoption of digital forensics beginning in 2008 as part of a project to 'develop and put into place the means with which to secure the personal archives of individuals in the digital era'.[18] Matthew Kirschenbaum, Richard Ovenden and Gabriela Redwine soon thereafter published the Council on Library and Information Resources report *Digital Forensics and Born-Digital Content in Culture Heritage Collections*, which advocated for cultural heritage institutions including archives to adopt digital forensics as part of its archival workflow to ensure the authentic transfer and appraisal of digital records. The report not only effectively illustrated the advantages of utilising digital forensics, but also provided case studies where digital forensics had been integrated as part of the archival workflow.[19]

Sounding the alarm on current archival practices, Christopher Lee and Kam Woods warned collecting institutions to improve their methods of discovery, identification and redaction of sensitive information, or risk losing the trust of donors and face a potential backlog of unprocessed material owing to the intensive manual procedures required to adequately process the collections.[20] They suggested that collecting institutions adopt digital forensics to image external media and discover private data as well as personally identify information.[21] Among the potential data that might contain personal and personally identifying information were 'emails, and email addresses that are personal and contain personally identifying (and in some cases private) information'.[22] It was suggested that the discovery and categorisation of such information using digital forensic software might save on the arduous, labour-intensive process of manually reviewing personally identifying data. Ben

Goldman and Timothy Pyatt further reiterated the advantages of digital forensic tools, and their ability to 'automate the identification and remediation of private or sensitive information found in digital files' including email, which has emerged 'as a hot zone for privacy risk'.[23] They advocated greater cooperation among collection institutions suggesting the need for defined policies, strategies and practice that could be shared.[24]

Digital forensics as a solution: FRED and EnCase

While it is unclear how many archives have implemented digital forensics into their workflows, few if any provincial archives in Canada have adopted it. As a first step we reviewed different hardware and software options, ultimately choosing to invest in Guidance Software's EnCase and a Forensic Recovery Evidence Device (FRED).

The FRED system was the preferred choice among archival institutions looking to conduct forensic investigations. Each FRED computer is equipped with an Ultrabay 3d write blocker to ensure information cannot be written to the external media, thus preserving the 'integrity of the file metadata, such as timestamps that may be relevant to the investigation'.[25] It is also equipped with various connectors, allowing the archives to acquire media cards, CD, DVD, Blu-Ray, M-Disks and USB connected devices.[26] It also has the ability to acquire internal hard drives.

In our review of digital forensic software EnCase offered many of the same features of its competitors (FTK, Bitcurator and so on). The driving force to adopt EnCase over other software, however, resulted from the fact that the Government of New Brunswick was already using EnCase to conduct internal investigations. By adopting the same software we could share ideas and collaborate internally if needed.

Although FRED came pre-installed with software to create a disk image of external media, PANB chose to use EnCase to acquire the data. EnCase provided the flexibility to create a disk image or to capture an individual folder or file.[27] The decision of when to create a disk image or capture a single record depends on how the record was transferred to PANB. When acquiring records from the government, PANB creates a disk image of any record transferred to us on 3.5 inch floppy disk, CD, DVD, USB connected devices or hard drives. Records transferred through a Secure FTP are captured as a logical evidence file.

While the decision to create a disk image of government records was easily made, some ethical concerns were expressed about creating a disk image of external and internal media received from private donors. By creating a disk image it may be possible to capture older records which the donor did not intend to donate and had previously deleted or hid. This information once ingested into EnCase could be recovered. While this has some benefits where potential donors have accidently deleted or lost records, archival institutions may unknowingly be acquiring personal information, or personally identifiable information. As Ben Goldman and Timothy Pyatt have advocated, archives need to develop clear policies and statements outlining their intentions and approach to the recovery and identification of information, such as the intention to discover deleted and hidden information.[28] They also advocated that archival institutions engage with donors during the early stages of donation to discuss privacy issues and evaluate potential personally identifiable information in the collections.[29] After some discussion, PANB decided not to create a disk image of any media without the explicit permission of the donor. We considered it unethical to review deleted information obtained from a private donor without permission.[30]

Similar to other forensic software, EnCase presents the records (evidence) in three viewing panes: tree pane, table pane and view pane. The tree pane provides a standard hierarchical view of the folder structure, similar to Windows Explorer. The table pane provides more detailed information (metadata) about the records, including the name of the file or folder, file extent, file size, the date the file was last accessed, created and written, as well as fixity check information.[31] The view pane provides multiple viewing options for records, including a report tab, which is used to review email or metadata, a text tab to display files in ASCII or Unicode text, a hex tab to display files in hexadecimal, a doc tab to display native views of formats supported by Oracle Outside In Technology, a transcript tab to display basic text files without formatting, and a picture tab to display graphic files.[32] Oracle Outside In Technology was designed to provide a solution for software developers to access (view) legacy, specialty and modern file formats.[33] By incorporating this technology EnCase can view over six hundred different file formats. With this capability we are able to quickly assess and select records for archival value. This led to some efficiency when selecting digital records for archival value, eliminating our need to migrate or normalise records to a long-term preservation format before selecting the records.

Whereas the majority of records can be reviewed upon ingest, the EnCase evidence processor is required before you can begin your evaluation of the records. It is designed to unpack compound files such as .zip and find email such as .pst files. It also creates thumbnail images to reduce the load time for larger images in a collection and finds Internet artefacts, such as browser histories and cached webpages.

When selecting records for archival value, we also utilise tags to assist in the selection and restriction process. As a first step, the information is reviewed and tagged as a ‘record’ or ‘non-record’ depending on its business or historical value. Once the information is selected as archival, we conduct a second review to determine if any restrictions need to be applied to the information. We created tags based on section 10(3) of the *Archives Act*, which provides the legislative framework for what information is restricted to the public. For example, section 10(3)(h.1) restricts records which contain opinions or recommendations provided to a minister. By labelling one tag as 10(3)(h.1) we are able to identify the records’ specific restrictions, and later export these records based on the restrictions.

EnCase also makes it possible to have multiple archivists review the same collection or multiple collections utilising the EnCase Review function. When using this function, EnCase will produce an HTML application (.hta file) which operates independently of the EnCase software. Attachments and records are exported into a separate folder and linked to the .hta file. When reviewing records in the .hta file, the reviewer can manage tags by adding or deleting tags to a collection. While this has improved our ability to analyse multiple collections without the need for additional EnCase licence agreements, there are some limitations to this feature, such as noticeable lag with larger collections. Once the reviewer has finished tagging the records they can export the tags as an HTML file and send them via email to the digital archivist. These can then be imported back into EnCase.

While tagging a record helps us during the appraisal process, bookmarks help us organise search requests and provide a means to organise records in a report. The user has the ability to create as many folders and sub-folders as needed. After an initial search is completed we use the bookmarks option to save our searches. In the bookmark tab we can move the location of folders or transfer records to other folders.

The bookmark tab also provides the option of adding comments to folders or individual records. During any search we add a note in the comment field indicating the request number and the search string used to complete the search. This has been particularly useful for us to maintain a record of the search and to provide valuable information to the researcher.

While we have used bookmarks primarily to organise search requests, it can also be used for arrangement. While we maintain the original file structure of all email accounts (for example, keeping inbox, sent items and deleted items separate), we plan to utilise this function to better organise and arrange our unstructured digital collections received from private donors, similar to how the STOP AIDS project has used bookmarks.[34]

Once the records have been arranged and described, EnCase has the option to create a forensic report. This report mirrors the structure that is created in the bookmarks tab. When the report is created, email messages become embedded into the report, whereas attachments or other records are added to the report as linked records. While basic metadata about these records is included in the report, the records themselves are exported into a separate folder for the researcher to review. The report also includes any comments created in the bookmarks tab such as restrictions.

Archivists also have the option of exporting the records out of EnCase. As part of the workflow we export archival records out of EnCase separately from records we have restricted. The export function allows you to maintain the original file structure or to export into a new folder hierarchy. While no migration or normalisation takes place during this process, records stored in a .zip or .pst will be exported out of the container and into the original format they were created in.

From a technology perspective, digital forensics has enhanced our ability to meet our mandate of acquiring and providing access to records relating to the history of the Province of New Brunswick. It has provided us a method to adequately view various file formats, including obsolete file formats. As we engaged in our first selection process we quickly discovered the value of item-level review and the challenges email messages present for archives on a broad level.

Discovering the good, the bad and the ugly

While digital forensics hardware and software can undoubtedly assist archivists in solving many of the technical challenges associated with software obsolescence, it can also be used to help archivists discover valuable information for clients. EnCase, for example, has the ability to complete a search pattern for personally identifiable information such as credit card information, social insurance numbers, phone numbers and email addresses. Records can also be indexed during the processing stage providing improved search functionality. Once processed, the text (including attachments in email messages) and metadata are discoverable. Searches can be completed on a single word, phrase or groups of words. When searching for a single word EnCase will populate suggestions found within the collection that contain the word. For example, when searching the word 'star' EnCase will suggest other words it found containing the word 'star', such as 'stars', 'starring', 'starting' and so on. It will also indicate the number of times (hits) the term was found and the number of records containing the term. The ability to create complex search strings has been particularly useful for conducting public inquiries. Whenever possible we work with the researcher to create the search string, explaining how Boolean operators can affect the search. We can further limit the results of a search string by utilising various filters that come pre-installed with EnCase. In some cases, we have downloaded additional filters or scripts which were created by the EnCase community.[35]

In the process of reviewing email messages for clients, we have discovered valuable business and historical information. Email is often used to communicate major decisions affecting various aspects of government including Memoranda of Executive Council, legal opinions, recommendations from lawyers or other civil servants, and other discussions. We also identified a number of email messages from constituents sent to their political representatives which, while valuable resources, raised issues of privacy.[36] While the discovery of the above-noted information may be considered 'good', we have also discovered the 'bad' and the 'ugly', reinforcing what M Taylor and others have already noted, that the corporate misuse of email may include various offences such as 'fraud, accessing or distributing pornography, harassment, and industrial spying amongst others'.[37] Although we have yet to find any illegal records, we have discovered information that could also be considered controversial and politically sensitive. As Andrew Waugh has stated, almost every investigation, royal commission, audit report or investigative journalism has discovered critical information in email. It is often the 'smoking gun'.[38] While the discovery of such information might be celebrated by some organisations, it is important to be cognisant of the power relations involved with collecting government email, particularly when the email relates to the senior management in your organisation. Although questions of power relations are beyond the scope of this article, Rodney Carter reminds us that 'groups display power over weaker elements in society', and 'where this power exists, there is an unequal relations between the groups'. These unequal relations can influence others 'through the control of resources, including information'.[39] While Carter was speaking in broader terms in relation to archives, government archives must consider their position within a power hierarchy before they assume the responsibility for collecting and appraising email messages.

With the potential discovery of illegal and politically sensitive records, we have taken a proactive approach in the discovery of such records.[40] Digital forensics has helped us discover this information in two ways. The first involves the review of images within a collection. The second involves utilising a search string to help identify potential flagrant text in an email message.

As part of the processing function, EnCase creates thumbnails of all images within an email message, including images located in an email attachment. These images can be reviewed separately from the record in the thumbnail tab. The ability to view these images separately has allowed PANB to quickly identify records which would normally be overlooked when reviewing them in the traditional email thread. During one investigation, for example, PANB discovered a small number of misogynistic images. These images were sent as part of a joke which degraded and sexually exploited women. Although the initial reaction was to delete these messages, the issue of their potential research value was raised in light of increased interest among historians, psychologists and sociologists in studying workplace sexual harassment and gender discrimination.

As Barbara Ritter has noted, the increased use of computer-mediated communication has provided new avenues for sexual harassment (including gender harassment, unwanted attention and sexual coercion) to pervade the workplace.[41] This has included 'active verbal (e.g. ask a coworker for personal, nonwork-related information online), active graphic (e.g. send your coworkers erotic pictures to their email), passive verbal (e.g. use an erotic term for user id at work), and passive graphic (e.g. view pornographic pictures on your office computer)'.[42] Although there are various factors relating to an individual's decision to engage in sexual harassment, archivists will be faced with evaluating this type of record particularly when acquiring email messages.

While active graphic forms of sexual harassment may be easier to discover, active verbal communication requires a more nuanced approach. To assist archives in the discovery of textual communications, an organisation can utilise the index search function similar to the way in which Laura Wilsey and others used digital forensics to identify and filter records for personal information relating to the STOP AIDS project. As they explain, 'we wanted to restrict all files containing personally identifiable information of program participants, such as name, address, telephone number, date of birth, gender identification and sexual orientation identification'.[43] They used the index search function to review a list of 18 search terms 'that might indicate the presence of sensitive content'.[44] Results returned hundreds and thousands of files which required a team of archivists to manually review and restrict.[45]

Using this same method, PANB created a generic search string to discover not only personally identifiable information, but also active verbal communications. For example, based on a current review of email messages PANB made the decision to include search terms such as 'bitch', 'slut', 'cunt', 'sex' and 'whore' to identify messages that contain sexual harassment or misogynistic content.

By utilising these terms in the search string, PANB has discovered a number of records that have been flagged as a result of their content. In one example, the sender had asked two male colleagues if they knew a particular female colleague. One responded with the question, 'Is she hot?', to which the sender simply replied, 'No. A bitch'. In another case, a sender emailed his colleague about meeting 'two sluts' and asked his friend if he was interested in meeting them. In yet another chain of email messages it was discovered that an employee was having an affair. These messages contained sensitive and sexually explicit dialogue between the two parties involved.

The discovering of active graphic or active verbal forms of sexual harassment in email messages is not that surprising given the misconception that email is a private communication tool. While it may be controversial in some circles and risky to keep any of the above-noted records, it was our belief that the email messages could be valuable to those interested in studying workplace sexual harassment and misogyny. As Terry Cook once reminded us,

> In many societies, certain classes, regions, ethnic groups, or races, women as a gender, and non-heterosexual people, have been de-legitimized by their relative or absolute exclusion from archives, and thus from history and mythology – sometimes unconsciously and carelessly, sometimes consciously and deliberately.[46]

Unfortunately, this could be extended to records promoting misogyny, sexual harassment or other forms of discrimination.

While the decision to keep these records for long-term preservation was based on the potential research value, we also recognised they could potentially embarrass the senders and receivers and lead to a political scandal.[47] This in turn could have negative consequences for PANB and result in undue pressure to destroy the records, or change existing retention and disposition schedules. It is therefore important that any archivist responsible for appraising email messages maintain an open dialogue with other archivists and senior government officials.

Bringing it all together: next steps for archives

By utilising digital forensics, an archive can reduce the length of time to complete a search for information or help in the process of selection and arrangement. This is particularly important as archives face decreasing revenues, and have fewer archivists to select and search an exponential growth of digital records. While the combination of hardware (FRED) and software (EnCase) provided a solution to some of our technical issues and improved other areas of service such as the ability to search for records in a timely fashion, we are aware that the traditional practice of reviewing records at an item-by-item level for restrictions may need to be reconsidered as the growth of records increases. Based on our existing resources and the exponential growth of email messages, we may need to leverage the forensic technology and its ability to discover records using a search string to find restricted information, without reviewing records item-by-item. Such an approach, however, means PANB must assume a certain amount of risk. How much risk an archive is willing to assume depends on the organisation.

Endnotes

1. Library and Archives Canada, William Lyon Mackenzie King Papers (MG26), available at <*http://heritage.canadiana.ca/view/oocihm.lac_mikan_108977?page=2*>, accessed 8 December 2016.
2. Jackie Dooley, *The Archival Advantage: Integrating Archival Expertise into Management of Born-Digital Library Materials*, OCLC Research, Dublin, OH, 2015, p. 9.
3. Donghee Sinn, Sue Yeon Syn and Sung-Min Kim, 'Personal Records on the Web: Who's in Charge of Archiving, Hotmail or Archivists?', *Library & Information Science Research*, vol. 33, no. 4, 2011, p. 320.
4. The Provincial Archives of New Brunswick (PANB), for example, is legislated by the *Archives Act* to preserve the records of the province. Section 5(1)(e), for example, specifically states that PANB has the responsibility to 'discover, collect and preserve records having any bearing upon the history of New Brunswick'. For further information see *Archives Act* (S.NB. 1977, c. A-11.1), available at <*http://laws.gnb.ca/en/showfulldoc/cs/A-11.1/20120313*>, accessed 8 December 2016. Similar legislation granting legislative powers to PANB exists throughout each of the provinces and territories.
5. The Radicati Group, *Email Statistics Report, 2015–2019*, The Radicati Group Inc., Palo Alto, CA, March 2015, p. 3.
6. The 122 emails sent/received per user per day were estimated to include 88 emails received (76 legitimate, 12 spam) and 34 emails sent. The Radicati Group, p. 4.
7. These figures are based on 252 working days.
8. The Radicati Group, p. 4.
9. While this article does not advocate a particular appraisal policy, the following articles, policies or guidelines may assist institutions in drafting such policy. Queensland State Archives, *Queensland State Archives' Appraisal Statement*, Queensland State Archives, Brisbane, 2013; Angus Whyte and Andrew Wilson, *How to Appraise and Select Research Data for Curation*, DCC How-to Guides, Digital Curation Centre, Edinburgh, 2010; National Archives and Records Administration, *Strategic Directions: Appraisal Policy*, NARA, College Park, MD, 2007; Ross Harvey, 'Instalment on "Appraisal and Selection"', in Seamus Ross and Michael Day (eds), *DCC – Digital Curation Manual*, HATII, University of Glasgow, Glasgow, 2007; National Archives of Australia, *Why Records Are Kept: Directions in Appraisal*, National Archives of Australia, Canberra, 2003; Maynard Brichford, *Archives & Manuscripts: Appraisal & Accessioning*, Society of American Archivists, Chicago, IL, 1977.

10. Provincial Archives of New Brunswick, 'Managing Email – What to Keep and What to Delete', available at <*http://www.archives.gnb.ca/Archives/RecMan.aspx?Section=2&culture=en-CA*>, accessed 8 December 2016.
11. Mark Brogan, 'Clipping Mercury's Wings: The Challenge of Email Archiving', *Archives and Manuscripts*, vol. 37, no. 1, 2009, p. 17.
12. Steve Whittaker and Candace Sidner, 'Email Overload: Exploring Personal Information Management of Email', in Ralf Bilger, Steve Guest and Michael J Tauber (eds), *CHI '96: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* ACM Press, New York, 1996, p. 276.
13. Whittaker and Sidner, p. 280.
14. Danyel Fisher, AJ Brush, Eric Gleave and Marc Smith, 'Revisiting Whittaker & Sidner's Email Overload Ten Years Later', in *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*, ACM Inc., Vancouver, 2006, p. 310.
15. Brogan, p. 19.
16. ibid. Brogan proposed the adoption of active email archiving, which involved capturing and storing all email sent and received, where it would then be indexed and provide improved search functionality. Single-instance storage could also be applied to the records, ensuring only a single copy of each message is stored. As he notes, however, active email archiving requires that all email be archived, including non-records.
17. Aside from unpacking the .pst containter to a .msg format, we also used this application to migrate the email to a .eml and .pdf format for preservation. For further information on Aid4Mail visit their website at <*http://www.aid4mail.com*>, accessed 8 December 2016.
18. Jeremy Leighton John, 'Adapting Existing Technologies for Digitally Archiving Personal Lives: Digital Forensics, Ancestral Computing, and Evolutionary Perspectives and Tools', paper presented at iPRES 2008: The Fifth International Conference on Preservation of Digital Objects, London, The British Library, 29–30 September 2008.
19. Matthew G Kirschenbaum, Richard Ovenden and Gabriela Redwine, *Digital Forensics and Born-Digital Content in Cultural Heritage Collections*, Council on Library and Information Resources, Washington, DC, December 2010.
20. Christopher A Lee and Kam Woods, 'Automated Redaction of Private and Personal Data in Collections', in Luciana Duranti and Elizabeth Shaffer (eds), *Proceedings of The Memory of the World in the Digital Age: Digitization and Preservation. An international conference on permanent access to digital documentary heritage, 26–28 September 2012, Vancouver, British Columbia, Canada*, UNESCO, 2013, p. 299.
21. Lee and Woods define private data as including, but not limited to, 'social security numbers, credit card numbers, financial records, medical records, employment information, education records, passwords, and cryptographic keys, and local and online account records'. Personal and personally identifying information included 'information that belongs to a specific individual (or group of individuals). Users are typically aware that contact names, telephone numbers, emails, and email addresses are personal and contain personally identifying (and in some cases private) information', pp. 299–300.
22. ibid., p. 300.
23. Ben Goldman and Timothy Pyatt, 'Security Without Obscurity: Managing Personally Identifiable Information (PII) in Born-Digital Archives', *Library & Archival Security*, vol. 26, nos 1–2, 2013, p. 44.
24. ibid., p. 52.
25. Sally Vandeven, *Forensic Images: For Your Viewing Pleasure*, SANS Institute, 2014, p. 10.
26. PANB uses an USB adapter to acquire 3.5 inch floppy disks, as well as zip disks. It is currently investigating the ability to create a USB connected device to read 5.25 inch floppy disks.
27. EnCase provides the ability to save image files as an EnCase evidence file or logical evidence file. 'With a logical evidence file, you can selectively choose which files or folders you want to preserve, instead of acquiring the entire drive. Unlike copying files from a device and altering critical metadata, logical evidence files preserve the original files as they existed on the media and include additional information such as file name, file extension, last accessed, file created,

last written, entry modified, logical size, physical size, MD5 hash value, permissions, starting extent, and original path of the file'. For further details see <*http://www.edrm.net/resources/glossaries/collection-standards/logical-evidence-file*>, accessed 8 December 2016.

28. Goldman and Pyatt, p. 47.
29. ibid., p. 48.
30. Ben Goldman has noted that some archives 'have developed policies explicitly stating their intention not to preserve hidden or deleted files accidently transferred to an archive, believing such steps to be a violation of donor privacy and intent. Such policies need not take a firm stance on such data, however. Some institutions may instead decide to take a more nuanced approach and simply acknowledge the likelihood of encountering such data and give an appropriate amount of attention'. ibid., p. 47.
31. EnCase provides over a thousand metadata fields that can be displayed in the table pane.
32. EnCase can view file formats supported by Oracle Outside In Technology. A list of file formats is available on the Oracle website at <*https://docs.oracle.com/outsidein/850/supportdocs/ds_oitfiles_8_4_1.pdf*>, accessed 8 December 2016.
33. For further information regarding Oracle Outside In Technology, see the company website at <*http://www.oracle.com/us/technologies/embedded/025613.htm*>, accessed 8 December 2016.
34. Laura Wilsey and colleagues have described their processes and the advantages of using digital forensics to capture and process digital files relating to the STOP AIDS project. The case study includes discussion on how they used tags and bookmarks as part of the process. Laura Wilsey, Rebecca Skirvin, Peter Chan and Glynn Edwards, 'Capturing and Processing Born-Digital Files in the STOP AIDS Project Records: A Case Study', *Journal of Western Archives*, vol. 4, issue 1, article 1, 2013.
35. PANB currently utilises various enscripts or filters in the process of discovery.
36. The discovery of constituency email provided some ethical challenges for PANB. Traditionally, constituency records were only acquired through a private donation from a Member of the Legislative Assembly (MLA). With the introduction of email, official records relating to the government business and records relating to their position of MLA were often intertwined. Generally the email sent or received by an MLA also contained personally identifiable information and sensitive information. Discussions over PANB's acquisition and selection of these email centred on whether we should retain or remove these email messages and request permission from the MLA to keep the records as part of a private donation. Upon further discussion we determined that constituency records were in fact business records relating to the province, since the MLA had used his or her government email account. MLAs had also been clearly instructed at the beginning of their term to separate constituency records from official government email. We believe it is still necessary to maintain some form of control over these records by identifying constituency records where possible utilising a separate tag.
37. M Taylor, J Haggerty and D Gresty, 'The Legal Aspects of Corporate Email Investigations', *Computer Law & Security Review*, vol. 25, no. 4, 2009, p. 372.
38. Andrew Waugh, 'Email – A Bellwether Records System', *Archives and Manuscripts*, vol. 42, no. 2, 2014, p. 216.
39. Rodney GS Carter, 'Of Things Said and Unsaid: Power, Archival Silences, and Power in Silence', *Archivaria*, no. 61, Spring 2006, p. 218.
40. Lee and Wood, pp. 299–300.
41. According to Ritter, 'Gender harassment involves misogynist behaviors that are insulting, hostile, or degrading towards women. Unwanted sexual attention corresponds closely with the legal notion of creating a hostile work environment, and may involve behaviours such as sexual comments about dress, touching, and display of sexual materials. Finally, sexual coercion is similar to quid pro quo [Sexual harassment] in which one is bribed or threatened to perform sexual acts in exchange for some job-related benefit', see Barbara Ritter, 'Deviant Behavior in Computer-Mediated Communication: Development and Validation of a Measure of Cybersexual Harassment', *Journal of Computer-Mediated Communication*, vol. 19, no. 2, 2014, p. 198.

42. ibid., p. 202.
43. Wilsey et al., p. 17.
44. ibid., p. 18.
45. ibid., p. 18.
46. Terry Cook, 'We Are What We Keep; We Keep What We Are: Archival Appraisal Past, Present and Future', *Journal of the Society of Archivists*, vol. 32, no. 2, 2011, p. 174.
47. PANB decided to delete the emails detailing the office affair. There was little archival or research value attributed to these emails.

Acknowledgements

I would like to thank my manager, Allen Doiron, for supporting my work in digital forensics. Thanks also need to be extended to Fred Farrell, Tom McCaffrey and Britany Sparrow. Appreciation is also extended to friends and colleagues who reviewed early drafts of the article and provided necessary feedback.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes on contributor

William P. Vinh-Doyle is currently the digital preservation archivist at the Provincial Archives of New Brunswick. He has a PhD in Canadian History from the University of New Brunswick. His interests include Canadian labor history, social history, digital humanities, and archival studies.