The Canadian Journal of Information and Library Science / La Revue canadienne des sciences de l'information et de bibliothéconomie

Volume 39, Number 2, June / juin 2015

Reviews are available online at Project MUSE, *Canadian Journal of Information and Library Science*, 39, 2:

Les comptes rendus sont disponibles en ligne à Project MUSE, *La Revue canadienne des sciences de l'information et de bibliothéconomie*, 39, 2 :

http://muse.jhu.edu/journals/canadian_journal_of_information_and_library_science/

The Canadian Journal of Information and Library Science / La Revue canadienne des sciences de l'information et de bibliothéconomie

The *Canadian Journal of Information and Library Science* continues the *Canadian Journal of Information Science* and is published by the Canadian Association for Information Science (CAIS). Its purpose is to contribute to the advancement of information and library science in Canada. The Journal is published on a quarterly basis in print format and is also available online on Project MUSE. Submissions for publication should be sent to the Editor. Orders for single copies should be sent to the University of Toronto Press, Journals Division, 5201 Dufferin Street, Toronto, ON M3H 5T8, Canada. Tel.: 416 667-7810. Beginning in 1986, with volume 11, the journal is published quarterly. For subscription rates and information, please visit www.utpjournals.com.

Personal or institutional members of CAIS receive the journal free of charge as benefit of membership.

ISSN 1195-096X (print)
ISSN 1920-7239 (online)

La *Revue canadienne des sciences de l'information et de bibliothéconomie* continue la *Revue canadienne des sciences de l'information* et est publiée par L'Association canadienne des sciences de l'information (ACSI). Son objectif est de contribuer à l'avancement des sciences de l'information et de la bibliothéconomie au Canada. La revue est publiée sur une base trimestrielle en format papier et est également disponible en ligne sur Project MUSE. Les manuscrits soumis pour publication doivent être adressés au rédacteur en chef. Pour commander un numéro s'adresser à University of Toronto Press, Département des revues, 5201 rue Dufferin, Toronto, ON M3H 5T8, Canada. Tél. : 416 667-7810. Depuis le volume 11 (1986), la revue est publiée sur une base trimestrielle. Pour les tarifs d'abonnement et autres renseignements, visitez www.utpjournals.com.

Les membres individuels et institutionnels de l'ACSI reçoivent la revue sans frais supplémentaires dans le cadre de leur abonnement.

Canadian Association for Information Science: Board of Directors / Conseil d'administration de l'Association canadienne des sciences de l'information

Editorial Address / Adresse éditoriale
Clément Arsenault
École de bibliothéconomie et des sciences de l'information, Pavillon Lionel-Groulx
C.P. 6128, succ. Centre-ville
Université de Montréal
Montréal, QC H3C 3J7
Canada
Tel.: 514 343-7400
Fax: 514 343-5753
E-mail: clement.arsenault@umontreal.ca

Subscriptions and Advertising / Abonnement et publicité
University of Toronto Press—Journals Divsion
5201 Dufferin Street
Toronto, ON M3H 5T8
Canada
Tel.: 416 667-7810
Fax: 416 667-7881
E-mail: journals@utpress.utoronto.ca

The Canadian Journal of Information and Library Science **is published four times per year in March, June, September, and December for the Canadian Association for Information Science, by University of Toronto Press. The *Journal* is available on microfilm through Micromedia Limited, Toronto.**

Periodical postage paid at Buffalo, NY and additional mailing offices. U.S. Postmaster: Send address changes to University of Toronto Press, Journals Division, 2250 Military Road, Tonawanda, NY 14150-6000. Canadian Postmaster: Send address changes to University of Toronto Press, 5201 Dufferin Street, Toronto, ON M3H 5T8.
Publications Mail Agreement No. 40600510
Printed in Canada

Preface

Luciana Duranti, Guest Editor

The cloud—on-demand access to a network of a shared pool of configurable computing resources—heralds unprecedented challenges for archivists as well as for records and information managers. Data, records, and archives are increasingly entrusted to Internet providers who offer on-demand online storage at a low cost, protecting materials with a level of security that no single organization can afford and maintaining them in formats compatible with any user's system. However, the cloud environment is neither transparent nor regulated. Those who create, manage, appraise, control, and preserve these stored materials encounter problems related to ownership, provenance, and jurisdiction, among others, as they remain responsible for such materials without control and are accountable without knowledge.

This special issue explores the challenges presented by keeping data, records, and archives in the cloud, reports on research into possible solutions, examines existing and proposed policies, procedures, regulations, and legislation, and describes cases of adoption of cloud models, law, contractual agreements, and technological infrastructure. Several of the articles contained in this issue discuss the preliminary findings of research conducted in the context of the InterPARES Trust (ITrust) project (http://www.interparestrust.org). ITrust is a five-year (2013–18) multi-national, interdisciplinary research project exploring issues concerning digital data, records, and archives entrusted to the Internet. Its goal is to generate theoretical and methodological frameworks capable of supporting the development of local, national, and international policies, procedures, regulations, standards, and legislation and to ensure public trust grounded on the evidence of good governance, a strong digital economy, and a persistent digital memory. ITrust builds on the findings of the International Research into the Preservation of Authentic Records in Electronic Systems, a project carried out in three phases from 1998 through 2012 (http://www.interpares.org).

This special issue begins with an analysis of the ideas, beliefs, and practices associated with the concepts of authenticity and trust. Corinne Rogers, the project coordinator for ITrust, discusses the concept of authenticity as it applies to digital records and reports on the findings of research on the way in which records professionals ensure and verify authenticity in practice. While Rogers describes the findings of a quantitative study, Erik Borglund, an archival scholar from Mid-Sweden University and a researcher for the ITrust control domain of the European team, presents the findings of a qualitative study conducted in Sweden on archivists' trust in the cloud. These two articles set the stage for the following four articles, which focus on specific issues raised by the adoption of a

La Revue canadienne des sciences de l'information et de bibliothéconomie 39, no. 2 2015

cloud environment for data, records, and archives storage and access, and describe research conducted in the context of ITrust and its preliminary results.

In the first of these four articles, Marie Demoulin, a scholar of both civil and common law from the University of Montreal, Jessica Bushey, a doctoral candidate in archival science at the University of British Columbia, and Robert McLelland, a professional archivist for the Delta Museum and Archives Society—all of them researchers for the ITrust legal domain of the North American team—discuss cloud service agreements and boiler-plate contracts in light of records and archival requirements, specifically authenticity and trust. In the second article, Valerie Léveillé, an archival graduate research assistant, and Katherine Timms, an information standards specialist with Library and Archives Canada, both researchers for the ITrust access domain of the North American team, analyse business processes, workflows, and documentation of open government in the Canadian jurisdiction in the context of exploring the possibility of a universal framework. In the third article, Pat Franks, a records management scholar from San José State University and a researcher for the ITrust control domain of the North American team, analyses the results of a survey conducted to identify cloud retention and disposition challenges and the way to mitigate them. In the fourth article, two Croatian scholars from the University of Zagreb, Hrvoje Stancic and Arian Rajh, and an information technology architect from the Croatian Financial Agency—all of them researchers for the security domain of the ITrust European team, discuss the long-term preservation of digitally signed records in a cloud environment, looking at the characteristics of two separate models of archives in the cloud in light of their sustainability and portability.

This special issue concludes with two case studies, thereby taking the reader from the discussion of research findings to the description of implementation. Both articles focus on the long-term preservation of archives using a cloud environment for storage and access. The first article, written by Anna Sobczak, a Polish researcher from Szczecin University, describes software developed by the State Archives of Baden-Württemberg that supports the appraisal, acquisition, description, dissemination, long-term storage, and preservation activities for public records in an online environment (or private cloud). The second article, co-authored by Bronwen Sprout, the digital initiatives coordinator for the University of British Columbia Library, and Mark Jordan, head of library systems for the W.A.C. Bennett Library of Simon Fraser University, presents a community cloud preservation service that is piloted by the Council of Prairie and Pacific University Libraries using the Archivematica digital preservation system. This final article discusses the Archivematica-as-a-service model generally and some aspects of implementation in greater detail, concluding with a discussion of future directions.

Although the risks involved in the adoption of a cloud environment for the access, storage, and preservation of data, records, and archives far outweigh the benefits, individuals and organizations, as well as memory institutions such as archives and libraries, increasingly look at the virtual environment of the Internet as the answer to the need for a secure, economic, and efficient lifecycle

management system for born digital materials—in particular, for those that were created and maintained in a cloud environment to start with and for those for which transfer to an in-house preservation system would require a tremendous number of resources. For this reason, it is essential to conduct careful research on each and every issue linked to the use of the cloud environment for data, records, and archives, to test possible solutions, and to share findings and results so that entrusting our materials to the cloud will become a viable option based on the trustworthiness of the providers rather than simply on the blind trust of the users. This special issue is a step in such a direction.

Préface

Luciana Duranti, Rédactrice en chef invitée

Avec l'informatique en nuage, c'est-à-dire l'accès sur demande à un pool partagé en réseau de ressources informatiques configurables, s'annoncent des défis sans précédent pour les archivistes, ainsi que pour les gestionnaires de documents et d'information. Les données, les documents et les archives sont de plus en plus fréquemment confiés à des fournisseurs de services Internet qui offrent le stockage en ligne à la demande pour un coût modique, avec une protection des matériaux et un niveau de sécurité qu'aucune organisation ne peut se permettre seule, et la possibilité de les maintenir dans des formats compatibles avec tous les systèmes de l'utilisateur. Cependant, l'environnement d'informatique en nuage n'est ni transparent ni réglementé. Ceux qui créent, gèrent, évaluent, contrôlent et préservent ces matériaux stockés font face à des problèmes de droits de propriété, de provenance et de compétence juridique, entre autres, car ils demeurent responsables de ces matériaux sans avoir à se soumettre à aucun contrôle et sans avoir les connaissances nécessaires.

Ce numéro spécial explore les défis que pose la conservation des données, des documents et des archives dans un nuage informatique; il offre aussi un compte-rendu des recherches de solutions possibles, examine les politiques proposées, les procédures, les règlements et la législation en vigueur, et décrit certains cas d'adoption de modèles d'informatique en nuage, avec l'aspect légal, les ententes contractuelles, et l'infrastructure technologique. Plusieurs des articles contenus dans ce numéro discutent les résultats préliminaires des recherches menées dans le cadre du projet InterPARES Trust (ITrust) (http://www.interparestrust.org). ITrust est un programme de recherche quinquennal (2013–18), multi-national et interdisciplinaire, qui a pour objectif d'explorer les problèmes que posent les données numériques, les dossiers et les archives quand elles sont confiées à l'Internet. Le but est de générer des cadres théoriques et méthodologiques capables de soutenir le développement de politiques locales, nationales et internationales, ainsi que des procédures, des règlements, des normes, de la législation, susceptibles de gagner la confiance du public, confiance fondée sur les preuves fournies par une bonne gouvernance, une économie numérique forte, et une mémoire numérique sans failles. ITrust s'appuie sur les conclusions de la recherche internationale sur la conservation des documents authentiques dans des systèmes électroniques, un projet réalisé en trois phases de 1998 à 2012 (http://www.interpares.org).

Ce numéro thématique commence par une analyse des idées, des croyances et des pratiques associées aux concepts d'authenticité et de confiance. Corinne Rogers, la coordinatrice du projet pour ITrust, aborde le concept d'authenticité tel qu'il s'applique aux documents numériques et elle rend compte des résultats

La Revue canadienne des sciences de l'information et de bibliothéconomie 39, no. 2 2015

de recherches sur la façon dont les professionnels s'assurent et vérifient l'authenticité dans la pratique. Alors que Rogers décrit les résultats d'une étude quantitative, Erik Borglund, chercheur archiviste de l'université Mid-Sweden, et chercheur pour le domaine du contrôle dans l'équipe ITrust européenne, présente les résultats d'une étude qualitative menée en Suède sur la confiance des archivistes en le nuage. Ces deux articles préparent le terrain pour les quatre articles suivants, qui se concentrent sur des questions spécifiques soulevées par l'adoption d'un environnement d'informatique en nuage pour le stockage et l'accès de données, de documents et d'archives, et ils décrivent des recherches menées dans le cadre de ITrust et les résultats préliminaires.

Dans le premier de ces quatre articles, Marie Demoulin, chercheure en droit civil et en common law de l'Université de Montréal, Jessica Bushey, doctorante en archivistique à l'Université de la Colombie-Britannique, et Robert McLelland, archiviste professionnel pour la société des archives et du Delta Museum — tous chercheurs pour le domaine légal au sein de l'équipe nord-américaine de ITrust — discutent des ententes de services et des contrats standards d'informatique en nuage à la lumière des exigences spécifiques des documents et des archives en termes d'authenticité et de confiance. Dans le deuxième article, Valérie Léveillé, assistante de recherche de troisième cycle en archivistique, et Katherine Timms, spécialiste des normes d'information auprès de Bibliothèque et Archives Canada, toutes deux chercheurs pour le domaine de l'accès au sein de l'équipe nord-américaine de ITrust, analysent les processus d'affaires, les flux de travaux et la documentation de gouvernement ouvert dans la juridiction canadienne et le contexte de l'exploration de la possibilité d'un cadre universel. Dans le troisième article, Pat Franks, chercheur en gestion des documents de l'Université d'État de San José et chercheur pour le domaine du contrôle au sein de l'équipe nord-américaine de ITrust, analyse les résultats d'une enquête menée pour identifier les défis que posent la conservation et l'élimination des documents dans le nuage, et pour trouver une façon d'en atténuer l'impact. Dans le quatrième article, deux chercheurs croates de l'Université de Zagreb, Hrvoje Stancic et Arian Rajh, et un architecte en technologie de l'information de l'agence financière croate, tous les trois chercheurs pour le domaine de la sécurité au sein de l'équipe européenne de ITrust, discutent de la conservation à long terme des documents signés numériquement dans un environnement de nuage informatique, en prêtant attention aux caractéristiques des deux modèles distincts d'archives présents dans le nuage, avec en vue leur durabilité et leur portabilité.

Ce numéro spécial se termine par deux études de cas, menant ainsi le lecteur de la discussion de résultats de recherche à la description de mises en œuvre. Les deux articles portent sur la préservation à long terme des archives en utilisant un environnement de nuage pour le stockage et l'accès. Le premier article, écrit par Anna Sobczak, chercheure polonaise de l'Université de Szczecin, décrit un logiciel développé par les Archives d'État du Bade-Wurtemberg qui permet l'évaluation, l'acquisition, la description, la diffusion, le stockage à long

terme, et les activités de préservation des documents publics dans un environnement en ligne (ou nuage privé). Le deuxième article, co-écrit par Bronwen Sprout, coordinateur des initiatives numériques à la bibliothèque de l'Université de la Colombie-Britannique, et Mark Jordan, responsable des systèmes de bibliothèque pour la bibliothèque W.A.C. Bennett de l'Université Simon Fraser, présente un service de conservation communautaire en nuage piloté par le Conseil des bibliothèques des universités des Prairies et du Pacifique, utilisant le système de conservation numérique Archivematica. Ce dernier article discute le modèle Archivematica en tant que service en général ainsi que certains aspects de sa mise en œuvre de manière plus détaillée, et conclut sur une discussion des orientations futures.

Bien que les risques liés à l'adoption d'un environnement d'informatique en nuage pour l'accès, le stockage et la conservation des données, des documents et des archives l'emportent de loin sur leurs avantages, les individus et les organisations, y compris les institutions mémorielles telles que les dépôts d'archives et les bibliothèques, se tournent de plus en plus vers l'environnement virtuel de l'Internet comme la réponse au besoin d'un système de gestion sécuritaire, économique et efficace du cycle de vie des matériaux initialement numériques, en particulier ceux qui ont été créés et maintenus dans un environnement de nuage dès le départ et ceux pour lesquels le transfert à un système de conservation en interne exigerait un nombre considérable de ressources. Pour cette raison, il est essentiel de mener des recherches approfondies sur chaque question liée à l'utilisation de l'environnement d'informatique en nuage pour les données, les documents et les archives, afin de tester les solutions possibles, et partager les conclusions et les résultats de sorte que confier nos matériaux au nuage deviendra une option viable basée sur la fiabilité des fournisseurs plutôt que simplement sur la confiance aveugle des utilisateurs. Ce numéro spécial est un pas dans cette direction.

Authenticity of Digital Records: A Survey of Professional Practice

L'authenticité des documents numériques : Un survol des pratiques professionnelles

Corinne Rogers
University of British Columbia
corinne.rogers@ubc.ca

Abstract: Authenticity of digital material is an enduring concern. However, while most people intuitively understand what authenticity is, few are able to identify exactly what is required to ensure, assess, and guarantee it. Heuristic and hermeneutic assessments of authenticity do not support any quantifiable measures of authenticity. Several important research projects have studied the means of ensuring that authenticity is protected throughout the life cycle of digital material, however, even as archival research and scholarship continue to offer insight into the nature of authentic digital objects and their preservation, new technologies, specifically distributed networked systems connected through the Internet, create new challenges to security and authenticity. This article reports on the author's research into the practices of records and information professionals to ensure, assess, and/or protect the authenticity of digital records and data.

Keywords: authenticity, trusting records, chain of preservation

Résumé : L'authenticité des matériaux numériques est une préoccupation qui dure. Mais alors que la plupart des gens comprennent intuitivement ce qu'est l'authenticité, peu sont en mesure d'identifier exactement ce qui est nécessaire pour assurer, évaluer et garantir l'authenticité. Les évaluations heuristiques et herméneutiques de l'authenticité n'appuient pas les mesures quantifiables de l'authenticité. Plusieurs projets de recherche importants ont étudié les moyens d'assurer que l'authenticité est protégée tout au long du cycle de vie du matériau numérique, cependant, tandis que la recherche d'archives et l'érudition continuent d'offrir une réflexion sur la nature des objets numériques authentiques et leur conservation, les nouvelles technologies, en particulier les systèmes distribués en réseau reliés par Internet, créent de nouveaux défis pour la sécurité et l'authenticité. Cet article rend compte de la recherche de l'auteur sur les pratiques documentaires des professionnels de l'information visant à assurer, évaluer, et / ou protéger l'authenticité des documents numériques et des données.

Mots-clés : authenticité, avoir confiance en des documents, chaîne de conservation

Introduction

The digital universe is predicted to grow by 40 percent a year into the next decade, and businesses and governments are strategizing to take advantage of new opportunities afforded by digital information in its many and diverse forms

La Revue canadienne des sciences de l'information et de bibliothéconomie 39, no. 2 2015

(Turner et al. 2014). Cloud computing offers significant economies of scale, and enterprises are rapidly moving, or considering moving, established computing practices, including digital records storage and management, to the cloud, where users relinquish a measure of control over the materials and technologies. Cloud computing now accounts for only 5 percent of total enterprise information technology (IT) spending, but this number is growing. As well as new technological infrastructure for traditional digital materials, we are entering the age of the "Third Platform," according to the International Data Corporation (IDC),[1] defined as "the next-generation compute platform that is accessed from mobile devices, utilizes Big Data, and is cloud based." Not all of the information created in the digital universe will be, or needs to be, preserved, but of the data and records that do warrant preservation, how will their authenticity be established and protected in these diverse and rapidly evolving technological contexts?

The 2014 IDC report states that maximizing opportunity requires certain imperatives for IT organizations but that "real transformation to a data-driven or software-defined enterprise is an "all-hands-on-deck imperative" not restricted to IT alone (Turner et al. 2014, 7). Those of us in records professions might well rejoice at this forecase. After all, much of the digital material being created by public and private sector organizations is in our purview to create, capture, manage, and preserve as records and data that form the foundation of business decisions and corporate and societal memory. Our professional codes of ethics hold us to principles of recordkeeping such as accountability, integrity, protection, compliance, availability, transparency, authenticity, preservation, security, protection, availability and use, privacy, and trust (Association of Canadian Archivists 1999; Society of American Archivists 2011; ARMA International 2014).

Preservation of trustworthy records regardless of medium—that is, records that can be proven authentic, reliable, and accurate—is at the core of the archival endeavour. It is about more than secure storage. Preservation encompasses all of the tools and techniques, policies, and procedures that ensure the target material remains trustworthy, accessible, and usable over time and across technological change. It requires the cooperation and collaboration of an inter-disciplinary team including archivists and records managers as well as IT professionals. In the digital universe, authenticity continues to be an enduring, if elusive, concern.

However, even though enterprises assume liability or responsibility for 85 percent of all data in the digital universe, the 2014 IDC report does not once mention issues of authenticity, trust, reliability, or integrity of this data. These qualities seem subsumed by the imperatives of security and privacy, the ability to enable and manage the explosive growth of mobile devices (and, presumably, the data they generate), and the ability to query data from wherever it may be stored (cross-boundary and cross-jurisdiction). Furthermore, the allocation of security budgets shows that prevention and protection from data breach and unauthorized access far outweighs money allocated to breach response (EMC Corporation 2013) and that the priority, in the event of a breach, is to restore

service at the expense of the protection of trace evidence that would aid investigation (Endicott-Popovsky, Frincke, and Taylor 2007). If, as David Weinberger is often quoted, transparency is the new objectivity, then security is the new authenticity.

The professional discourse of archivists and records managers and the concerns for data in the information technology sector may overlap in the areas of privacy, access, and security. However, they too often occur as conversations rather than as collaborations. And while the trustworthiness of records and data may still be viewed by the enterprise as the responsibility of records professionals, the primacy of security often places IT ahead of records management. The authority of records professionals may be overshadowed or undermined by their reliance on IT for systems implementations as well as the focus on information and data analytics as business drivers (Richards 2014).

This article reports on the results of a survey of records professionals designed to explore their practices in ensuring and assessing the authenticity of records, documents, and data for which they are responsible. The survey is not limited to questions about the management of records and data in the cloud, and, in fact, most of the respondents are working primarily in the "Second Platform"—the distributed world of LAN/Internet and client/server architectures (EMC Corporation 2013). In 2013, less than 20 percent of the data in the digital universe was stored or processed in the cloud, but, by 2020, that figure is predicted to double. Understanding how authenticity is assessed and protected for digital records and data in current enterprise service architectures is foundational to understanding the challenges records professionals face in cloud platforms. If businesses and governments rely on information and data increasingly coming from third parties, mobile devices, and sensors, authenticity will continue to be a critical issue.

The issue of authenticity

Most people intuitively understand authenticity to be the quality of genuineness, but few are able to identify exactly what is required to ensure, assess, and guarantee it. Authenticity, the quality of a record that is what it purports to be, has historically been understood as deriving from the circumstances of a document's creation, if known, and from the manner and place of its preservation. The presence of a signature indicated the agreement of the author with the content of the document and authenticated the transaction recorded therein. Signatures of witnesses or countersigners further verified the document's authenticity. Signers and countersigners could be questioned if necessary, and their testimony used as a guarantee of genuineness. Such determination of authenticity was based on observation and testimony.

This understanding of authenticity, elegant in its simplicity if challenging to apply, is at the root of archival, diplomatic, and legal theory and has been codified in records management standards. According to archival science, record authenticity is "the trustworthiness of a record as a record, i.e. the quality of a record that is what it purports to be and that is free from tampering or corruption"

(InterPARES Glossary, n.d.). *Black's Law Dictionary* (n.d.) defines "authentic" as "Genuine; true; having the character and authority of an original; duly vested with all necessary formalities and legally attested; competent, credible, and reliable as evidence." The Society of American Archivists defines "authenticity" as "the quality of being genuine, not a counterfeit, and free from tampering, and is typically inferred from internal and external evidence, including its physical characteristics, structure, content, and context." This definition closely associates a record's authenticity with the creator of the record, something that can be verified by testing the physical and formal characteristics of the record. Authenticity as a record does not automatically imply reliability of the content of the record (Duranti 1998, 46; Pearce-Moses 2005). Finally, we see this concept of authenticity codified in ISO 15489, the international records management standard as follows:

> An authentic record is one that can be proven
> a. to be what it purports to be,
> b. to have been created or sent by the person purported to have created or sent it, and
> c. to have been created or sent at the time purported" (International Organization for Standardization 2001, s. 7.2.2).

In common law legal systems, documentary evidence must be authenticated to be admissible at trial. Authenticity, established through processes of authentication, is codified in our legal systems through statute and common law. Authentication of documentary evidence is accomplished through witness testimony, expert analysis, non-expert opinion, or, in the case of public documents or other special types, circumstances of record creation and preservation.[2]

These heuristics, based heavily on the appearance of documents, have developed over centuries and are still operational today, often misguidedly applied to digital documents. In interviews conducted in 2011 with lawyers, digital forensics experts, and records managers during the Digital Records Forensics Project (a three-year collaboration from April 2008 to April 2011 between the University of British Columbia's School of Library, Archival and Information Studies, the UBC Faculty of Law, and the Computer Forensics Division of the Vancouver Police Department), one respondent from the legal domain answered the question about determining the authenticity of documents: "You can tell just by looking at it" (Rogers 2011).

Authenticity is also contextual: "The meanings of 'authenticity' are relative to the concept of authentic that is held by different disciplines" (Lauriault et al. 2007, 140). This idea has been explored in recent literature (MacNeil and Mak 2007; Duncan 2009; Mak 2012). At the root of these explorations is the concept of authenticity as a social construction dependent on the context or discipline within which authenticity is defined, interpreted, and required. If we subscribe to the view that digital resources are "in a continuous state of becoming" as they are created, used, migrated, preserved, and accessed over time, then so too is the nature of their authenticity (MacNeil and Mak 2007,

26). In both cases, the questions remain about how to define the necessary elements of authenticity within a given context and how to assess them.

Much research has been conducted by records professionals on the nature of digital records and their attributes that may support the presumption of their authenticity. Still, current means of assessing authenticity do not offer any quantifiable measures. There is a pressing need for measures such as our financial, governmental, health, critical infrastructure, and social network systems to increasingly rely on complex integrated, interdependent (although not necessarily interoperable), distributed networked systems.

Digital technology's many benefits and challenges in respect of documentary material are well known. The benefits, including the ease of creation, search, access, and sharing, are offset by the ease of alteration, loss of integrity that may be difficult or impossible to detect, difficulty in establishing ownership and authorship, and difficulty in enforcing intellectual rights. The advent of cloud computing has increased the challenges, introducing, in particular, all of the issues arising from third party handling of material and jurisdictional questions about material created, stored, and transmitted around the globe, to name but two examples.

Records, defined as documents created or received in the course of practical activity and set aside for further action or reference, are the raw material of archival research and scholarship (Duranti 1993, 9; Eastwood 1994, 125; Duranti and Michetti 2015). In the digital environment, research agendas in information management communities focus on authenticity as an integral value that must be protected over time and across technological change through digital preservation (joining values of sustainability, accessibility, and understandability), broadening the scope of enquiry beyond records as narrowly defined by archival theory to documents, data, and digital objects of all types.

Archival science and the science of diplomatics have supported archivists in their understanding of the authenticity of traditional records. Authentic records are those whose identity can be established and whose integrity can demonstrated through an unbroken chain of custody over time. Diplomatics posits that all records can be analysed, understood, and evaluated in terms of a system of formal elements that are universal in application and decontextualized in nature (Duranti 1998). The InterPARES Project (International Research into Permanent Authentic Records in Electronic Systems, Phase 1 and 2), adopted the theoretical framework of diplomatics for the study of digital records and successfully developed the chain of preservation: "A system of controls that extends over the entire lifecycle of records and ensures their identity and integrity in any action that affects the way the records are represented in storage, or presented for use" (InterPARES Glossary, n.d.). The concept of a chain of preservation extends the controls implicit in the idea of chain of custody to address the susceptibility of digital records to corruption or loss. Requirements for establishing the authenticity of digital records are articulated in the benchmark requirements supporting the presumption of authenticity of electronic records and the baseline requirements supporting the production of authentic copies of electronic

records (Duranti 2005b) and the Creator and Preserver Guidelines (Duranti and Preston 2008).

Digital diplomatics is ideally suited to the analysis of authenticity of digital records as defined by archival science, but it is limited when the subject of analysis is broadened to include digital objects that may not satisfy that precise, but narrow, definition (MacNeil and Gilliland-Swetland 2005, 52; Duranti and Endicott-Popovsky 2010, 2). Archivists are now creating research alliances with digital forensic practitioners to develop and extend the applicability of digital diplomatics in the field of digital preservation and the focus on authenticity, reliability, and accuracy (Duranti 2009; Kirschenbaum, Ovenden, and Redwine 2010; John 2012).

In the traditional environment, a record is presumed authentic if it is relied on by its creator for the conduct of business and maintained in an unbroken chain of custody by the creator or his legitimate successor(s) (Duranti 1997, 214; Eastwood 1994, 127). Archival documents, deemed authentic by virtue of the circumstances of their creation and maintenance as part of the aggregate of records unified by the archival bond, are also therefore presumed reliable and their contents accurate.

In the digital environment no such automatic presumption of authenticity should exist. Digital technology has upset the traditional systems of control that have ensured the creation of authentic records and the means of presuming their continued authenticity over time and across technological change (MacNeil and Gilliland-Swetland 2005, 21; Lauriault et al. 2007, 140). Digital records differ significantly from paper records. Records, documents, and data created and stored in computer media are volatile and subject to loss, intentional or unintentional alteration, contamination, or corruption, even when they are still in the custody of their creator. Their authorship, provenance, or chain of custody may be difficult or impossible to determine. They may be transmitted, shared, and copied with ease. Their accessibility is subject to hardware and software obsolescence and incompatibility. Even if the creator relies on a digital record in the course of business and maintains its unbroken chain of custody, the fragility and vulnerability of digital records demands explicit action to protect the record's authenticity. Furthermore, reliability and accuracy are no longer directly linked with authenticity and may be compromised together or separately (Duranti and MacNeil 1997, 48; Duranti 2005a, 1; MacNeil and Gilliland-Swetland 2005, 21; Duranti and Thibodeau 2006, 54).

Survey: indicators of authenticity

As part of a broader research project exploring concepts and practices of authenticity of digital records and data, the author conducted a web-based survey from 3 March to 1 May 2014. The purpose of the survey was to gather basic information about how records, information, or systems professionals ensure, assess, and/or protect digital records' authenticity, what metadata they employ or rely on, and what indicators of authenticity they consider to be important. The

survey was posted on the major English-speaking archival and records management listservs.

It is my contention that despite large-scale, significant, and influential research into the topic of authenticity, primarily in the context of long-term preservation, the theoretical results of these projects are not being consistently applied in the practice of records professionals. This hypothesis is explored in the survey. The broad research questions motivating the survey interrogate notions of authenticity of digital records and data and investigate how records professionals interpret, ensure, and assess authenticity.

The survey consisted of seventeen questions, designed to explore work practice and belief about record authenticity. Demographic questions asked respondents to identify their job or position and the sector in which they work, their age, level of education, and discipline of their degree(s). Subsequent questions explored their main professional responsibilities, the means they used to ensure authenticity, what metadata they routinely applied or relied on for that purpose, whether they had ever been called upon to make a formal attestation of authenticity in a legal or administrative proceeding and, if so, what indicators had been most important in that attestation, and whether their organization explicitly defined authenticity in its policy instruments. The survey sought to explore the relationship between practice and belief—that is, what records professionals relied on in their work and whether that matched their belief or trust in authenticity indicators, identified from the perspective of archival science. It also sought to distinguish between what I have termed "social" and "technological" tasks and indicators of authenticity.

Social tasks are those conducted on the records or digital objects as conceptual objects, while social indicators of authenticity are instruments developed by an organization to support the creation, management, or preservation of records (for example, classification schemes, retention and disposition plans, and policies and procedures documents). They are based on domain knowledge and created and implemented by the intention of human actors (records professionals, management, legal counsel, and so on). They may or may not be present within a given organization; they may be mandatory or voluntary in their application or use, and, even when mandatory, they may be circumvented or adapted in practice. They include the foundational instruments of archival and records management practice: policy instruments, classification schemes or file plans, retention and disposition schedules, and archival description or other descriptive measures (which may be captured in varieties of descriptive metadata). Technological tasks treat the records as logical objects and involve technical aspects of preservation or curation, monitoring or enforcing security, or designing records systems. Technical indicators are non-discretionary in their creation—that is, they are the result of a work process or state change in the records (for example, system metadata capturing date created and date modified), are algorithmically generated or implemented by the technological components (for example, computer, network) of the overall record system (for instance, checksums, audit logs), are created to manage and control system access and security, or are created by a

third party as specifications to a part of the technological system (for example, documentation about software). Technological indicators may be used to control the records but are more focused on controlling the system in which the records reside. They include audit logs, access controls and security measures, cryptographic validation techniques, and system metadata as well as technical documentation. These distinctions were explored in a series of ranking, Likert-style questions. They were further supported by open-ended opinion questions asking respondents to give their own definition of authenticity and identify the indicators they felt were most important.

Preliminary findings

The survey received 441 responses. Of these, 148 did not answer any questions beyond those gathering demographic information and were discarded. Of the remaining 293 responses, participants self-identified primarily as archivists (45 percent) and records or information managers (33 percent). The remaining 22 percent were split between information professionals (librarians, administrators, 10 percent), educators (6 percent), and other (6 percent). Industry sectors most represented were information and cultural industries (including libraries and archives, broadcast, and telecommunications) and government (see Figure 1). Industry sectors were condensed from the North American Industry Classification System (Statistics Canada and Standards Division 2012).

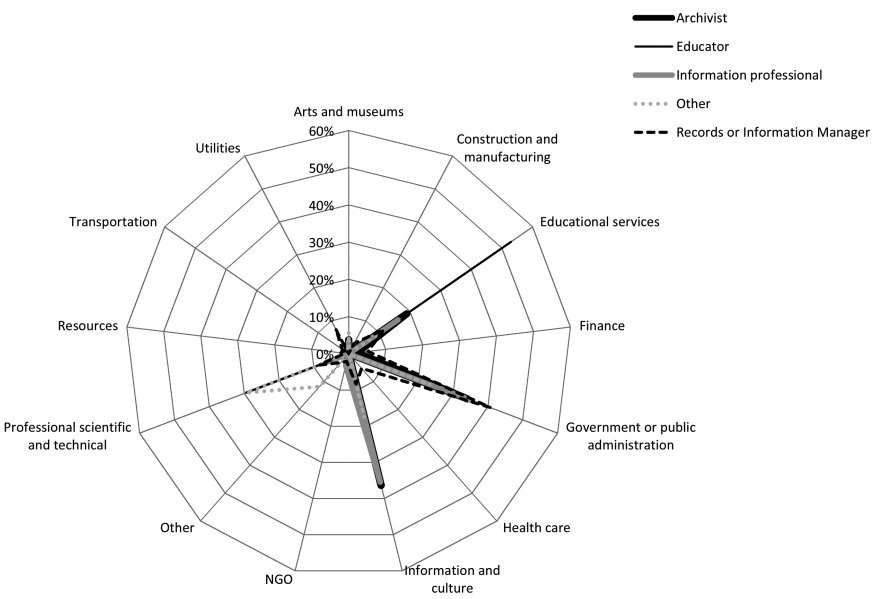

Figure 1: Respondents by profession and industry sector (n = 293)

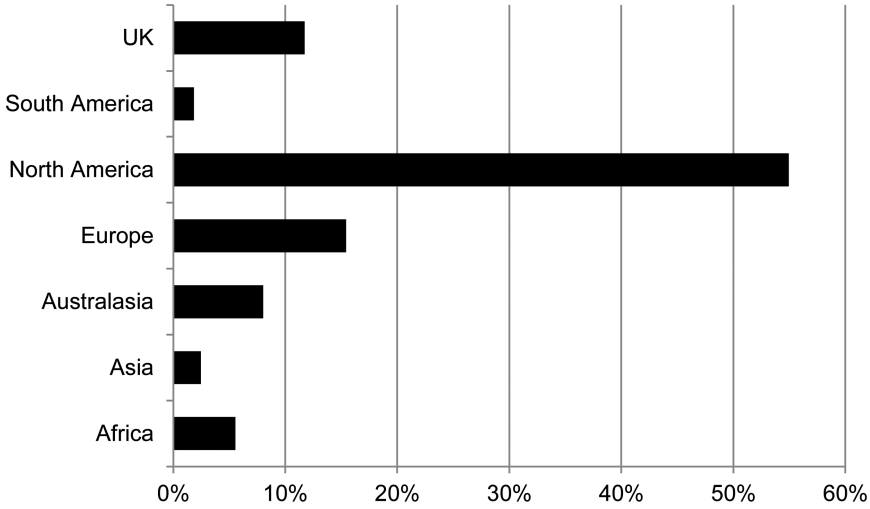

Figure 2: Percentage of respondents by continent

The survey achieved limited global reach, with the majority of respondents coming from North America (55 percent), followed by Europe (15 percent), and the United Kingdom (12 percent) (see Figure 2).

Respondents were asked a series of questions to determine their job responsibilities, how they ensured authenticity of records, and if they had even been required to attest to responsibility. These questions can be found in Appendix I.

Seventy-seven percent of respondents said they managed records or information frequently or very frequently, and 67 percent said they conducted retrieval and access, managing or designing metadata (56 percent), and designing information or records policies (51 percent). Monitoring security or access privileges and designing records management systems were the least common activities: 35 percent and 37 percent of respondents respectively reported that they never or infrequently performed these tasks.

When asked to rate social or technological indicators of authenticity according to how frequently they were used to ensure authenticity, more than 50 percent of respondents reported that they relied on traditional (social) archival and records management tools "most of the time" or "always" for managing authenticity, specifically policies governing records (55 percent) and record systems (60 percent), documentation about records systems (51 percent), classification schemes (61 percent) and retention and disposition schedules (51 percent). Fifty-three percent of respondents used access controls and security measures, and 54 percent employed standardized metadata "most of the time" or "always." However, 51 percent of respondents never or rarely relied on audit logs in the course of their work, and 61 percent never or rarely used cryptographic validation techniques.

Of specific cryptographic techniques employed, digital signatures were the least relied upon. Only 11 percent of respondents used digital signature technology

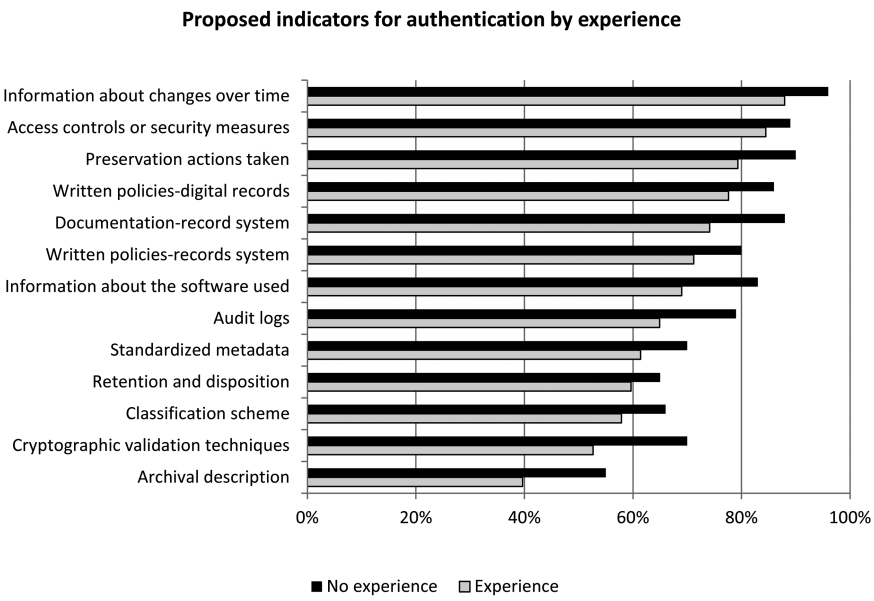

Figure 3: Comparison of indicators of authenticity (n = 247)

to ensure authenticity, and 61 percent never used this technology. Only 5 percent of respondents had been required to attest to authenticity of records in a court proceeding, and 10 percent had been involved in e-discovery or legal hold. Seventy percent of respondents had never been required to attest to the authenticity of records in their care for any purpose, including reference or access.

Respondents who had been required to attest to the authenticity of records or data were asked what indicators they had used in that attestation, while those who had never been required to attest to authenticity were asked what they believed would be necessary in the event that they were so called. The indicators most frequently used by respondents who had been required to attest to authenticity were policies about the digital records in question, information about access controls, and information about changes to the material. Respondents who had not been required to attest to authenticity inflated their belief in the value of all indicators. However, the greatest difference in choice of indicator was found in information about the preservation activities, use of standardized metadata, documented changes to information, information about software, and use of cryptographic techniques (see Figure 3).

When asked to rate their belief in the importance of indicators of authenticity if required to make an attestation of authenticity in a legal or administrative action, 68 percent of respondents said that standardized metadata would be very important or extremely important, cryptographic validation techniques were deemed very or extremely important by 66 percent, audit logs were favoured by 76 percent, and access controls and security measures were considered very

or extremely important by 88 percent of respondents. Despite their current practice, therefore, technical means of validating authenticity were considered as important as traditional means. With respect to organizational records and information policies, 54 percent of respondents said their organization did not define authenticity of digital material, and 17 percent did not know if their organizational policies contained such definitions.

Preliminary analysis of the narrative responses to the final two questions (what is your definition of authenticity of digital records and what do you believe is essential to proving the authenticity of digital records?) reveal that authenticity is still generally assessed according to traditional social heuristics. In response to the first question, several respondents noted that records produced in the usual and ordinary course of business could be presumed authentic, thus reflecting statute and precedent law governing business records in common law traditions. Most respondents noted integrity as a means of establishing authenticity, and several stated that bitwise integrity was necessary after the moment a record was "fixed"—that is, chosen to be kept as evidence of the action represented in the record or preserved for long-term reference in an archives. Responses generally indicated a pragmatic approach to authenticity, for example, one respondent answered:

> Is [the record] sufficient for the purposes it may be used for? Would it satisfy a judge or adjudicator? Whatever I can claim about it, can I back that up with facts? / The basic definition of an authentic record is "Can it be used as an authentic record in a situation where an authentic record would be needed?" This is not a yes/no answer (though the question is), but rather a range. I want the records as authentic as they need to be for future uses. They needn't be the MOST authentic—just authentic enough.

The final question explored respondents' beliefs about essential indicators of authenticity. Answers focused on chain of custody, controls on creation and management, policies on access and ensuring provenance information, and the addition or presence of metadata about the creator and context of creation. Several respondents noted the importance of cryptographic validation techniques, and several specifically stated that security and access controls were paramount (although one respondent noted the importance of these controls in the context of using public cloud-based email and document sharing).

This early exploration of the survey data points the way to further research to explore in greater depth the importance of social versus technical indicators of authenticity and how these are used when authenticity is questioned in legal or administrative hearings. Next steps will include further coding and analysis, particularly of the open-ended survey questions, followed by semi-structured interviews with many of those who indicated their willingness to provide more information. The applicability and authority of indicators of authenticity of digital records and data will be assessed in different environments, in particular, when records and data are created, maintained, or preserved in cloud-computing applications. This will be of increasing importance as more organizations turn to cloud service providers to support their operations and as courts continue to face the increasing challenge of evidence presented in digital form.

Limitations

Web-based surveys are convenient ways in which to reach a broad population quickly, but they do have limitations. Primary among these limitations is the inability to be assured of a representative sample. Even when using professional listservs (where the sample members can be reasonably assured of common purpose, training, and responsibility), respondents choose to reply, and while all respondents may be members of the target population, not all members of that population are members of, or have access to or read, these listservs. Generalizability of the results, therefore, is limited, and validity cannot be objectively measured. However, as an indicator of general practice, such surveys provide useful information.

Conclusions

Preliminary findings indicate that records professionals still tend to rely on traditional heuristics for ensuring authenticity, even when they claim to put their trust in more technical solutions if required to attest to authenticity. Records and information professionals—archivists and records managers—have traditionally been the trusted professionals who keep records safe, authentic, and reliable. As complex technology increasingly mediates between the record and the record user, records professionals necessarily place their trust in information technology professionals. It appears that the trusted records professional is now becoming the trusting technology user—the trustee has become the trustor. However, each discipline has unique and complementary knowledge. The records professional knows what information in the form of records and data has value and must be preserved and the information technology professional understands how to protect and secure that information. If our documentary heritage is at the root of democracy and accountability, both professions are necessary in its authentic preservation.

Notes

1. International Data Corporation, http://www.idc.com, is a prominent global provider of intelligence for the information technology, telecommunications and consumer technology markets.
2. *Federal Rules of Evidence*. http://www.law.cornell.edu/rules/fre.

References

ARMA International. 2014. *ARMA Generally Accepted Recordkeeping Principles*. http://www.arma.org/r2/generally-accepted-br-recordkeeping-principles.

Asssociation of Canadian Archivists. 1999. *Code of Ethics: The Association of Canadian Archivists*. http://archivists.ca/content/code-ethics.

Black's Law Dictionary. n.d. 2nd ed. http://thelawdictionary.org/.

Duncan, C. 2009. "Authenticity or Bust." *Archivaria* 68: 97–118.

Duranti, L. 1993. "The Archival Body of Knowledge: Archival Theory, Method, and Practice and Graduate and Continuing Education." *Journal of Education for Library and Information Science* 34 (1): 8–24. http://dx.doi.org/10.2307/40323707.

——. 1997. "The Archival Bond." *Archives and Museum Informatics* 11 (3/4): 213–18. http://dx.doi.org/10.1023/A:1009025127463.
——. 1998. *Diplomatics: New Uses for an Old Science*. Lanham, MD: Scarecrow Press.
——. 2005a. "The Long-Term Preservation of Accurate and Authentic Digital Data: The InterPARES Project." *Data Science Journal* 4: 106–18. http://dx.doi.org/10.2481/dsj.4.106.
——. 2005b. *The Long-Term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*. San Miniato: Archilab.
——. 2009. "From Digital Diplomatics to Digital Records Forensics." *Archivaria* 68 (2): 39–66.
Duranti, L., and B. Endicott-Popovsky. 2010. "Digital Records Forensics: A New Science and Academic Program for Forensic Readiness." *Journal of Digital Forensics, Security and Law* 5 (2): 1–12.
Duranti, L., and G. Michetti. 2015. "The Archival Method: Rediscovering a Research Tradition." In *Research in the Archival Multiverse*, ed. Anne Gilliland, Sue McKemmish, and Andrew Lau. Melbourne, Australia: Monash Publishing.
Duranti, L., and H. MacNeil. 1997. "The Preservation of the Integrity of Electronic Records: An Overview of the UBC-MAS Research Project." *Archivaria* 42 (1): 46–67.
Duranti, L., and R. Preston. 2008. *Research on Permanent Authentic Records in Electronic Systems (InterPARES) 2: Experiential. Interactive and Dynamic Records*. Padova: Associazione Nazionale Archivistica Italiana.
Duranti, L., and K. Thibodeau. 2006. "The Concept of Record in Interactive, Experiential and Dynamic Environments: The View of InterPARES." *Archival Science* 6 (1): 13–68. http://dx.doi.org/10.1007/s10502-006-9021-7.
Eastwood, T. 1994. "What Is Archival Theory and Why Is It Important?" *Archivaria* 37 (1): 122–30.
EMC Corporation. 2013. *Leaders Edge: Highlights from CIO Summit 2013*. Atlanta, GA: EMC Corporation. http://www.emc.com/microsites/cio/articles/cio-summit-2013/cio-summit-2013-atlanta.pdf.
Endicott-Popovsky, B., D.A. Frincke, and C. Taylor. 2007. "A Theoretical Framework for Organizational Network Forensic Readiness." *Journal of Computers* 2 (3): 1–11. http://dx.doi.org/10.4304/jcp.2.3.1-11.
International Organization for Standardization. 2001. *ISO-15489 (2001) Information and Documentation-Records Management*, No. ISO-15489 (2001). http://www.iso.org/iso/catalogue_detail?csnumber=31908
InterPARES Glossary. n.d. *InterPARES 2 Project: Terminology Database*. http://interpares.org/ip2/ip2_terminology_db.cfm.
John, J.L. 2012. *Digital Forensics and Preservation*: Digital Preservation Coalition. http://www.dpconline.org/component/docman/doc_download/810-dpctw12-03pdf.
Kirschenbaum, M.G., R. Ovenden, and G. Redwine. 2010. *Digital Forensics in Born Digital Cultural Heritage Collections*. Washington, DC: Council on Library and Information Resources.
Lauriault, T.P., B.L. Craig, D.R.F. Taylor, and P.L. Pulsifer. 2007. "Today's Data Are Part of Tomorrow's Research: Archival Issues in the Sciences." *Archivaria* 64 (2): 123–80.
MacNeil, H., and A. Gilliland-Swetland. 2005. "Authenticity Task Force Report." In *The Long-term Preservation of Authentic Electronic Records*, ed. L. Duranti, 19–65. San Miniato, Italy: Archilab.

MacNeil, H., and B. Mak. 2007. ''Constructions of Authenticity.'' *Library Trends* 56 (1): 26–52. http://dx.doi.org/10.1353/lib.2007.0054.
Mak, B. 2012. ''On the Uses of Authenticity.'' *Archivaria* 73: 1–17.
Pearce-Moses, R. 2005. *A Glossary of Archival and Records Terminology*. http://www.archivists.org/glossary/.
Richards, L. 2014. *Conversation with the Author*. Washington, DC, 13 August.
Rogers, Corinne. 2011. ''Trust Me! I'm a Digital Record: Findings from the Digital Records Forensics Project.'' Presented at the Archives 360, Society of American Archivists, Chicago, IL, 27 August.
Society of American Archivists. 2011. *SAA Core Values Statement and Code of Ethics*. http://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics.
Statistics Canada, Standards Division. 2012. *North American Industry Classification System (NAICS) Canada*. Ottawa, ON: Statistics Canada. http://www.census.gov/eos/www/naics/.
Turner, V., J. Gantz, D. Reinsel, and S. Minton. 2014. *The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things*, White Paper no. IDC 1672. Toronto: IDC Corporation. http://idcdocserv.com/1678.

Appendix A. Survey

How often do you conduct the following tasks with respect to digital records (e.g. electronic documents, images, data, data sets, database records, electronically stored information [ESI], web pages, etc.)? If self-employed or retired, please refer to the job or contract you feel is most relevant. (Never; Rarely; Sometimes; Often; Very Often)

- Conduct retrieval and access
- Monitor or enforce security/access privileges
- Monitor or enforce privacy of personal information
- Monitor or enforce compliance with record keeping regulations/policies (including e-discovery)
- Conduct preservation or curation
- Design systems for storage and management of records
- Design information/records policies
- Manage records or information
- Manage/design metadata
- Other

When you create or manage digital records, how often do you rely on or apply the following to ensure their authenticity? (Never; Rarely; Sometimes; Most of the time; Always)

- Written policies and procedures governing the management of the records system
- Documentation about the record system (design, operation, management, etc.)
- Written policies and procedures governing digital records

- Information about the software used to create and manage the digital records
- Information about changes made to the digital records over time (e.g. migration, normalization, etc.)
- Information about actions taken to preserve the digital records
- Classification scheme and/or file plan
- Retention and disposition schedules
- Archival description
- Access controls/security measures
- Audit logs
- Cryptographic validation techniques (e.g. digital signatures, hash digests, etc.)
- Standardized metadata

How frequently do you use the following cryptographic validation techniques?

- Digital signatures
- Trusted time stamps
- Checksums
- Hash digests
- Secure transmission

What metadata do you routinely use or manage? Please check all that apply.

- A metadata schema or guideline (e.g. Dublin Core, PREMIS, MoReq, etc.) Please list: _____
- A modification of a schema, customized for your organization. Please describe: _____
- A custom-built metadata schema (designed without elements from existing schemas). Please describe: _____
- Metadata generated by the software or record system in use only
- Not sure

Have you ever been required to guarantee or attest to the authenticity of digital records in any of the following circumstances?

- Providing testimony in court or administrative hearing
- Pending litigation or administrative action (e-discovery process)
- Authenticating copies of digital records for research or in response to reference requests
- Other _____
- I have never been required to guarantee or attest to the authenticity of digital records

When you were required to guarantee or attest that digital records are authentic, how important were the following in making your assessment? (Not at all important; Very unimportant; Neither important nor unimportant; Very important; Extremely important)

- Written policies and procedures governing the management of the records system

- Documentation about the record system (design, operation, management, etc.)
- Written policies and procedures governing digital records
- Information about the software used to create and manage the digital records
- Information about changes made to the digital records over time, (e.g. migration, normalization, etc.)
- Information about actions taken to preserve the digital records
- Classification scheme and/or file plan
- Retention and disposition schedules
- Archival description
- Access controls/security measures
- Audit logs
- Cryptographic validation techniques (e.g. digital signatures, hash digests, etc.)
- Standardized metadata

If you needed to assess that digital records are authentic, how important would the following be in making your assessment? (Not at all important; Very unimportant; Neither important nor unimportant; Very important; Extremely important)

- Written policies and procedures governing the management of the records system
- Documentation about the record system (design, operation, management, etc.)
- Written policies and procedures governing digital records
- Information about the software used to create and manage the digital records
- Information about changes made to the digital records, (e.g. migration, normalization, etc.)
- Information about actions taken to preserve the digital records
- Classification scheme and/or file plan
- Retention and disposition schedules
- Archival description
- Access controls/security measures
- Audit logs
- Cryptographic validation techniques (e.g. digital signatures, hash digests, etc.)
- Standardized metadata

Based on a consideration of storage only, how much confidence would you have in the authenticity of records in the following storage options, all else being equal? (No confidence; Little confidence; Neither confidence nor lack of confidence; Considerable confidence; Total confidence)

- Digital records stored by their creator on removable media (i.e., a USB key/external hard drive, optical or magnetic media)
- Digital records stored by their creator on stand-alone computers
- Digital records stored by their creator in network drives/filing system
- Digital records in cloud storage maintained by a third party cloud service provider
- Digital records stored by an archives

- Traditional (e.g. paper, microfilm) records stored on- or off-site by their creator
- Traditional records stored by a third party that is not an archives
- Traditional records stored by an archives

Do your organization's records policies define authenticity of digital records?

- Yes
- No
- Don't know

What is your definition of authenticity of digital records?

What do you believe is essential to proving the authenticity of digital record?

What About Trust in the Cloud? Archivists' Views on Trust

La question de la confiance dans le nuage : Le point de vue des archivistes sur la question

Erik A.M. Borglund
Mid Sweden University
erik.borglund@miun.se

Abstract: More and more information is "going to the cloud," including records and archives. This article focuses on understanding trust-in-cloud solutions from an archivist's perspective, exploring whether cloud computing has changed the archivist's role and how archivists respond to cloud-related problems and challenges. Twelve archivists in Sweden were interviewed in Swedish. They describe changes in their role due to cloud computing and services in the domain of archival science. Their role has changed from being reactive to becoming proactive, guarding not only the organization's needs and assets but also its archival records. Working proactively implies guaranteeing that requirements are updated and that contracts and agreements between the organization and cloud service provider are correct. The research shows that trust consists of several dimensions and cannot be easily achieved with technical solutions. Organizations' risk-tolerance levels have also changed to take advantage of the benefits and savings that cloud services provide for organizations.

Keywords: archivists, cloud computing, records, trust

Résumé : Des quantités de plus en plus importantes d'information vont « dans le nuage », y compris des dossiers d'archives. Cet article se propose de comprendre le point de vue des archivistes concernant la confiance qui peut être accordée aux solutions informatiques en nuage, d'examiner si l'informatique en nuage a changé le rôle des archivistes et comment les archivistes réagissent aux problèmes et aux défis liés aux nuages informatiques. Douze archivistes en Suède ont été interrogés en suédois. Ils décrivent les changements dans leur rôle dûs à l'informatique en nuage et dans les services propres au domaine de l'archivistique. Leur rôle est passé de réactif à proactif, se faisant les gardiens des besoins et des actifs de leur organisation, et non seulement de ses documents d'archives. Travailler de manière proactive implique de garantir que les exigences sont mises à jour, et que les contrats et les accords entre l'organisation et le fournisseur de service informatique en nuage sont corrects. La recherche montre que la confiance se compose de plusieurs dimensions, et qu'elle ne peut pas être facilement réalisée avec des solutions techniques. Les niveaux de tolérance au risque des organisations ont également changé, afin de tirer profit des avantages et des économies que les services d'informatique en nuage apportent aux organisations.

Mots-clés : archivistes, informatique en nuage, documents, confiance

La Revue canadienne des sciences de l'information et de bibliothéconomie 39, no. 2 2015

Introduction

In the last couple of years, more and more information is "going to the cloud," including records and archives, yet very little research has been undertaken to assess the impact of cloud computing from an archival science perspective (Ferguson-Boucher and Convery 2011). "The cloud" is the short term for cloud computing, a metaphor for various services available through a network, which, in most cases, is the Internet. In cloud computing, a range of different computing resources may be accessed, and one way to present and understand the cloud is to use the model defined by the US National Institute of Standards and Technology (NIST) (National Institute of Standards and Technology 2009; Mell and Grance 2011). Service models are central to the NIST model: software as a service (SaaS); platform as a service (PaaS); and infrastructure as a service (IaaS).

As a result of its low cost, organizations are increasingly moving their records into the cloud and delegating to cloud providers the responsibility for their security, accessibility, disposition, and preservation. However, how high is the price that these organizations pay in terms of having control over their records or, as is the case with archives, of the records entrusted to them for permanent preservation? We have seen cloud providers go bankrupt, disappear, or be sold; records lost, retained when they should have been destroyed, or mixed up in shared servers; failed back-ups; and unauthorized access by sub-contractors and hackers. Further, it is impossible to pinpoint the geographical location of the records at any given time as well as the jurisdiction under which they fall; to prove the chain of custody and the authenticity of the records; to ensure protection of legal privilege or trade secrets when using a third party; to isolate documents for legal hold; to conduct audits; and to guarantee that the records that need to be permanently preserved are kept according to archival standards. These are only a few of the problems encountered by organizations using the cloud as if it were a recordkeeping or a record-preservation system. Yet the number of those who choose to use the cloud for these purposes is growing exponentially by the day. If this phenomenon cannot be stopped, we must at least try to reduce its risks to an acceptable level.

In the existing literature about archives and the cloud, several different focuses and trends can be identified. There are examples of archives being presented as cloud solutions and of the archive being presented as a service (Askhoj, Nagamori, and Sugimoto 2011; Askhoj, Sugimoto, and Nagamori 2011). However, a literature search in scientific bibliographic databases and outlets found nothing about electronic archives being developed using PaaS or IaaS, which are similar to more traditional outsourcing. Another identifiable trend is the management of records. Business based on modern web 2.0 encourages cloud usage, and it is almost impossible to talk about online work without talking about the cloud (Stuart and Bromage 2010), which implicitly also makes the cloud a topic of interest for the archival community. Katherine Stuart and David Bromage (2010) present a set of problems related to cloud computing and records management: (1) trust in records; (2) general problems related to the management of records; and (3) the fact that the storage location of the records is unknown.

This last problem of where the files are stored is not solely an archival problem (Benson, Dowsley, and Shacham 2011).

Societal changes provide another perspective on the relationship between archives and recordkeeping and computing. Since the early 1990s, changes in society due to the rapid development of information technology (IT) have been highly debated in archival science (see, for example, Cook 1997; Dollar 1992). Technical evolution does not stop, and the modern online culture and the adoption of available technologies requires new methods to be able to manage archives and records (Upward, McKemmish, and Reed 2011). Cloud computing is one of these new technologies affecting the archival domain. The Internet and new technologies have also established a more mobile work trend, which Sari Mäkinen (2013) takes as the departure point for her work. She argues that the mobile worker and the new ways of using mobile technology put archives and record management to the test. One can argue that modern mobile workers are also a driving force for cloud storage and cloud usage within the records management domain (Mäkinen and Henttonen 2011; Mäkinen 2013). Archival theory rests upon the idea of provenance, and, according to Mohamed Sakka, Bruno Defude, and Jorge Tellez (2010), provenance is even more challenging to achieve in the cloud compared to relational databases, for example. This survey of the field indicates that there are several challenges for archival science regarding cloud computing and its components. One of the most obvious problems is how to trust digital records (see, for example, Duranti and Rogers 2012) and those in the cloud are even more problematic.

This article will focus on how trust in various cloud solutions can be understood from the perspective of archivists, who have previously been seen as guardians of trustworthy records. However, when more records are stored in the cloud, archivists cannot be the same kind of "guardians" that they were with analogue and paper-based records—they have a different role. This article does not aim to focus on how to make digital records trustworthy; trust comprises more than a technical solution. Trust involves actors, and this article investigates the archivists as actors. The purpose of this article is therefore to explore whether cloud computing has changed the archivist's role and how modern archivists relate themselves and their work to problems and challenges that spring from the cloud.

Cloud service perspectives

In this article, the model defined by the NIST has been used, serving as a guide for characterizing the cloud: "This cloud model is composed of five essential characteristics, three service models, and four deployment models" (Mell and Grance 2011, 2). The essential characteristics include on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. The model presents three different services from the cloud: SaaS; PaaS; and IaaS. Finally, the four deployment models are private cloud, community cloud, public cloud, and hybrid cloud. The service models' and the deployment models' internal relationship with each other, together with the characteristics of the

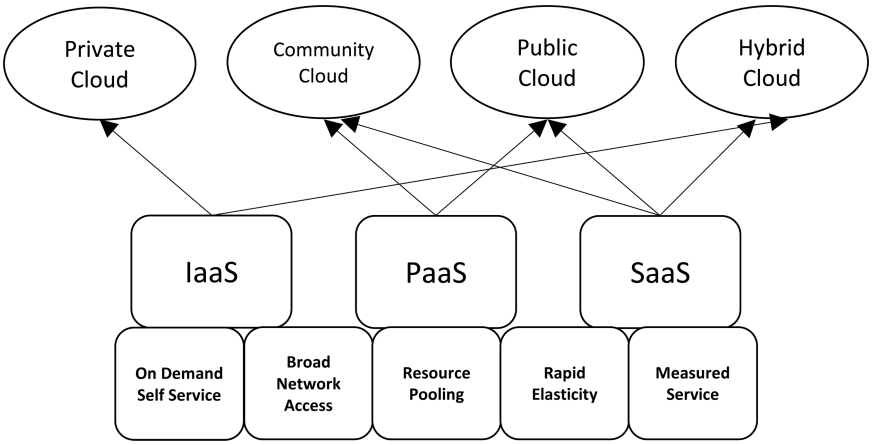

Figure 1: Cloud characteristics and the relationship between deployment models and service models (Vizcayno n.d.)

cloud (derived from the earlier definition) have been used. Their relationship is presented in Figure 1.

According to the NIST, five essential characteristics make the cloud what it is: (1) on-demand self-service, which allows users to access as many computing capabilities as they need; (2) broad network access, so that a user can access the cloud from any machine that has a connection to the Internet; (3) resource pooling, which makes the cloud into a multi-tenant model, supporting multiple users at the same time; (4) rapid elasticity, so that users can change the amount of computing resources they need at any time and the cloud will instantly expand to support their needs; and (5) measured service, so that how much a user utilizes is precisely measured in terms of storage, processing, bandwidth, and so on, and these resources can be monitored, controlled, and reported to the users, who are only charged for what they need, using a pay-as-you-go model, which reduces costs in most cases.

Perspectives of trust in archival science

Although this article does not focus on making digital records in the cloud trustworthy, it is relevant to understand the perspective of trust that has influenced this research. Historically, archives have been seen as the guardians of evidence, which requires trustworthy records (Duranti 1996). It is possible to interpret the archives as black boxes in which the records are always kept in a secure environment, guaranteeing the trustworthiness and evidential value of every single record. The trustworthiness of records is seldom questioned, even when the records are removed from their origin, as demonstrated in the cases of WikiLeaks and the whistleblower Edward Snowden.

The impact of Snowden's information leaks seems to be related to citizens' trust in government records, which is in turn rooted in a long tradition of

managing and keeping records and archives. No one actually questioned the correctness of the leaked information. This example illustrates citizens' trust in public records, despite the fact that these records were extremely vulnerable to manipulation outside their original context and custody. In the digital world, records are no longer necessarily in the custody of the archive. Only part of the record is captured as a record, and digital signatures or other technical measures used to guarantee authenticity often do not exist. This situation makes it more pressing to understand how trust is created in the digital world. It also helps to understand how trust can be lost. The current move toward open data, reuse of public sector information, and cloud storage increases the importance of understanding how trustworthy and reliable records can be ensured whenever and wherever they move in the new, networked environment.

Trust is a fundamental concept in archival science, and it is extremely important that records can be guaranteed to be trustworthy. However, researchers interested in all three dimensions of trust (individual, organizational, and temporal) are rarely found. The relationship between records, trust, and evidence has been an issue for discussion among archival scholars influenced by cultural, technological, legal, and philosophical trends. When viewing records as impartial evidence, the records derive their value from the manner in which they were created as "by-products of activity rather than conscious players in the activity itself" (Trace 2002, 139). However, trust must also be understood and seen in relation to what constitutes the record. Anneli Sundqvist (2011, 277) explains that records are both instruments of trust and objects to be trusted. The main difference between records that are digital and those that are paper/analogue is that the electronic records are only logical, not physical, objects. According to Sundqvist, trust is relational since it always involves someone who trusts, and trust, rather than being the result of rational calculations, works as "a substitute for explicit knowledge" (279). Time itself is a challenge since records must be able to be trusted even after their original context is gone, which is why formalities have been developed: for example, date, signature (internal), requirements of custody (external), and so on (284). In the digital environment, trustworthiness is often achieved by technical solutions such as digital signatures (Duranti and Rogers 2012), but there are also various forensic technologies that can demonstrate whether records are trustworthy.

In a digital environment such as the Internet, it is natural that trust in information has become more and more relevant (Kelton, Fleischmann, and Wallace 2008), particularly since the use of Internet technologies is now fully embedded in modern society. Kari Kelton, Kenneth Fleischmann, and William Wallace (2008) list one simple way of categorizing trust in four levels: individual, interpersonal, relational, and societal. Morten Hertzum et al. (2002) show that trust in information is an intertwined mix of people, documents, and virtual agents affecting trust. In more management-focused literature, trust is seen as being between individuals and is often presented as an outcome of a process in which actors come to trust each other (Blomqvist 1997). It is obvious that trust is a very multifaceted concept, and a large sample of research measuring trust

between organizations was completed by Risto Seppänen, Kirsimarja Blomqvist, and Sanna Sundqvist (2007).

Research method

The research was based upon a qualitative approach (Taylor and Bogdan 1998; Myers 2009), using twelve in-depth interviews with archivists. Two criteria were used for selecting interviewees: (1) their interest in being interviewed and (2) their experience with modern archival management, which had to include digital records management. An invitation letter was sent out to available Swedish archival list servers, and an announcement was made through Twitter and Facebook, resulting in a total of fifteen replies. Of these replies, twelve were interviewed. The selection of the archivists can be characterized as adapted selection (Hartman 1998). The study was carried out in Sweden, and the interviews were done in Swedish. The interview questions were very open and focused on triggering thoughtful responses from the archivists that were interviewed. The results from the interviews were partly transcribed to support the inductive analysis process, which was used to identify patterns within the responses to each of the questions. The analysis was conducted with the help of qualitative data analysis software, Nvivo 10.

Only Swedish archivists were interviewed. Although the national bias of the research at the design phase of the study was seen as minimal, all context dependencies cannot be avoided—the cloud and usage of the cloud are not bounded by national borders. However, it should be borne in mind that this is qualitative research that does provide generalizable results, but it aims instead to present results that can be transferred to similar conditions and contexts.

Results

The results of the interviews are structured according to the general topics elicited by the interview questions. Each topic aims to cover the overall content of the responses given by the archivists during the interview.

Trust and the cloud in general terms

The interviews all started with a general question about trust and the cloud and how each archivist spontaneously thought about trust and the cloud. This was a question that triggered a lot of response from the interviewees, and a couple of common areas have been identified. "Can we trust the information we store?" was the reply from one of the archivists. Others gave similar answers, with problems related to trust being a common theme. Can we trust that the information stored in a cloud service is stored in a correct way; can we trust the service provider; and can we trust that the information is kept within the national borders? In this case, trust is a multifaceted term focusing on both the records and the service provider. Many archivists considered it important to set up a contract or agreement that increases trust. Focusing on information security and the management of information security is one strategy that implicitly increases trust.

Another more general problem that was identified is the fact that the cloud is intangible since physical components of the cloud service are not visible to the customer (that is, the servers are placed at a location hidden from the customer). This intangibility makes trust more difficult to define and discuss. In a physical archive, the door could be opened to inspect the archival records yourself, and if it was well kept it is easier to trust the archival provider. With the cloud, you do not fully know what to trust because there is limited competence and knowledge about what a cloud is and how it works. Some of the archivists who were interviewed also said that the cloud is probably more trustworthy than other alternatives just because the service providers are professionals, and the problem with trust is more related to the fact that the providers are not clearly identifiable.

Challenges

New technical solutions and new technical innovations adopted by organizations and archivists might result in different challenges for the archival profession to solve. Many of these are related to problems, which the interviewed archivists defined as 'challenges'. They are listed below because they are difficult to categorize.

- *Long-term perspective.* How long can cloud service providers guarantee the service? How can we guarantee that the information stored in a cloud service today can also be accessible and useful in the future?
- *Knowledge.* Archivists do not have enough competence and knowledge about cloud services and IT to be able to define the requirements for a cloud service. There is a risk that due to this low level of knowledge the organization's IT department will assume responsibility for the written contract between the cloud service provider and the organization.
- *Black box syndrome.* Even if the cloud service provider gives several customer business references, they still resemble a black box service. For a potential customer, it is very difficult to verify and check that the cloud service provider really can do what they claim. It is also very difficult to check that the cloud service provider has the relevant technical systems for long-term preservation.
- *Information security.* Managing information security is a challenge when information is managed in a cloud service and the "information owner" does not have physical access to the information.

Trust in the cloud provider

The introduction question presented the major problem of trust in the cloud provider. However, trust is not a universal term, so the ways in which trust in the cloud provider can be understood will be examined. After analysing the interviews, a set of very clear categories related to trust became visible. First, trust is related to something that can best be described as a relationship with the cloud provider—a relationship that exists between the user and the provider of the cloud service. If the organization has had some previous and historically successful business with the cloud service provider, it was claimed that trust will

increase. In other words, a common history is often positive. Trust is also expected to be easier to achieve if the cloud service provider is close to the client organization. In other words, the perception of trust can be related to physical distance between the customer (the organization) and the cloud service provider.

Second, trust as concept can also be divided into sub categories. "Can we trust the cloud service provider not to give away the stored information to someone else?" was one of the archivists' spontaneous replies to the question. Another question was: "can we really trust that the cloud service provider knows what they are doing, that they have the relevant competence about Swedish archival regulations?" These two different trust perspectives relate to the cloud service provider but have different foci. One-third of the interviewed archivists mentioned the whistle-blower Edward Snowden as one example of why it is problematic to trust a cloud service provider. You do not really know with whom they will share the information.

Trust and the record

When talking about cloud services, it is natural to include the aspects of trust related to the artefact, the information object that is kept in the cloud—that is, the records that are managed. The interviewed archivists all agreed that trust in relation to records in the cloud is very similar to the problems that exist with all digital records. The problem with trustworthy records is that in cloud services an external partner manages the records. The record is digital, and, therefore, the same kind of problems relating to trust can be identified in these records as in other digital records. It is challenging to guarantee that the record fulfils the quality criteria of authenticity, integrity, completeness, and usability. However, there is almost a paradox with trust because if you do not trust the record, then the whole business idea of the cloud is useless. Some of the archivists also said that the trustworthiness of records stored in the cloud might even be higher than with records stored and managed in-house by an IT department with doubtful competence.

Competence needs

Every interviewed archivist agreed that the phenomenon of the cloud was here to stay and that it would be very problematic to manage archival issues inside the cloud without specific and new knowledge and competences. Three major competence areas were identified during the interviews. These competences exist side by side and are not mutually exclusive; on the contrary, they are intertwined:

1. *IT knowledge.* Archivists should increase their IT knowledge. Some of the archivists proposed that the modern archivist should have a more IT-based toolbox—that is, they should have basic knowledge in information systems, system science, databases, and so on. Knowledge in information security was also identified as being relevant. Increased IT competence would allow archivists to communicate better with cloud service providers, including supporting communication between archivists and IT professionals.

2. *Requirement engineering.* Requirement engineering is seen as the competence to identify the user, organization, and IT requirements for the management of digital records in a cloud environment. All archivists agreed that they needed competence for specifying requirements. This competence requires knowledge about cloud services as technology.
3. *Agreement/contract design.* Competence to work with agreements and contracts is also necessary. If the organization aims to use cloud solutions, it is important to finalize an agreement between the organization and the cloud service provider. This cannot be left to lawyers or IT professionals. Archival requirements also need to be embedded in the agreement.

Archivists' new responsibilities

Given the new competences that are needed, a relevant follow-up question is whether archivists in modern organizations have acquired new responsibilities as a result of the increased use of cloud services. Previously, archivists were seen and presented as guardians of trustworthy records, but it is absolutely natural that this picture can and must be changed. One of the archivists said: "The archivist has become more of a guardian of the entire organization's archival interests than merely guardian of the records." This quotation is a comprehensive summary of the new responsibilities that became visible during the interviews. Proactivity is another word that explains the archivists' expanded responsibilities. Such proactivity becomes operational when the archivist needs to specify requirements, draft agreements, and contribute to writing the contract between the organization and the cloud service provider. He or she may also need to develop rules and regulations supporting the organization's work with various cloud services. The new proactive approach makes the archivist a generalist as well as an expert.

Will the role of archivists change?

The need for archivists to be proactive and broaden their responsibilities beyond that of a guardian changes their role. The archivist is now responsible for how information is managed and controlled—they are no longer merely guardians but, rather, more of a controller responsible for information management. When it comes to cloud services, this audit function becomes more important, ensuring that information in departments in an organization is managed and controlled according to the regulations and requirements. Some of the interviewed archivists saw an opportunity for the archivist to work together with the IT department to set up a new cloud audit service to guarantee that organizational information assets kept in the cloud are kept according to organizational requirements.

Even if the cloud changes the role of the archivist, all archivists interviewed in this study claimed that responsibility for providing records upon request will still be the archivist's responsibility. Another role that will not change is responsibility for appraisal and archival description, although its craftsmanship will change due to the cloud environment.

Reasons for using the cloud

The primary reason for using different cloud services relates to costs. Half of the archivists replied spontaneously that the primary motive for using cloud services was because of their low cost and the need to save money. However, after discussing this topic with them in more depth, a more nuanced picture became clear. The price of the service is still important, of course, but the cost is also related to the service level. When organizations set up their own services in-house or hosted on controlled servers, it is not completely clear what the cost will be in in relation to the service. However, by using cloud services, the majority of the interviewed archivists said that they knew what they got for a defined cost. Service/cost is clearer with cloud services.

Yet cost is not only related to cost for the service—that is, the storage and management of stored data. Competence is not cheap, and by using cloud services organizations can minimize the internal competence required. An IT department does not need to be an expert in setting up advanced storage solutions that will fulfil archival requirements. However, organizations that do not have this competence can also choose the cloud merely because they do not have to have the competence in-house. The technical evolution regarding advanced IT and data and records storage is growing quickly, and it can be impossible to guarantee that smaller organizations will have the right competence for managing their archives.

Citizens and external users were also cited as a reason for using the cloud. The archivists argued that there is a trend in public organizations to be more service oriented, and trends such as open data make public organizations more eager to test and use cloud-based services. Many citizens are used to accessing services they use themselves, such as DropBox, Box, iCloud, and GoogleDrive, from any device, and this is another reason why public authorities are going for cloud solutions. The citizens request easy access and this access motivates cloud usage, which makes it easier to access and reach data, information, or records that are stored in such services. Staff within public organizations also use cloud solutions privately and this use creates an organizational-bounded desire for such services. Easiness, smooth access, and flexibility are arguments that were presented during the interviews.

The last reason for using the cloud, which was presented by many of the archivists, was ideological. Many public authorities have decided that they should not host any IT in-house and that all such technology should be bought in as a service. There is a trend to streamline public work, contracting out as much as possible. IT, economy, human resource management, and archives are all examples of support processes that can be put in the cloud instead of being managed by the organization itself.

Risk taking

The interviewed archivists were asked whether they thought that organizations and individuals tend to become more willing to take risks concerning cloud services. They were asked to compare this idea both to other digital record

management and archival management technologies and also to purely analogue management. Eight of the archivists agreed that they tend to take more risks, as individuals and in their organizations, when it comes to the use of various cloud services. It is hard to clearly identify the underlying reasons, but some tendencies can be presented. First, risk taking is argued to result from low competence and juridical and IT knowledge among archivists as well as among decision makers primarily. Some of the archivists discussed risk taking as an effect of a more negligent use of information in modern society. Another aspect of risk taking is that those who have become used to cloud services privately have adopted more risk-tolerant behaviour since the easiness of cloud services has made them willing to take more risks.

Some of the archivists argued that cloud services could, on the contrary, be more secure than other alternatives because the cloud service providers are experts in what they do, while small organizations' IT departments may not have the necessary expertise. The long-term perspective was a common challenge for all archivists, who thought that none of their organizations really tried to understand the risks connected to the requirements of preserving records over the long term. The concept of the cloud is not easy to grasp, and, therefore, the risks that organizations and individuals are willing to take may be interpreted as more risky than they are in reality. However, in-house digital storage is also risky, and the risks related to physical archives are very seldom discussed. In the worst case, physical archives can be more risky than the cloud. But the archivists who were interviewed all claimed that risks with the cloud are also more fuzzy and difficult to understand.

Rules and regulations

Opinion was divided among the interviewees on whether the regulations and the current National Archives of Canada Act support the archivists in their work with the cloud.[1] Two clear opposing perspectives became visible. The first was that the rules and regulations are good enough, and those problems that exist depend entirely on how each organization applies the regulations. This perspective rests upon assumptions wherein regulations are seen in general terms. The second perspective is totally opposite, where the new phenomena of cloud service and digital records are seen as being so radically new and different that current rules and regulations are extremely out of date. Proponents of both perspectives made it clear that they considered it necessary to design practical guidelines based on the current regulations.

SaaS, PaaS, or IaaS

The last section of the interviews focused on trying to see which of the three service models might be most popular. The archivists all said that the service models do not, in reality, have borders that are as clear as the NIST (National Institute of Standards and Technology 2009) states. They said that their experience is rather that these service models are intertwined. Not one of the interviewed archivists worked at an organization that had used PaaS and IaaS on their own.

On the other hand, all had some experience in using SaaS, but this was often combined with IaaS. In the Swedish context, the terms Saas, PaaS, and IaaS have not been fully adopted by the archival community, and the interviewed archivists could neither say whether their IT departments had used these terms.

Concluding remarks

The purpose of the research presented in this article was to investigate whether cloud computing has changed the archivist's role and how modern archivists relate themselves and their work to problems and challenges that spring from the cloud. Based upon the interviews carried out in this research, new knowledge is presented. The archivists interviewed for this article described how their role has changed due to the effects of cloud computing and the introduction of cloud services in the domain of archival science. Their role has previously been more reactive—that is, to act when the information has already been created, which is impossible with digital records in general but even more problematic when it comes to cloud services. A proactive approach is proposed in which the archivist protects the organization rather than the archival records. The proactive archivist makes sure that requirements are updated and that the contract and agreement between the organization and the cloud service provider is correct.

Cloud services and cloud computing are different from other digital records management and archival management techniques. This distinction has had an impact upon trust as well. Trust in relation to cloud services is complicated and this research shows that trust consists of several dimensions, and, therefore, trust is not something that can be easily achieved with technical solutions alone. This research also indicates that there has been a change in organizations' willingness to take on risk and that the cloud services currently on offer provide easier and cheaper solutions for organizations.

The problems presented by Stuart and Bromage (2010) that relate to cloud computing and records management have been only partly confirmed by this research. The problems they outline include (1) trust in records, (2) general problems with the management of records, (3) the unknown location of the stored records. The first problem that Stuart and Bromage (2010) present about trust in records has not been fully confirmed by our research. As described previously, the problem of trust in the cloud service provider is still seen as a larger issue. The problem with the general management of records has also not been confirmed other than by several comments about general challenges in records management. The third problem, however, has been confirmed by our research.

The research presented in this article will be followed by a larger questionnaire-based study that will aim to reach a larger sample of archivists and further explore how it is possible to interpret issues concerning the cloud and trust. This future research will also aim to identify the many dimensions of trust in relation to cloud services. However, instead of seeing the cloud as a problem, this research will support the perspective that the cloud is actually a starting point for the creation and establishment of new, proactively driven archival practice that in turn supports the development of new, relevant, and up-to-date methods.

(V9 15/7/15 14:39) UTP (6"×9") AGaramond pp. 114–127 1659 CJILS 39.2_03_Borglund (p. 125)

Notes

1. National Archives of Canada Act, RSC 1985, c 1.

References

Askhoj, J., M. Nagamori, and S. Sugimoto. 2011. ''Archiving as a Service: A Model for the Provision of Shared Archiving Services Using Cloud Computing.'' Proceedings of the 2011 iConference, Seattle, WA. http://dx.doi.org/10.1145/1940761.1940782.

Askhoj, J., S. Sugimoto, and M. Nagamori. 2011. ''Preserving Records in the Cloud.'' *Records Management Journal* 21 (3): 175–87. http://dx.doi.org/10.1108/09565691111186858.

Benson, K., R. Dowsley, and H. Shacham. 2011. ''Do you know where your cloud files are?'' Proceedings of the Third ACM Workshop on Cloud Computing Security Workshop, Chicago, IL. http://dx.doi.org/10.1145/2046660.2046677.

Blomqvist, K. 1997. ''The Many Faces of Trust.'' *Scandinavian Journal of Management* 13 (3): 271–86. http://dx.doi.org/10.1016/S0956-5221(97)84644-1.

Cook, T. 1997. ''What Is Past Is Prologue: A History of Archival Ideas since 1898, and the Future Paradigm Shift.'' *Archivaria* 43: 17–63.

Dollar, C.M. 1992. *Archival Theory and Information Technologies: The Impact of Information Technologies in Archival Principles and Methods*. Macerata, Italy: University of Macerata.

Duranti, L. 1996. ''Archives as a Place.'' *Archives and Manuscripts* 24 (2): 242–56.

Duranti, L., and C. Rogers. 2012. ''Trust in Digital Records: An Increasingly Cloudy Legal Area.'' *Computer Law and Security Review* 28 (5): 522–31. http://dx.doi.org/10.1016/j.clsr.2012.07.009.

Ferguson-Boucher, K., and N. Convery. 2011. ''Storing Information in the Cloud: A Research Project.'' *Journal of the Society of Archivists* 32 (2): 221–39. http://dx.doi.org/10.1080/00379816.2011.619693.

Hartman, J. 1998. *Vetenskapligt tänkande: från kunskapsteori till metodteori*. Lund: Studentlitteratur.

Hertzum, M., H.H.K. Andersen, V. Andersen, and C.B. Hansen. 2002. ''Trust in Information Sources: Seeking Information from People, Documents, and Virtual Agents.'' *Interacting with Computers* 14 (5): 575–99. http://dx.doi.org/10.1016/S0953-5438(02)00023-1.

Kelton, K., K.R. Fleischmann, and W.A. Wallace. 2008. ''Trust in Digital Information.'' *Journal of the American Society for Information Science and Technology* 59 (3): 363–74. http://dx.doi.org/10.1002/asi.20722.

Mäkinen, S. 2013. '''Some records manager will take care of it': Records Management in the Context of Mobile Work.'' *Journal of Information Science* 39 (3): 384–96. http://dx.doi.org/10.1177/0165551512471934.

Mäkinen, S., and P. Henttonen. 2011. ''Motivations for Records Management in Mobile Work.'' *Records Management Journal* 21 (3): 188–204. http://dx.doi.org/10.1108/09565691111186867.

Mell, P., and T. Grance. 2011. *The NIST Definition of Cloud Computing*. Gaithersburg, MD: National Institute of Standards and Technology; http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.

Myers, M.D. 2009. *Qualitative Research in Business and Management*. London: SAGE.

National Institute of Standards and Technology (producer). 2009. *Definition of Cloud Computing*. http://www.nist.gov/itl/cloud/upload/cloud-def-v15.pdf.

Sakka, M., B. Defude, and J. Tellez. 2010. ''Document Provenance in the Cloud: Constraints and Challenges.'' In *Networked Services and Applications: Engineering, Control and Management*, vol. 6164. ed. F. Aagesen and S. Knapskog, 107–17. Berlin: Springer. http://dx.doi.org/10.1007/978-3-642-13971-0_11.

Seppänen, R., K. Blomqvist, and S. Sundqvist. 2007. ''Measuring Inter-Organizational Trust: A Critical Review of the Empirical Research in 1990–2003.'' *Industrial Marketing Management* 36 (2): 249–65. http://dx.doi.org/10.1016/j.indmarman.2005.09.003.

Stuart, K., and D. Bromage. 2010. ''Current State of Play: Records Management and the Cloud.'' *Records Management Journal* 20 (2): 217–25. http://dx.doi.org/10.1108/09565691011064340.

Sundqvist, A. 2011. ''Documentation Practices and Recordkeeping: A Matter of Trust or Distrust?'' *Archival Science* 11 (3-4): 277–91. http://dx.doi.org/10.1007/s10502-011-9160-3.

Taylor, S.J., and R. Bogdan. 1998. *Introduction to Qualitative Research Methods: A Guidebook and Resource*, 3rd edition. Chichester, NY: Wiley.

Trace, C.B. 2002. ''What is recorded is never simply 'what happened': Record keeping in modern organizational culture.'' *Archival Science* 2 (1-2): 137–59. http://dx.doi.org/10.1007/BF02435634.

Upward, F., S. McKemmish, and B. Reed. 2011. ''Archivists and Changing Social and Information Spaces: A Continuum Approach to Recordkeeping and Archiving in Online Cultures.'' *Archivaria* 72: 197–237.

Vizcayno, D.C. N.d. Danielito C. Vizcayno Blogs. http://dcvizcayno.wordpress.com/2012/04/13/cloud-computing-tips-for-financial-industry/.

Cloud Service Contracts: An Issue of Trust

Les contrats de service d'informatique en nuage : Une question de confiance

Jessica Bushey
School of Library, Archival and Information Studies, University of British Columbia
jbushey@mail.ubc.ca

Marie Demoulin
École de bibliothéconomie et des sciences de l'information, Université de Montréal
marie.demoulin@umontreal.ca

Robert McLelland
Delta Museum and Archives
robertmclelland@gmail.com

Abstract: This article compares cloud service contracts with records management and archival needs to determine whether or not those needs are met by currently available, boiler-plate contracts. It finds that, in general, the requirements of storing and preserving authentic records are not met by current cloud service agreements. It ends by proposing a checklist of requirements for recordkeeping professionals to utilize in negotiating or choosing contracts to better support the needs of authentic records in the cloud.

Keywords: cloud, authenticity, contracts

Résumé : Cet article envisage les contrats de services d'infonuagique au regard de la gestion des documents d'archives et des besoins archivistiques, afin de déterminer si ces besoins sont satisfaits par les contrats standards actuellement disponibles. Il en ressort qu'en général, les exigences de stockage et de conservation des documents sous une forme authentique ne sont pas satisfaites par les accords actuels de services d'infonuagique. L'article se termine en proposant une grille de lecture à l'attention des professionnels de la gestion et de la conservation des documents, à utiliser dans les négociations ou la sélection des contrats, afin de mieux répondre aux besoins de conserver des documents authentiques dans l'infonuagique.

Mots-clés : infonuagique, authenticité, contrats

Introduction

Cloud-based services are increasingly becoming a key part of how organizations worldwide conduct business activities. Encompassing a large array of possible models, the term "the cloud" describes a service wherein a client may purchase scalable access to information technology (IT) infrastructure for the creation, use, management, and/or storage of information. The ease with which large amounts of information can be affordably stored and accessed from anywhere

(V9 15/7/15 14:40) UTP (6"×9") AGaramond 1323 CJILS 39:2 pp. 128–153 1659 CJILS 39.2_04_Bushey (p. 128)

with an Internet connection has made these services an attractive option for organizations of various types. Despite this advantage, the risks of adopting cloud-based services are largely unrecognized and not well understood, which can result in cloud customers using services that do not meet with the best practices and legislation governing the management and storage of information and business records.

Research milestones

This article reports on research being conducted by the InterPARES Trust (ITrust) project on current cloud service contracts from a records management, archival, and legal perspective. ITrust (2013–18) is a multinational, interdisciplinary research project exploring issues concerning trust of digital information and records in the online environment (InterPARES Trust 2014). The research discussed in this article builds upon an earlier study conducted by ITrust entitled Project 10: Contract Terms with Cloud Service Providers, which explored the extent to which cloud service contracts met general records management requirements. The findings of this study revealed that the majority of selected cloud providers' contracts did not meet records management requirements. With this foundation, the current study aims to examine the issues further from the archival point of view and incorporate a stronger legal framework. The study is guided by the overarching research question: how effective are cloud service providers' terms and agreements at addressing the needs of records managers and archivists?

At this early stage in the research, the qualitative content analysis of selected cloud service providers' terms and agreements, recordkeeping standards, and legal requirements have been conducted and the results are discussed in this article, along with a preliminary set of recommendations/draft checklist for both cloud providers and customers to consider. The research that is described does not constitute the entirety of the project's work but, rather, reports on the research conducted thus far, in an effort to share preliminary findings with records managers and archivists as well as lawyers and cloud service providers and encourage feedback.

Terminology

There are several terms used throughout this article, which may be interpreted differently depending upon the disciplinary viewpoint. In an effort to provide consistency across ITrust research projects, definitions available in the *InterPARES 2 Dictionary* have been used whenever possible (InterPARES 2 Project 2015). In the context of recordkeeping standards and/or legal acts, which utilize specific terms and provide definitions, all attempts have been made to clarify usage to the reader and include citations. There is an absence of standardized terminology across cloud service contracts that refers to customer content (that is, data, information, and records); therefore, throughout this article the term data refers to the smallest meaningful units of information. The term information refers to an assemblage of data intended for communication either through

space or across time. The term record refers to a document made or received in the course of a practical activity as an instrument or a by-product of such activity that is set aside for action or reference.

In archival science, records are considered trustworthy if they are reliable (that is, they can stand for the facts they are about), accurate (that is, the records are precise, correct, and free of error or distortion), and authentic (that is, the records are what they purport to be and are free from tampering or corruption). It is important to understand that authenticity is established by assessing the identity (that is, the whole of the characteristics of a record that uniquely identify it and distinguish it from any other record) and the integrity of a record (that is, the quality of being complete and unaltered). The recordkeeping standard, *ISO 15489-1: Information and Documentation—Records Management Part 1*, which was issued by the International Standards Organization (ISO) in the fall of 2001, addresses authenticity and integrity separately, in that an authentic record is what it purports to be and a record has integrity if it can be proven that it has remained complete and unaltered after being set aside (ISO 2001). There is an absence of standardized terminology across cloud service contracts to refer to confidentiality, privacy, and security. Admittedly, it is beyond the scope of this article to discuss these terms in depth, therefore working definitions in the context of the ITrust project and its focus on cloud computing are provided. Confidentiality refers to the expectation that private facts entrusted to another will be protected and not shared without consent. Privacy refers to control over access and use of personal information. Security is the state of being protected from unauthorized access.

Methodology

Taking an interdisciplinary approach, the authors explored literature on cloud services and the terms and agreements in the areas of archival science, records management, and law. The literature review (discussed in second section of this article) identified key concerns regarding cloud contracts and balancing the specific needs of cloud service customers within the complex infrastructure and service delivery of large cloud providers. Furthermore, it identified several initiatives and guidelines, which address the challenges being faced by agencies and organizations considering the adoption of cloud-based services.

The second phase of the study involved the comparative analysis of selected cloud provider terms and agreements with recordkeeping standards and legal framework (discussed in the third section of this article). The analysis (presented in the fourth section) reveals several issues, in particular, gaps in the existing cloud provider contracts regarding the availability of metadata assigned to data, the ability to audit data, uncertainty about where the data are stored, difficulties in destroying and migrating data, and difficulties in establishing the authenticity of the data stored within these services. The recommendations (presented in the fifth section) are an attempt to address the gaps. The aim of the research is to provide a checklist to guide records managers and archivists through the process of assessing cloud provider contracts and determining if the agreements meet

recordkeeping standards and legal requirements specific to their organization and/or institution. The authors recognize that larger organizations can, and probably do, negotiate better contracts for their services, but many smaller organizations or members of organizations cannot do so or choose not to do it. Potentially, the checklist could be used within an organization to communicate the needs of records mangers and archivists to administration and IT support.

Literature review

The literature review undertaken in the course of this research was done to establish primarily a foundation for what the requirements for a recordkeeping system should be regardless of the medium, to review current research on cloud service agreements and their legal frameworks, and to determine what current standardization efforts exist for cloud service agreements.

Summary of the ITrust's Project 10

As stated earlier, the findings of Project 10: Contract Terms with Cloud Service Providers presented the impetus for the current study. Project 10 selected nine cloud service providers from Canada, the United States, and Europe in an effort to address a wider range of jurisdictions. Providers' online contracts were analysed, and a set of fifteen categories of contract terms was identified. Passages of the cloud provider contracts were classified as meeting, or not meeting, the general records management needs. The findings of the study reveal that most of the selected cloud provider contracts do not meet general records management requirements.

Recent studies on cloud computing

As part of the literature review, the authors explored recent studies on cloud computing, which stem from the areas of archival science, records management, and law. Through this review, the authors learned that cloud service terms and agreements tend to be broken into several legal documents (Bradshaw, Millard, and Walden 2011). The agreements include a general document for services, such as the terms of service, a document for more specific service such as the service level agreement, and documents covering other general areas such as the privacy policy and the acceptable use policy (ibid.). In addition, research has found that there very little standardization of terms exists across providers' agreements (Baset 2012). The authors of this article also took into account the literature by legal scholars, which examines privacy and security issues raised by the cloud's adoption. The authors noticed that several studies have been conducted to examine the legal framework for the use of cloud-based services by the federal and provincial authorities (Vermeys, Gauthier, and Mizrahi 2014).

Recordkeeping standards

In seeking to identify general recordkeeping requirements, the authors of this article consulted standards and guidelines that have been established by international bodies and organizations. Their primary source for identifying these requirements was the ISO's report *ISO 15489-1* (ISO 2001, 1), which became

their standard for records management. This source provides an overview for how records management should be carried out in public and private organizations, regardless of how the records are being kept and what form they take. The *ISO 15489-1*, however, does not include archival preservation of records in its requirements, and so additional sources such as the ISO's report *ISO 14721* were used to augment this need (ISO 2012).

ISO 14721 is the ISO's reference model for an open archival information system, which permits a designated community to preserve records and information that is created and kept in a digital environment (ISO 2012). The authors used this standard to approach the analysis of cloud service agreements from an archival perspective, in which information may need to be preserved indefinitely. This is an important issue, especially in light of what clauses need to exist within cloud provider contracts to support this type of long-term preservation.

In addition to these standards, the authors of this article also consulted the Association of Records Managers and Administrators (ARMA) International's *Generally Accepted Recordkeeping Principles®* (ARMA International 2013), which create an outline of concepts that a recordkeeping program should meet to be effective. These principles share a good deal with the standards created by the ISO, specifically *ISO 15489-1*, but provide less description of an ideal records management environment.

Also included in our literature review was the European Commission's *Model Requirements for the Management of Electronic Records* (2009). These requirements were included because they provide guidelines specific to records in a digital environment, such as how a system should implement audit trails, access restrictions, destruction, backup, and so forth for digital records.

Cloud computing contract standards

The authors included in their literature review efforts to standardize cloud service agreements. One effort, in particular, is the recently published guide from the European Commission entitled the *Cloud Service Level Agreement Standardization Guidelines* (European Commission 2014). This guide identifies topics for concern relating to information being stored with cloud providers and makes recommendations of service level objectives for inclusion in service level agreements (15–36). Many of the service level objectives identified by this guide express similar areas for concern identified by both the ITrust Project 10 and this article.

An additional set of guidelines can be found in the policy produced by the Public Records Office of Victoria, Australia (2012, 7) on the use of cloud computing. This policy identifies the leaking of sensitive information and the loss of information as the primary risks of using the cloud. It requires that agencies of the Victorian government conduct risk assessments on the implementation of the cloud before engaging in it; on all legislation, standards, and policies; and on all agreements between agencies and cloud service providers to ensure the security of data, that the ownership of the data remains with the agency, and that the agency is established as the controller of the data (7–8).

Comparative analysis recordkeeping standards, legal requirements, and cloud contracts

Recordkeeping requirements are identified in legislation, regulations, policies, and standards. Depending upon the nature of the business or organization and the applicable law, the requirements for maintaining control over records created, received, and stored, along with the systems that facilitate and support them, will vary. Records, which serve as documentary evidence of legal transactions and support the critical operations of the organization, have a high value and must be managed properly throughout their lifecycle. In addition, records that contain personal or sensitive information must be identified when they are created or when they are received and managed (including disposition) in compliance with applicable privacy and freedom of information legislation and statutes. The adoption of cloud-computing services to facilitate and support organizational activities involving the creation, receipt, use, and storage of information and records must be approached with caution due to the potential risks associated with the cloud.

In brief, the records management and recordkeeping community has identified the following risks to business use of cloud-computing services: unauthorized access to information and records stored in the cloud; privacy breaches; loss of access to, and management of, information and records stored in the cloud; alteration of information and records stored in the cloud (impacting record authenticity and integrity); and the lack of transparency regarding account management, server locations, data destruction, and data recovery (Ferguson-Boucher and Convery 2011; Public Records Office Victoria 2012). Therefore, before implementing cloud-based services, agencies, organizations, and institutions should carefully review the contractual agreements of the cloud-service providers, assess the risks, and determine the degree to which they meet the organization's strategy with regard to records management and recordkeeping.

Unlike traditional approaches to outsourcing information technology services, in which the services were negotiated directly with the provider, cloud computing introduces IT services on a grand scale. Cloud computing utilizes online platforms for delivery, circulates customer data throughout server farms scattered across the globe, and relies on generic terms and conditions to regulate their contractual relationships with customers. As a result, customers may be unaware of where the service infrastructure is located and if there are sub-contractors involved. In addition, the distributed characteristic of cloud computing may present obstacles to enforcing breaches of contract, especially in cases that involve security and data privacy (Public Records Office Victoria 2012).

Formally, the terms and conditions may be contained in a single document hosted on the provider's website or in a set of documents containing the terms governing the relationship between the customer and the cloud service provider (Bradshaw, Millard, and Walden 2011, 192). In general, these documents may include a service level agreement (SLA), terms of service, acceptable use policy, and privacy policy. If a cloud service is provided for free, the SLA is not included (ibid.). For the purposes of this study, all available documents were consulted and are referred to by their title throughout the following analysis.

At present, a standardized SLA for cloud computing does not exist at an international level. However, at a regional level, we already mentioned the existence of the European cloud SLA standardization guidelines. In addition, an initiative by the ISO is underway, entitled *ISO / International Electrotechnical Committee New Proposal 19086: Information Technology—Distributed Application Platforms and Services—Cloud Computing—Service Level Agreement Framework and Terminology* (ISO 2013). In the absence of an international standard for cloud-computing SLAs, organizations and archival institutions considering adoption of cloud computing to facilitate and support records management and/or digital preservation must assess cloud service providers' terms and conditions before implementation, taking into account not only their records management and recordkeeping needs but also their legal duties.

Recordkeeping requirements and cloud provider terms and conditions

The following analysis utilizes the *ISO 15489-1: Information and Documentation – Records Management Part 1*, which was issued by the ISO in the fall of 2001, to identify recordkeeping requirements that should be taken into consideration when an organization assesses cloud-computing services for managing and storing their records. *ISO 15489-1* is technology neutral and includes sections on records system design and implementation (section 8) and records management processes and controls (section 9), which support the creation and maintenance of authentic, reliable, and useable records and protect the integrity of those records for as long as required (ISO 2001, 6). The comprehensive nature of the standard makes it suitable for addressing current records (that is, in use by the organization) and non-current records (that is, no longer in use but set aside for future reference).

ISO 15489–1 also identifies the characteristics of authoritative records, which are records that correctly reflect what was communicated, decided, or the action taken and support the needs of the business, and they can be used for accountability purposes (ISO 2001, 7). In addition to content, authoritative records should also contain, or be linked to, metadata that documents the structure of a record, the business context, and the links between documents that participate in the same activity (7). According to the ISO standard, the characteristics of authoritative records are authenticity (that is, an authentic record is what it purports to be), reliability (that is, a reliable record is one whose contents are accurate and the persons responsible for its creation have the authority to do so), integrity (that is, a record has integrity if it can be proven that it has remained complete and unaltered after being set aside), and usability (that is, a useable record is one that can be located, retrieved, presented, and interpreted). Throughout this article, reference will be made to these characteristics, as defined by *ISO 15489–1*.

In addition, *ISO 14721: Space Data and Information Transfer Systems—Open Archival Information System Reference Model*, which was issued by the ISO in 2012, is used in the following analysis to address the roles, responsibilities, and expectations of cloud providers and their clients as well as the specific

requirements related to the preservation environment (ISO 2012). *ISO 14721* provides a framework as well as the concepts needed by non-archival organizations (that is, cloud providers) to be effective participants in the preservation process.

An analysis of cloud-computing terms and conditions documents in the context of recordkeeping standards reveals several key issues for discussion: data ownership; availability, retrieval, and use; data retention and disposition; data storage and preservation; security; data location and data transfer; and end of service—contract termination. For the purposes of this project, the following cloud providers were identified in the ITrust's Project 10 and selected for further analysis: the Google Cloud Platform (United States), the Pathway Communications CloudPath (Canada), and the GreenQloud (Iceland).[1] The rationale for their selection is based on international representation, online access to terms and agreements, and limited resources. Every attempt has been made to consult the most current version of the terms and conditions documents available on the selected cloud providers' websites; however, it is common practice for terms and conditions to be updated. The cloud provider reserves the right to vary contract terms and amend its terms and conditions by posting an updated version to their website, noting that the continued use of the service by the customer is considered to demonstrate acceptance of the new terms and conditions (Bradshaw, Millard, and Walden 2011, 202).

The key issues will be addressed using an interdisciplinary approach, in which the specific recordkeeping requirement and legal framework will be identified and contrasted with selected sections from the cloud providers' terms and conditions documents. The degree to which the terms and conditions meet recordkeeping requirements will be discussed, along with the implications for recordkeeping activities within the organizations and archival institutions.

Data ownership

The authors of this article recognize that information in digital form accessed and stored in the cloud cannot be owned in the same manner as physical objects, at least not in the way as information transcribed onto a physical medium.[2] However, it can be controlled at a similar level by intellectual property rights, confidentiality or privacy, and contracts (Reed 2010, 1). For simplicity, this article will operate under the assumption that data ownership does not require a physical medium. The recordkeeping standards approach data ownership by stating that records may be physically stored with one organization, but the responsibility and management control may reside with either the creating organization or another appropriate authority. Records stored in electronic systems require arrangements that distinguish between the ownership of the records and the storage of the records (ISO 2001, s. 8.3.4).

However, data ownership in the cloud is a complex issue, not only because of the intangible nature of digital information but also because of the infrastructure of cloud computing itself, in which an individual or organization may entrust their information and records, along with others, to a cloud provider

and use the provider's platform and applications in the cloud to create further information and records, while the provider may create a great deal of information related to these operations for several purposes (for example, data processing, management, marketing, and so on).

It can be reasonably understood that information generated by the customer and stored in the cloud does not belong to the service provider (Reed 2010, 17) but, rather, that the provider is authorized to do specific operations with it to provide the service. Metadata generated by the service provider about the customer's information and operations in the cloud can raise more issues. These metadata can be important for the customer to further demonstrate that the integrity and the security of the data have been preserved. However, this information is owned by the service provider, who generated it for internal purposes—that is, to manage the cloud and ensure the use and quality of the service (Reed 2010, 9). Beyond the ownership issue, the contract terms and conditions should determine whether and how the customer has the right to access and use this metadata for recordkeeping purposes, during the contractual relationship but also at the end of the service (see the discussion later in this article).

Analysis of the terms and conditions documents for terms that declare ownership or responsibility for customer information and content reveals a lack of consistency in terminology and placement, which may lead to confusion when organizations are trying to evaluate several different service providers. Google is the most declarative and places the notice of being a data processor at the outset of their terms of service, whereas Pathway Communications makes a distinction between client data and information generated during the process of providing the cloud service. In doing so, Pathway Communications is imposing ownership of intellectual property via the terms and conditions. GreenQloud does not seek to assert intellectual property rights over customer content accessed and stored in their services. None of the three providers mention in their terms and conditions the right of the customer to access internal system metadata or the conditions to use metadata under license, for instance. As explained earlier, if the customer needs to access internal system metadata for recordkeeping purposes, the provider has the right to deny access to this metadata or to ask for additional fees to facilitate access and/or use.

Google Cloud Platform's terms of service includes section 1 on the provision of the services, in which Google is identified as "merely a data processor" (section 1.3). In doing so, Google identifies as being the service provider/data processor, who only acts upon instructions from the customer. The customer/data controller determines the purposes and means of processing personal information and customer content. This appears to be an oversimplified approach to the relationship between Google and its customers, especially as the cloud service provider often makes important decisions about the processes of managing and storing customer information and content. In section 3 on customer obligations, responsibility for customer data are assigned to the customer (section 3.1), specifically the management of intellectual property (section 3.6) and protecting the privacy and legal rights of end users (section 3.2). In direct reference to

the Digital Millennium Copyright Act, Google relies on copyright holders to manage their intellectual property online (section 3.6).[3]

Pathway Communications CloudPath's terms of service include section 8 on client data, in which responsibility for the storage, care, custody, and control of client data are assigned exclusively to the customer (section 8.3). Towards the end of the terms of service, there is section 20 on ownership of intellectual property, in which the cloud provider claims ownership of any intellectual property developed by Pathway during the performance of cloud services (section 20.1).

GreenQloud's end-user license agreement and terms of service includes section 5 on your responsibilities, in which the customer is assigned responsibility for the technical operation of customer content with the provided service (section 5.1a), managing customer content in a manner that complies with Icelandic laws on privacy and trade secrets (section 5.1b) and addressing any claims related to customer content (section 5.1c).

Availability, retrieval, and use

The importance of having information and records available to the organization to fulfil their immediate and future business needs is one of the driving forces behind the adoption of the cloud. Recordkeeping standards, such as ARMA International's *Generally Accepted Recordkeeping Principles* (2013), emphasize that records must be available for access and retrieval in a timely and efficient manner. Moreover, availability and retrieval is not only a question of efficiency but also a legal issue, as it is closely linked to statutory or constitutional rights to have access to certain data. To be more precise, availability is a fact and access is a right, but the latter cannot be satisfied without the former (Vermeys, Gauthier, and Mizrahi 2014, 86). Another issue is to control who can access the data and to protect the data's integrity and confidentiality, which is more a security issue that will be examined later in this article.

According to the data protection laws in Canada (see Privacy Act, the Personal Information Protection and Electronic Documents Act (PIPEDA), and similar provincial statutes),[4] in the United States, or in Europe, individuals have a right to access their own personal information held by an organization, whether public or private (except, in the latter case, for the US system, which provides for self-regulation by industry). Similarly, a lot of countries provide a general right of access to information held by public bodies and government organizations. In Canada, such a right is granted by the Access to Information Act and by equivalent provincial statutes.[5] Similar legislation has been adopted in the United States and in Europe. According to these laws, organizations must be able to provide access to the requested information within a period that may vary, depending on the legislation, from twenty to thirty days. This may seem quite reasonable from a technological point of view, but one has to consider the time needed to process the request from an administrative point of view, identify all of the requested documents, and evaluate whether some information should fall under one of the exemptions from access stated by law. Therefore, this administrative process cannot be retarded by technical difficulties to retrieve

and access the recorded information. In this respect, the availability of the stored data implies also the availability of the infrastructure, hardware, and software, which facilitates the retrieval and readability of the data (Vermeys, Gauthier, and Mizrahi 2014, 88). Of course, the fact that an organization is using a cloud-based service provided by a third party is not a reason to justify any delay in the processing of the request. In this case, if the organization is unable to provide access to the requested data, they remain liable and expose themselves to a complaint that could lead to specific sanctions.

Analysis of the terms and conditions documents for terms regarding availability, retrieval, and use of customer content reveals the use of SLA to present monthly uptime percentages (that is, total number of minutes in a month minus the number of minutes of downtime experienced in a month, divided by the total number of minutes in a month) and assures customers that cloud services are reliable and continuous. All three of the selected cloud service providers claim service availability of 99.99 percent. Service credits are supplied in the event of failure to meet performance standards; however, the list of exceptions is long and the onus is on the customer to determine which types of outages, downtime, unavailability, losses, delays, or problems actually constitute a failure and qualify for service credit.

Google Cloud Platform provides a separate document entitled *Data Processing and Security Terms*, in which they agree to make customer data available to the customer in accordance with the terms of the agreement. There is an additional clause, in which Google will assist the customer in the deletion and migration of customer data in the event that the customer is unable to do so, but this service comes with a fee. Pathway Communications CloudPath's SLA includes section 4 on performance standards, in which Pathway provides target percentages and time periods for each of their cloud-based services (that is, cloud server hosts, cloud storage, network, and cloud migration). GreenQloud's SLA addresses availability in their uptime section. Divided into three areas: data centre power, public network, and cloud instance uptime, GreenQloud guarantees 100 percent uptime. In the event of downtime, credit is allotted to the customer's account. The durations that qualify for credit are twenty minutes of data centre downtime, one hour of cloud instance downtime, and any length of public network downtime.

Data retention and disposition

Records management divisions within organizations and preservation activities conducted by archival institutions rely on data retention and disposition schedules to perform information governance and remain compliant with increasingly complex legal and regulatory environments. Recordkeeping standards suggest that decisions made by the organization regarding the retention and disposition of records should be carried out and implemented by the electronic system. The electronic system should be capable of producing audit trails to track disposition activities (ISO 2001, section 8.3.7).

In some cases, disposition actions may require transfer of the records from one electronic system to another. The transfer should not alter the authenticity, reliability, integrity, or usability of the records. Authorized records destruction must be performed in a manner that preserves the confidentiality of the information. The process of record destruction should include all copies throughout the system and related metadata (ISO 2001, section 9.9). This can raise difficulties for the metadata generated, which is owned by the service provider in relation to the customer's data and operations in the cloud. Having ownership of such metadata (see discussion earlier on data ownership), the provider could refuse to destroy his own metadata if they are still useful for internal systems management purposes (for example, statistics, service improvement, and so on).

Analysis of the selected cloud providers' terms and conditions reveal an absence of terms that address data retention or deletion according to customer-stipulated schedules or recordkeeping requirements. Google Cloud Platform's data processing and security terms include section 5 on data correction, blocking, exporting and deletion, in which Google provides the customer with the ability to delete customer data in accordance with the functionality of the selected service. Terms in the terms and conditions assert that once the customer deletes their data and it is no longer recoverable by the customer, Google will delete or render permanently inaccessible the customer-deleted data within a maximum period of 180 days. In the case of data whose destruction is required by law under a specific schedule, the legal schedule could be overruled by up to six months. The customer would remain liable for such an infringement, as it is his legal duty to use procedures or services that ensure the destruction of the data at the right time. In the context of organizations that are required by law to delete certain types of records, more information about how customer data are rendered permanently inaccessible is required. In addition, the terms in their terms and conditions do not clarify if "inaccessible" data would be available to law enforcement through an e-discovery request.

Data storage and preservation

The manner in which records are stored after they are no longer in active use by the organization impacts the quality of the records and their capacity to be used for accountability purposes. In addition, evidence law directly or indirectly imposes certain precautions on the processing of the data to ensure a strong evidentiary value of the information brought before the court. This is the case in civil law jurisdictions (such as Quebec, France, or Belgium) where the integrity of the electronic record is a formal condition to recognize it as the legal equivalent of a paper record—that is, as "writing" within the hierarchy of the means of evidence. This integrity must be preserved throughout the lifecycle of the record.

Determining what actions are required by a system that stores records for the long term and provides preservation of digital information is challenging for organizations, especially if cloud providers are not transparent about the infrastructure and processes involved in providing cloud-based storage. The

task of maintaining information and records throughout changing technologies, new data formats, and evolving requirements for use requires knowledge of, and adherence to, recordkeeping standards aimed at digital preservation.

Recordkeeping standards state that systems selected by an organization for storing electronic records should ensure that the records held within the system remain accessible, authentic, reliable, and useable throughout any changes made to the system. If the systems provider implements changes, then audit trails and process metadata should be made available to the organization (ISO 2001, section 9.6). Planned migration and/or emulation of hardware, software, and/or operating systems by the electronic records system provider should not impact the authenticity, reliability, and usability of the records held within the system (section 8.3.5).

Analysis of the selected cloud providers' terms and conditions reveal terms that state that the customer is responsible for backing up the application, project, and customer data (Google 2014). In general, activities aimed at storing data and records for any length of time are referred to by cloud providers as backup procedures. The actions to preserve or the activity of preservation are absent from all terms and conditions documents.

Pathway Communications CloudPath terms and conditions agreement includes section 8 on client data, in which the provider states that it is the responsibility of the client to ensure the proper storage, care, custody, and control of client data, including regular backups of client data to non-Pathway systems to "ensure against loss or corruption" (section 8.3). Although Pathway Communications admits to creating backups of their systems on a periodic basis, the cloud provider does not guarantee customer access to "snapshots" (section 8.1). Alternatively, Pathway Communications CloudPath provides data backup as a fee-based service (section 4.3.1 and section 5.1.4), which includes integrity checks on backup sessions (section 4.2.4) and support for restoring client data due to a failure of the Pathway's backup system (section 4.6.3). However, there are number of limitations listed in relation to backup services and Pathway's backup system (section 4.6). In addition, the cloud provider includes terms that make it clear that scheduled maintenance may impact customer data; therefore, customers are required to back up their data to a non-Pathway location before scheduled maintenance occurs (section 5.1.4).

GreenQloud's end-user license agreement and terms of service include section 10 on other security and backup, in which the customer is deemed responsible for maintaining appropriate backup of customer content. The terms include a reference to the customer's responsibility to protect their content by performing "routine archiving."

Security

Security is a control measure implemented throughout the electronic records system that prevents unauthorized access, destruction, alteration, or removal of records. Among the security measures to be taken, the protection of the con-

fidentiality of the data through access control is of crucial importance. Access to records stored in electronic systems should be managed through controls on access to ensure the integrity of the records and protect against unauthorized access, use, alteration, or destruction. Any change in the format of records transferred to the system and/or delivered to the user should be specified. The electronic system should be capable of producing audit trails and/or access logs to demonstrate that records are being protected from unauthorized access, use, alteration, or destruction (ISO 2001, section 8.3.6). The electronic system should capture and maintain metadata associated with the access, retrieval, and use of records within the electronic system. This includes metadata that is embedded or linked to records as well as metadata generated by the electronic system during processes associated with the management of records (section 8.3.2). In the case of a system malfunction or security breach, the cloud service provider should notify the client organization immediately and demonstrate the integrity of the system by providing access to tracking that reveals the movement and uses of records within the record system (section 8.2.3 and section 9.8.1).

From a legal perspective, such security measures are requested under data protection legislation. Sectorial regulations at a provincial, national, or international level must also be considered—for instance, those related to the financial markets (such as the Sarbanes-Oxley Act or the Basel Accords).[6] As mentioned earlier in the discussion on data preservation), the evidentiary value of the record will depend on the actions taken on the data throughout its entire lifecycle to preserve its integrity and authenticity, which includes security measures. More specifically, the duty to ensure the confidentiality of the data is a very common legal requirement that can be found in hundreds of different statutes and regulations (Vermeys, Gauthier, and Mizrahi 2014, 95 n401). In the following considerations, the authors mainly focus on security requirements with regard to personal data.

According to the principles set out in the Model Code for the Protection of Personal Information, included in Schedule 1 of PIPEDA, "an organization is responsible for personal information in its possession or custody, including information that has been transferred to a third party for processing. The organization shall use contractual or other means to provide a comparable level of protection while the information is being processed by a third party" (Principle 4.1.3). Such a principle can be found in most regulations to ensure the protection of personal data. The fact that the data have been transferred to a third party processor does not transfer the accountability of the organization. In such a situation, it is interesting to note that the contract is considered to be a key element to ensure security (Office of the Privacy Commissioner of Canada 2009, 9). Therefore, organizations considering the use of cloud-based services should pay special attention to the service provider's contract terms related to security and check if they explain how the security of the data is ensured through technical, physical, and organizational measures.

Analysis of the cloud providers' terms and conditions reveal different degrees of addressing security issues. Of the three selected cloud providers, Google is the only one that has a separate document entitled *Data Processing and Security Terms*, which is made available through a hyperlink buried deeply in section 15 of Google's terms of service. In this document, security terms are discussed at length, pertaining to both the physical infrastructure providing the services and customer content and account information (that is, personal data). The degree to which cloud providers will deliver security measures to customers appears to be reliant on the types of services being offered (for example, managed or non-managed) and whether or not the customer pays additional fees. Moreover, concerning controls on access and use of customer data, the selected cloud provider terms and conditions focus on assigning responsibility to the customer for managing access restrictions to their account and their content.

Google Cloud's data processing and security terms include section 1 on the provision of the services, in which the provider states that all facilities that store and process the application and customer data must adhere to security standards set forth by the "industry" (section 1.3). Later, in section 4 on data security, the cloud provider states the implementation of appropriate technical and organizational measures to protect customer data from accidental loss, unlawful deletion, alteration, or unauthorized access (section 4.1). In the event of a "data incident," Google will notify the customer after the incident has been identified and measures to secure the customer's personal data has been performed (section 4.3). The security terms are discussed further in Appendix 2 on Security Measures, in which data centre and network security (section 1), access and site controls (section 2), and data (section 3) are listed. These security measures are both physical and virtual, addressing infrastructure security and measures taken to protect unauthorized persons from gaining access to the system and data centres, the multi-tenant environment on Google-owned servers, access controls for administrators and end users, logging capabilities available to the customer (that is, audit trails), as well as the process for handling hardware failure and performance errors.

Concerning the control of access and confidentiality, Google considers customer data to be the customer's confidential information (section 15.15). Google will not disclose a customer's information, except to the persons who need to access it to fulfil Google's obligations under the agreement and who have agreed to keep it confidential (section 7). In Appendix 2, Google also identifies the multi-tenant environment used by Google-owned servers and states that the customer will be given control over specific data-sharing policies (section 3a). Furthermore, Google states that the combination of policies and the functionality of selected services will enable the customer to determine the product-sharing settings applicable to end users for specific purposes. Google also makes available certain logging capabilities to the customer. The wording seems to imply that customers must shape their access controls to the existing functionality of Google services, which may not accommodate customization based on requirements promulgated by recordkeeping standards.

By comparison, Pathway Communications CloudPath's terms of service include section 4 on scope and limitations of the services, in which the cloud provider includes terms for non-managed services. Specific to security, Pathway communications takes responsibility for the physical security of the hardware (networking, storage, and servers) and the software that hosts the cloud services (section 4.1.5). The terms for fee-based managed services include support for server monitoring and response (section 4.3.2) and firewalls (section 4.3.5). Yet there are additional services deemed "specialty services" that are excluded, such as migration services and restoring customer data (section 4.4 and section 4.6.3). The responsibility for monitoring access to customer data are addressed in Pathway Communications CloudPath's terms of service under section 9 on unauthorized access, in which Pathway declines responsibility for any unauthorized access to customer data (section 9.2) and states that the customer is responsible for maintaining the security of their access credentials and for all activities that occur under their account (section 9.1).

GreenQloud's end-user license agreement and terms of service include section 10 on other security and backup, in which the provider assigns responsibility for maintaining appropriate security protection of customer content to the customer. In section 2 on the customer's account and section 3 on acceptable conduct, it is mentioned that access to GreenQloud's services through a customer account is the responsibility of the customer, regardless of whether the activities are undertaken by the account holder or their employees. There is no mention of audit trails or access logs.

Data location and cross-border data flows

In cloud computing, the processing and storage services can be provided on-demand by using several the cloud provider's resources throughout the globe. As a result, legal concerns regarding cloud computing focus on the issue that the customer's data may be stored or processed in different locations and unknown jurisdictions (Bradshaw, Millard, and Walden 2011, 206). From a legal perspective, the main issue raised by the location of data is the storage of data outside the customer's jurisdiction. This can be a concern with regard not only to data protection laws but also to foreign laws that allow investigation agencies access to any data stored in a provider's jurisdiction. The most famous example is the US Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act, also known as the Patriot Act.[7] Nevertheless, this major concern is often based on wrongful assumptions as to the application of such laws and needs to be examined in more detail.

First of all, contrary to a common misunderstanding, in Canada neither the Privacy Act nor PIPEDA prohibit the use of cloud-based services by public or private bodies, even if it implies a transfer of data outside the country. Provincial laws themselves do not restrain cross-border data flows, except for British Columbia, Nova Scotia, and Quebec (Klein 2008, 4 and 14; Vermeys, Gauthier, and Mizrahi 2014, 45, 112):

> Much of the confusion stems from the mistaken belief that Canadian privacy laws require Canadian organizations to shield personal information from a foreign government's ability to lawfully access that information. Most countries, including Canada, have laws permitting government agencies to access personal information within their jurisdiction for national security and law enforcement purposes. Despite the fact that some of these laws potentially permit broader government access than the USA Patriot Act (such as in the United Kingdom), transfers that may be subject to the USA Patriot Act are the source of the most confusion and misinformation." (Klein 2008, 4)

One common misunderstanding seems to be that only data stored in the United States are subject to the Patriot Act. In fact, according to this act, the US government has widespread powers to access data not only stored on servers located within the United States but also stored anywhere with a cloud-service provider that is registered in the United States or that conduct continuous and systematic business in the United States (Van Hoboken, Anrbak, and Van Eijk 2012, 36; Vermeys, Gauthier, and Mizrahi 2014, 49). In addition, as already mentioned, the US Patriot Act is certainly not a unique piece of legislation, as similar laws have been adopted by other governments, including Canada. Therefore, wherever the data are stored, whether or not in the cloud, organizations may be subject to similar types of orders to disclose information to governmental authorities (Office of the Privacy Commissioner of Canada 2005; Vermeys, Gauthier, and Mizrahi 2014, 49). One must also mention the fact that according to the Patriot Act, "a company subject to a section 215 order cannot reveal that the FBI has sought or obtained information from it" (Office of the Privacy Commissioner of Canada 2005). Nevertheless, an appropriate level of transparency can be reached if the service provider mentions in the contract that the data stored in the cloud may be subject to such disclosure orders. In addition, if an organization chooses to store personal data in the cloud of a service provider, it should inform individuals "that their information may be processed in a foreign country and that it may be accessible to law enforcement and national security authorities of that jurisdiction" (Office of the Privacy Commissioner of Canada 2009, 8 and 9).

However, even if the law does not prohibit the transfer of personal data outside Canada, organizations should assess the risks of jeopardizing the integrity, security, and confidentiality of personal information entrusted to third-party service providers, wherever they are located (Office of the Privacy Commissioner of Canada 2009, 7 and 9). It has also been noted that certain countries, provinces, or regions restrict the possibilities to transfer certain data outside their jurisdiction. In Canada, British Columbia and Nova Scotia require public bodies to ensure that personal information in their custody or under their control is stored and accessed only in Canada, which would prohibit the use of cloud-computing servers based outside the country. Nevertheless, without entering into details, these restrictions provide for several exceptions and do not apply to private bodies (Klein 2008, 11; Vermeys, Gauthier, and Mizrahi 2014, 51). In Quebec, restrictions are imposed for the storage of personal data outside the province. In short, public and private bodies must ensure that the personal data will receive

an equivalent level of protection under local privacy laws than under Quebec privacy laws. While it has been recognized that such an equivalent protection is offered by other provincial privacy laws and by federal laws in Canada, as well as by European laws, some doubts might be raised for the storing of personal data in the United States (Vermeys, Gauthier, and Mizrahi 2014, 117; compare Klein 2008, 11). This issue can also lead to difficulties for a Canadian provider having servers located in the United States or for servers located in Quebec under the control of a foreign provider. However, considering the practical difficulties raised by such a restrictive approach, Nicolas Vermeys, Julie Gauthier, and Sarit Mizrahi (2014, 129) suggest that it could be possible to abide by the spirit of the act by using encryption technologies to protect data before storing them in the cloud, wherever the servers might be located (see also Canellos 2013).

Finally, it is well known that the European Union has also adopted a restrictive legal framework with regard to the transfer of personal data outside Europe, requiring that the privacy laws of the country of destination offers the same level of protection as the EC Directive 95/46 on the Protection of Individuals with Regard to the Processing of Personal Date and on the Free Movement of Such Data.[8] In this context, the European Commission has officially considered Canada as providing an adequate level of protection for personal data transferred from the European Union to recipients subject to PIPEDA.[9] In the United States, companies may comply on a voluntary basis to the Safe Harbor international privacy principles, a program settled by the US Department of Commerce in consultation with the European Commission and officially recognized as offering an adequate level of protection.[10]

In addition to these issues, the location of data might also be a criterion (among others), according to the rules of conflict of law, in determining the law that applies in the case of litigation if the parties have not chosen the law governing the contract (Goh 2014, 59). However, most cloud providers include terms that state the choice of forum for settling disputes between the provider and customer. In general, cloud providers select a jurisdiction compatible with their own legal system (for example, Pathway Communications is based in Ontario, Canada).

Recordkeeping standards do not address jurisdiction and limit their discussion to location. The electronic records system should be able to track the location of records as they move throughout the system (ISO 2001, section 9.8.3). Google Cloud Platform's terms of service include section 1 on the provision of the services, in which the cloud provider asserts the right to transfer, process, and store "an application and customer data in the United States or any other country in which Google or its agents maintain facilities" (section 1.3 and section 8.1). Google also mentions the fact that it is, and will remain, enrolled in the Safe Harbor program and will adopt a solution that achieves compliance with the terms of EC Directive 95/46 (section 1.5). The terms of service expressly state that the customer has the obligation to protect the privacy and legal rights of its end users under all applicable laws, which includes the communication of

a privacy notice, the obtaining of any required consent and the obligation to inform end users that the data will be processed by Google (section 3.2). The terms also stipulate that notwithstanding any non-disclosure agreement, Google will disclose confidential information to the extent required by the applicable legal process and under certain conditions (section 7).

Google Cloud Platform's data processing and security terms include section 8 on data transfers, in which the provider states that the customer may select where "certain customer data will be stored permanently, at rest" (section 8.2). These terms appear to be linked to specific services, yet it is unclear exactly what storing data permanently entails or what data at rest means. In addition, if a customer is not a US city, county, or state government entity, then all claims related to the cloud services will be governed by California law and litigated in the federal or state courts of Santa Clara county in California (section 15.10).

Pathway Communications CloudPath's terms of service include section 28 on governing law, in which the provider states that the agreement is governed by the laws of the province of Ontario and that all disputes arising from cloud-based services will be addressed in that specific jurisdiction (section 28.1). GreenQloud's end-user license agreement and terms of service include section 5 on the customer's responsibilities, in which compliance with Icelandic law is required.

End of service: contract termination

In the event that the relationship with a cloud provider ends, the organization needs assurance that it can gain access to its information and records and that any data it leaves behind in the third-party system will be deleted by the cloud provider (Bradshaw, Millard, and Walden 2011, 203). There are several reasons why the services may be terminated, some of which relate to actions taken by the cloud provider or the customer or simply to the scheduled end of the contract. It is important that organizations are aware of contract termination procedures before adopting cloud services. An additional consideration is whether the provider is offering a paid service or a free one (196). Contracts for paid services address the duration of the service and the necessary steps to terminate the contract, whereas free services do not have fixed durations and may reserve the right to close inactive accounts.

Recordkeeping standards address the discontinuation of a records system as an event that should not preclude ongoing access to those records formerly held by the system. System providers should ensure the removal of all records and associated metadata from the system in a manner that does not impact record authenticity, reliability, usability, and integrity. In cases of account termination, the records system provider should ensure that all records and associated metadata are transferred to the organization in a manner that does not impact record authenticity, reliability, usability, and integrity (ISO 2001, section 8.5). Archival organizations using third-party services for long-term preservation of their archival records must have a formal contingency plan in case the archives or the third provider ceases to operate (ISO 2012, section 3.2.5).

Analysis of the selected cloud provider terms and conditions reveal two related, but different, activities: suspension of services and termination of services. Suspensions are typically in response to customer violations of the cloud service and require investigation by the cloud provider to determine restoration of the service and access to customer content or deletion of the account and customer content. Termination of services may be the final result of a suspension, the result of account inactivity, or the response to end of contract term.

Google Cloud Platform's terms of service include section 8 on term and termination, in which the cloud provider presents three types of termination: termination for breach (section 8.2), termination for inactivity (section 8.3), and termination for convenience (section 8.4). The effects of termination include terms that require the customer to delete the software, any application, instance, project, and customer data and that, upon request, each party will return or destroy confidential information of the other party (section 8.5). Google reserves the right to terminate services in the event of account inactivity exceeding 180 days (section 8.3).

Pathway Communications CloudPath's terms of service state in section 8 on client data that the customer will not have access to their data during a suspension or following termination (section 8.1). In addition, unless written modification is agreed upon, the cloud provider is free to delete client data from the system within seven days of termination of the account (section 8.4). Later in section 14 on service suspension or termination, the cloud provider includes several reasons for which the cloud provider can suspend or terminate services without liability, including unauthorized access by a third party (section 14.1.4) and overdue payment (section 14.1.6). The cloud provider will give "reasonable advance notice" of suspension of service. However, in the event of termination, the cloud provider is not obligated to refund payment and may prevent customers from accessing their data (section 14.2). In the case of a breach of contract, notice of account termination will be sent to the customer (section 16).

GreenQloud's end-user license agreement and terms of service include section 6 on suspension and termination, in which violations of the agreement will result in suspension or termination of the customer's account. During investigation of the suspected violation, all accounts are suspended. The cloud provider will not refund the customer for suspending or terminating accounts that are a result of violations of the agreement. GreenQloud states that it will try to notify the customer before suspension or termination. In the event of account suspension without cause, the cloud provider will provide fourteen days advanced notice. In section 7 on effect of termination, the customer is responsible for all fees and charges for in-process tasks that were completed by the cloud provider after the date of termination. Retrieval of customer data following termination is only available to clients that have paid for post-termination use of the provider's services.

Findings and discussion

Based on a thorough analysis of selected cloud providers' terms and agreements, the findings reveal that some boilerplate contracts, without additional fee-based services, are ineffective at meeting the recordkeeping needs of organizations and institutions operating within specific legal requirements. While some of the agreements do touch on the needs of records management and preservation, these sections of the agreements clearly aim, unsurprisingly, to protect the service provider rather than the client and its records needs. This is likely due to the reality that, because boilerplate agreements can be easily entered into by anyone, they have the potential to expose the service provider to a large amount of risk, which is further complicated by the fact that many of the companies offer similar terms, but the terms differ in their implementation. It is particularly true in the case of the uptime percentage terms of SLAs, which differ in how uptime is measured and how recompense is offered. It is also true with the terms on copyright and ownership, which may guarantee that the clients own their own data, but not the data created by the service provider in provisioning the service. This would likely mean that metadata assigned to records during their storage and use within the service would be unavailable, making proper audits and preservation extremely difficult.

Thus, records managers and archivists need to identify and establish the relevant regulatory and legal framework, in which the agency, organization, and/or institution operates within before adoption of cloud-based services. Areas such as public records requirements, freedom of information, and protection of privacy (POP) requirements, accountability requirements, security requirements, data location requirements or restrictions to cross-borders data flows, evidentiary requirements, and intellectual and copyright protection necessitate degrees of compliance and should be considered as part of the organization's recordkeeping strategy (Public Records Office Victoria 2013, 6). Private organizations, which do not handle public records are not subject to as rigorous a regulatory environment, except for POP; however, records managers and archivists still need to base their decisions on the availability of service required, the ability to execute records scheduling and disposition, the assurance of record reliability and authenticity, data privacy, long-term access, and system security. In any case, the provisions related to the end of the contract should be carefully examined to ensure a complete restitution of the data in a format that preserves their authenticity, with all of the associated metadata that ensure their traceability as well as the warranty that all of the customer's data are permanently and immediately destroyed after such a restitution.

However, it is possible that within the context of cloud services, some needs of records management may not be possible given the nature of the cloud. The purported benefits of cloud technology in sharing hardware to decrease costs, for example, cause extreme difficulties in ensuring that information has been irrecoverably destroyed when necessary. This is because the infrastructure that this information is stored on likely contains information from other clients or even information from the same client that is still needed, making degaussing

(that is, eliminating a magnetic field), physical destruction, or even wiping impossible. Another example of a difficulty that may emerge is the difference in where data may be stored permanently at rest, as Google refers to it (section 8.2). While the service provider may be able to guarantee that the client's data will be stored in a particular jurisdiction, it may be difficult to ensure that the data will not pass through jurisdictions that the client may be unaware of or that are unwanted by the client. It is conceivable that this data may be compromised or backed up during the transfer process, as the service providers do not specify what the transfer process to a state of permanent at rest may be or how long it may take.

Recommendations and further research

Any effort to aid records managers and archivists in entering into agreements with cloud-service providers should attempt to account for the needs of record-keeping systems as described by recordkeeping standards. In addition, it is important to recall that an organization that decides to opt for a cloud-based solution with a service provider must still fulfil its legal duties and remains accountable for the compliance with the requirements imposed by law. In this respect, specific issues that need to be addressed in cloud provider contracts are listed in the following checklist.

1. *Data ownership*
- Who owns the data stored, transmitted, or created in the cloud by the customer (that is, you)? Does the service provider have the right to use them and, if so, to what extent?
- Who owns the metadata generated by the system during procedures of upload, management, download, and migration? Do you have the right to access them for recordkeeping or legal purposes during the contractual relationship and at the end of the contract (see also section 7 on end of service)?

2. *Availability, retrieval, and use*
- Are SLA using precise indicators and providing clear information about the availability of the service?
- Does the degree of availability of the data fit with your business needs and allow you to comply with the freedom of information legislation (if you are a public body), the right of a person to access her own personal data, and the right of authorities to legally access your data for investigation, control, or judicial purposes?

3. *Data retention and disposition*
- Are your data (and all their copies) destroyed in compliance with your data retention and disposition schedules? If so, are they immediately and permanently destroyed in a manner that prevents their reconstruction, according to a secure destruction policy ensuring confidentiality of the data until their complete deletion?

- What is the nature and content of the associated metadata generated by the provider? Considering their nature and content, do they need to be destroyed at the same time and in the same manner as your data to comply with your internal or legal destruction policies? If yes, will the service provider proceed to such destruction?
- Does the system provide and give you access to audit trails of the destruction process? Will you receive an attestation, report, or statement of deletion from the provider, if requested by your internal or legal destruction policies?

4. *Data storage and preservation*
- Who is responsible for creating backups of customer data and recovering deleted or corrupted data?
- Are records migrated or emulated in a way that preserves their authenticity, reliability, integrity, and usability? Does the system provide and give you access to audit trails concerning the migration/emulation process?
- How will the service evolve? Will you be notified of any evolution of the service that could impede the authenticity of your data?

5. *Security, confidentiality, and privacy*
- Does the system prevent unauthorized access, use, alteration, or destruction of the data through technical, physical, and organizational measures? Does the system provide and give you access to audit trails, metadata, and/or access logs to demonstrate this?
- Will you be notified in the case of security breach or system malfunction?
- Does, or will, the service provider use the services of a subcontractor? Does the service provider provide information about the identity of the subcontractor and its tasks?
- What is the confidentiality policy of the service provider in regard to its employees, partners, and subcontractors?
- Is there a special confidentiality or security policy for sensitive, confidential, personal, or other special kinds of data?
- Is the service provider accredited and/or is he audited on a systematic, regular, and independent basis by a third-party to demonstrate that he complies with his security, confidentiality, and privacy policies? Is such a certification or audit process documented and do you have access to information such as the certifying or audit body and the expiration date of the certification?

6. *Data location and cross-border data flows*
- Where is the location of the data (and their copies) while they are stored in cloud-based services? Do they comply with the location requirements that might be imposed on your organization's data by law, especially by applicable privacy law? If not, are you considering the use of encryption technologies before storing the data in the cloud?
- Will you be notified if the data location is moved outside your jurisdiction?

- Does the contract mention that the data stored in the cloud may be subject to disclosure orders by national or foreign security authorities? Will the provider inform you and ask for your consent before disclosure (if such information or consent is allowed by law)?
- What is the legal jurisdiction in which the agreement is enforced and how dispute settlement will be resolved?

7. *End of service: contract termination*

- What is the duration of the contract? In what circumstances and how can it be terminated? Will there be any prior notification before the termination of the contract?
- At the end of the contract, whatever the reason, do you have the warranty that your data will be restored in a usable and inter-operable format? What is the time, procedure, and cost of such a restitution? Does the provider provide assistance for the restitution?
- At the end of the contract, will you have the right to access the associated metadata generated by the system for recordkeeping and legal purposes, notably to demonstrate that the confidentiality, integrity, authenticity, and reliability of your data have not been altered during their storage in the cloud?
- At the end of the contract and after complete acknowledgement of the restitution of your data, will your data and associated metadata be immediately and permanently destroyed in a manner that prevents their reconstruction (see also section 3 on data retention and disposition)?

In this study, the authors analysed the available cloud providers' terms and agreements. In doing so, the authors did not enter into contract negotiations with individual cloud providers, which limits the findings of this study to what is available online, which typically include services deployed in public clouds. The recommendations provided in this article will assist records managers and archivists in assessing existing cloud provider contracts and identifying gaps, but they can also be used to customize a contract and supplement existing fee-based services.

At the same time, it should be acknowledged that adherence to a checklist may not completely ensure that the client of a cloud service is entering into an agreement that places them in full compliance with recordkeeping and legal needs, obligations, and requirements. As can be seen in the agreements looked at in section 3 of this article, service providers may offer terms related to a recordkeeping need but may differ in how that need is addressed and which party is protected most by the language used. As a result, clients will still need to actively engage in the agreement process since the need for recordkeeping is addressed to some degree by the agreement, but it may not mean that it is addressed as much as it should be for the security and well-being of the organization. Organizations that utilize a checklist in creating or choosing cloud agreements to enter into should use it as a guide for navigating recordkeeping needs in the cloud, but they should still conduct risk assessments for the terms of the

agreement to determine whether the terms offered are agreeable (Public Records Office Victoria 2012, 7).

Despite this precaution, the authors of this article still believe that a checklist is more useful to records managers and archivists than a model contract. While it is true that records managers and archivists generally strive to meet the same standards, differing legal frameworks and cultures, organizational contexts, capabilities, and risk appetites make model contract terms difficult to produce. Based on the research presented in this article, the authors have devised a checklist of issues that should be addressed in any cloud service contract. This list should be considered only as a draft, as additional research will be necessary to test the checklist and identify gaps or weaknesses that may exist within it.

Notes

1. Google Cloud Platform: Data Processing and Security Terms, https://developers.google.com/cloud/terms; Pathway Communications CloudPath, http://cloudpath.pathcom.com; GreenQloud, https://www.greenqloud.com.
2. See *Oxford v Moss*, [1979] 68 Cr App R 183.
3. Digital Millennium Copyright Act, Pub L 105-304.
5. Privacy Act, RSC 1985, c P-21; Personal Information Protection and Electronic Documents Act, SC 2000, c 5.
5. Access to Information Act, RSC 1985, c A-1.
6. Sarbanes-Oxley Act, Pub L 107-204, 116 Stat 745; Basel I (1988), Basel II (2004), and Basel III (2010) accords are a set of international recommendations for banking regulations issued by the Basel Committee on Banking Supervision.
7. Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act, Pub L 107-56, 115 Stat 272.
8. EC Directive 95/46 on the Protection of Individuals with Regard to the Processing of Personal Date and on the Free Movement of Such Data, [1995] OJ L281.
9. EC Decision 2002/2 pursuant to Directive 95/46/EC on the Adequate Protection of Personal Data Provided by the Canadian Personal Information Protection and Electronic Documents Act, [2002] OJ L002, 13.
10. EC Decision pursuant to Directive 95/46/EC on the Adequacy of the Protection Provided by the Safe Harbour Privacy Principles and Related Frequently Asked Questions Issued by the US Department of Commerce, [2000] OJ L215, 7.

References

ARMA International. 2013. *Generally Accepted Recordkeeping Principles*. http://www.arma.org/r2/generally-accepted-br-recordkeeping-principles.

Baset, Salman. 2012. "Cloud SLAs." *Operating Systems Review* 46 (2): 57–66. http://dx.doi.org/10.1145/2331576.2331586.

Bradshaw, Simon, Christopher Millard, and Ian Walden. 2011. "Contracts for Clouds: Comparison and Analysis of the Terms and Conditions of Cloud Computing Services." *International Journal of Law and Information Technology* 19 (3): 187–223. http://dx.doi.org/10.1093/ijlit/ear005.

Canellos, David. 2013. *Adopting the Cloud While Adhering to Domestic and Foreign Government Regulations*. http://www.safegov.org/2013/10/2/adopting-the-cloud-while-adhering-to-domestic-foreign-government-regulations.

European Commission. 2009. Model Requirements for the Management of Electronic Records. Brussels, Belgium.
——. 2014. Cloud Service Level Agreement Standardisation Guidelines. Brussels, Belgium.
Ferguson-Boucher, Kirsten, and Nicole Convery. 2011. ''Storing Information in the Cloud: A Research Project.'' *Journal of the Society of Archivists* 32 (2): 221–39. http://dx.doi.org/10.1080/00379816.2011.619693.
Goh, Elaine. 2014. ''Clear skies or cloudy forecast? Legal Challenges in the Management and Acquisition of Audiovisual Materials in the Cloud.'' *Records Management Journal* 24 (1): 56–73. http://dx.doi.org/10.1108/RMJ-01-2014-0001.
InterPARES 2 Project. 2015. *InterPARES 2 Dictionary*. http://interpares.org/ip2/display_file.cfm?doc=ip2_dictionary.pdf.
International Organization for Standardization (ISO). 2001. *ISO 15489-1*. http://www.wgarm.net/ccarm/docs-repository/doc/doc402817.PDF.
——. 2012. *ISO 14721*. http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284.
——. 2013. *ISO/IEC NP 19086*. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=63902.
Klein, Kris. 2008. *Applying Canadian Privacy Law to Transborder Flows of Personal Information from Canada to the United States: A Clarification*. Industry Canada. https://www.ic.gc.ca/eic/site/ecic-ceac.nsf/eng/gv00508.html.
Office of the Privacy Commissioner of Canada. 2005. *Bank's Notification to Customers Triggers PATRIOT Act Concerns*. PIPEDA Case Summary no. 2005–313. https://www.priv.gc.ca/cf-dc/2005/313_20051019_e.asp.
——. 2009. *Processing Personal Data across Borders. Guidelines*. https://www.priv.gc.ca/information/guide/2009/gl_dab_090127_e.pdf.
Public Records Office Victoria. 2012. *Cloud Computing: Implications for Records Management, V.1.0*. State of Victoria, Australia. http://prov.vic.gov.au/wp-content/uploads/2012/04/Issues-Paper-Cloud-Computing.pdf.
——. 2013. *Cloud Computing Decision Making, V. 1.0*. State of Victoria, Australia. http://www.unimelb.edu.au/unisec/privacy/pdf/PROVCloud_Computing_Guideline_1.pdf.
Reed, Chris. 2010. ''Information 'Ownership' in the Cloud.'' Legal Studies Research Paper no 45. School of Law, Queen Mary University of London. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1562461.
Van Hoboken, Joris, Axel Anrbak, and Nico Van Eijk. 2012. *Cloud Computing in Higher Education and Research Institutions and the USA Patriot Act*. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2181534.
Vermeys, Nicolas, Julie M. Gauthier, and Sarit Mizrahi. 2014. ''Étude sur les incidences juridiques de l'utilisation de l'infonuagique par le Gouvernement du Québec.'' Working paper. Laboratoire de cyberjustice, Université de Montréal. http://www.cyberjustice.ca/wordpress/wp-content/uploads/2014/08/%C3%89tude-sur-les-incidences-juridiques-de-lutilisation-de-linfonuagique-par-le-gouvernement-du-Qu%C3%A9bec.pdf.

Through a Records Management Lens: Creating a Framework for Trust in Open Government and Open Government Information

Les objectifs visés par les systèmes de gestion documentaires : La mise en place d'un cadre de confiance et de la transparence de l'information dans un gouvernement ouvert

Valerie Léveillé
School of Library, Archival and Information Studies, University of British Columbia
valerieleve@gmail.com

Katherine Timms
Information Standards Specialist, Library and Archives Canada
katherine.timms@bac-lac.gc.ca

Abstract: Through an analysis of current business processes, workflows, and documentation that guide the creation, management, and control of open government information as well as the policies, procedures, and structures in place that help instruct these processes and establish open government initiatives, this article will offer a preliminary exploration of the possibility of establishing a universal framework around these initiatives that would help ensure that the information being distributed is accurate, authentic, and trustworthy. A preliminary investigation of open government initiatives in Canadian jurisdictions represents the first phase of research to situate and explore this discussion in context.

Keywords: open government, open data, open government data, open information, records management, business process

Résumé : En s'appuyant sur une analyse des processus d'affaires courants, des flux de travail et de la documentation qui guident la mise en place, la gestion et le contrôle de l'information dans un gouvernement ouvert, ainsi que les politiques, les procédures et les structures qui sont en place pour aider à mettre en marche ces processus et à établir des initiatives gouvernementales ouvertes, cet article offre une exploration préliminaire de la possibilité d'établir un cadre universel concernant ces initiatives, cadre qui aiderait à assurer que l'information distribuée est exacte, authentique et digne de confiance. Une enquête préliminaire d'initiatives de gouvernement ouvert dans les administrations canadiennes constitue la première phase de la recherche visant à situer et à explorer cette discussion dans son contexte.

Mots-clés : gouvernement ouvert, données ouvertes, données gouvernementales ouvertes, information ouverte, gestion documentaire, processus d'affaires

La Revue canadienne des sciences de l'information et de bibliothéconomie 39, no. 2 2015

Introduction

Open government has become widespread within modern bureaucracies. The advent of modern technologies, social media platforms, and the like have played a significant role in creating a sense of accessibility between citizen and government, which in turn has led to an influx of public demand for access to government information. While, for some jurisdictions, open government is used as a means to respond to this influx, it is primarily representative of a shift towards greater government transparency and public accountability. By encouraging an open dialogue between their administration and its constituents, governments hope to meet the ever-increasing challenge of establishing trust between it and its citizens. However, this trust can arguably only be cultivated under strict conditions, which include the dissemination of complete, accurate, and reliable open government information that has been generated in a trusted records management environment.

The premise of this research is reflected in the objectives of the InterPARES Trust (ITrust) project, an international research collaboration that seeks to

> generate theoretical and methodological frameworks to develop local, national and international policies, procedures, regulations, standards and legislation, in order to ensure public trust grounded on evidence of good governance, a strong digital economy, and a persistent digital memory (InterPARES Trust 2014).

As a product of this research, this article will explore concepts of trust in open government initiatives and the open government information (for example, government data and data sets, open information, metadata components, and so on) that support these initiatives by touching on the records management concerns that underlie them both. The first section of the article will begin by defining the concepts of open government and open government data as individual but complimentary entities and conclude by exploring examples of other types of open government initiatives, including those that are currently being adopted across various Canadian jurisdictions. The second section of this article will then refer to the analysis of business processes as an approach to exploring the impact that these initiatives are having on the creation, management, and control of government records. It will follow by offering guidance on how these challenges may be reconciled through the adoption of an enterprise-level perspective. The Canadian examples will be revisited at the end of this section, with some preliminary observations given on how they relate to the records management framework just explored.

This article offers an introductory exploration of the issues pertaining to the intersection of open government initiatives and records management requirements.[1] Outlined in the form of a literature review and preliminary analysis, it presents a summary of the issues and ideas that have been explored to date within the parameters of this research. It must be acknowledged, however, that while this article proposes one way of addressing these issues, it is most certainly not the only way that records professionals may be choosing to approach this matter. As such, the next phases of this research will not only serve as a means

of verifying the effectiveness of the approach suggested in this article but also explore additional strategies that are being adopted by those currently pursuing open government initiatives.

Open government, open government information, and the Canadian context

Open government and open government information have become increasingly important tools for achieving "democratic accountability and deliberation" in government (Janssen, Charalabidis, and Zuiderwijk 2012, 260). While there is a convergence between open government and open government data, the latter of which is one of several examples of what can constitute "open government information," the two are distinct concepts with separate objectives and backgrounds. To properly understand the records management requirements that are attributable to these concepts, their underlying features and characteristics must be understood and properly contextualized. For this purpose, this section of the article will also explore various Canadian jurisdictions that currently support open government initiatives.

Open government

In its simplest form, open government means increasing public access to government information through modern digital technologies. It is often interpreted as a contemporary extension of e-government or government 2.0, with a focus shifting away from the delivery of government information and services exclusively via communications technology and towards the delivery of government information via innovative technological platforms. The overarching goal of open government is to create a sense of openness, sharing, and collaboration between different government departments and between government and the public (Francoli 2011, 152–53).

Beyond e-government, the fundamental principles that ground the concept of open government share strong ties with those that underpinned the early "right-to-know" or right to information movements that eventually led to the creation of access to information legislation. Prior to these specific efforts, similar beliefs contributed to the rise of related movements in support of freedom of expression and freedom of the press.[2] At the federal level, access-to-information (ATI) legislation was developed as early as 1966 in the United States. It was intended as a way of not only increasing access to information on government activities and procedures but also reducing corruption, malfeasance, and bribery inside government.[3] Canada's first Access to Information Act was later introduced in 1983.[4] As a countermeasure to the adverse consequences that ATI-like legislation risked having on the privacy of the average citizen, Privacy Acts were introduced shortly thereafter as a means for protecting personal information that could be found in government records.[5] The advent of the Internet, social media, and innovations in communication technologies have each had significant repercussions on amplifying demands for ATI, many of which government and legislation have had difficulty keeping up with. While it may not be

based on new ideas, open government is reshaping the way government now approaches ATI and, thus, its relationship with the public.

An open government strategy aims to increase government transparency by giving citizens access to both public sector information (that is, publicly funded information that should be publically available) and information that informs the public, to a reasonable degree, on government processes, activities, and procedures. Increased transparency should not only result in a surge of the amount of information available to citizens about their community, but this information should also empower citizens to hold their government to account (Francoli 2011, 154; O'Hara 2012, 226; Ubaldi 2013, 13). Accountability can only be truly achieved when, first, citizens are given free, unrestricted, and unbiased access to information that permits them to hold their government to account and, second, when that government is willing to accept being held accountable, taking responsibility for any failures, losses, or shortcomings that are brought to light as a result of the release of that information (Veljković, Bogdanović-Dinić, and Stoimenov 2014, 279). True accountability, in turn, encourages citizen engagement and participation, a third objective of open government. An open government initiative is unique in that it seeks to achieve true democratization of knowledge creation and dissemination as a way of encouraging effective public oversight of government activities by civil society. In this regard, it also aims to enable citizens to influence government service development and public policy drafting (Scassa 2014, 398). This action ultimately supports the long-term objective and benefit of open government: enhancing, building, and nurturing, if not, in some cases, renewing, a trust relationship between a government and its citizens.[6]

With conceptual roots in ATI legislation, open government has traditionally adopted a more reactive, rather than proactive, approach to information dissemination—that is, only once the information is requested will it then be released, providing that the information is eligible for disclosure and distribution within the limits of the law. However, this particular strategy has often made government a target of significant criticism from those who doubt the true intentions of open government strategies. Critics will regard such an approach as a tactic that helps government increase, rather than give up control over public information. They would argue that this results in the reinforcement of existing structures and demonstrates a further resistance to change toward a more democratically accountable model of government (Janssen, Charalabidis, and Zuiderwijk 2012, 266). This stance is slowly shifting, however, with governments now creating extensions of their open government initiatives to include different platforms and tools (for example, open data portals, forums for open dialogue)—to be explored in the next sections of this article—that allow for the proactive or routine distribution of specific types of government information. Ultimately, transparency and accountability can only truly be achieved when, first, the gap that exists between a proactive and a reactive approach to the release of information is narrowed and, second, when a governance model that supports more routine dissemination of information is adopted. Ideally, the concept of an

"open" government recognizes the public's right to information that belongs in the public domain or that has been gathered as a result of public funds, and, as such, it supports the open and unrestricted distribution of this information and encourages its reuse by the public. In turn, this supports the objectives of open government: transparency, accountability, and citizen engagement.

Open government data

Open government is strictly a concept, an ideology. When put into practice, a government's commitment to such a concept is usually reflected in the form of policy and action plans and set in motion via specific tools, platforms, or initiatives. While various governing bodies will approach this task differently depending on the availability of staff, time, and resources, open data and open data portals are nevertheless popular examples of tools commonly used to support the opening up of government information. To understand how these tools are used, one must first understand what they encompass.

There are varying definitions attributed to "open data." Simply put, data, or "the smallest meaningful units of information," become by definition "open" when they can be "freely used, re-used and redistributed by anyone, only subject to (at most) the requirements that users attribute the data and that they make their work available to be shared as well" (InterPARES 2 Project n.d.; Ubaldi 2013, 6). This definition is attributable to many different data types and is a concept that is reused and adopted across many different communities, both public and private, that support the free distribution of open information, which can include scientific data, environmental/meteorological data, mathematic data, or government data. The concept of open data falls under an umbrella of similar "open" movements, including open access, open source, and open hardware. These movements all have the common goal of generating open knowledge—that is, "any content, information or data that people are free to use, re-use and redistribute—without any legal, technological or social restriction" (Open Knowledge 2014).

There are four characteristics that are attributable to open data.[7] First, users must be able to access the data easily and effortlessly in a machine-readable format and via a web-based interface and platform that will not impose any technical or educational barriers. Second, data must be distributed and made available in a format that will allow users to reuse it, including manipulate it or "mash" it up with other data sets. Third, open data must account for quantity. In order for it to carry meaning and be of value to its user community, the data must be distributed in large quantities, either in numbers (that is, large amounts of data on one topic), in time (data accumulated over an extended period), or in total amount or size. Lastly, open data must be absent of membership, bias, exclusivity, or special privilege. It must account for universal participation by excluding such restrictions or controls as unwarranted licenses, copyright restrictions, patents, trademarks, and, as much as possible, charges for data access and reuse.[8] The concept of free distribution, from both a legal and technological standpoint, also compels a legal component to the definition of open data, thus

requiring distributing parties to apply an appropriate open license that will allow for free and fair distribution of the data. This article, however, will focus specifically on open government data, which distinguishes itself from other open data types as "any data or information produced or commissioned by public bodies" (Ubaldi 2013, 6), and, in this case, produced or commissioned by a government body.

Opening up government data is the practice of identifying government data and/or information that is public in nature and making this data available to the public so that it can be reused and redistributed for purposes other than those for which the information was originally compiled. Unlike ATI, publications of open government data, usually in the form of structured data sets, are normally done on a more proactive or routine basis and are released onto a web-based data portal that eases discovery and user interaction. These data types are varied and can include information on demographics, health, and safety, geographic and ecological issues, and/or financial details, among many other types. The make-up of the government data user community is equally as diverse—users stem from both public and private sectors, academia, and civic organizations and can consist of information technology (IT) professionals/developers, entrepreneurs, advocates, and individual citizens alike (Ubaldi 2013, 11). By processing, mashing, and distributing these data sets, these users play an integral role in generating value from this data.[9] Whether this value translates into the development of new public policy, the creation of new products and services, or economic growth, organizations inevitably seek to generate a type of return on investment when opening up government information (Commission to the European Parliament, the Council, the European Economic and Social Committee, and the Committee of the Regions 2011; Janssen, Charalabidis, and Zuiderwijk 2012, 260). Above all, open government data are a tool used to help achieve the ultimate goal of any open government strategy: to establish a trust relationship between the government and citizens.

Beyond open data platforms: other types of open government initiatives

This article's focus with regard to open data will expand beyond open government data and open data platforms. It will aim to encompass many forms of open government information, including open government data and datasets, which are important and valuable components of open government information, as well as open information (for example, statements of ministerial expenses, completed ATI requests), metadata components, and so on. Today, open government strategies, the open government information they distribute, and the different platforms through which this information is disseminated can take many forms:

- developing web-based applications that help improve the delivery of information and government services;
- modernizing ATI-like legislation;
- designating specific government information for routine release (for example, expense reports, awarded tenders or contracts, and so on) through an "open information" portal, including summaries of ATI requests once completed;

- creating and using social media-like platforms that enable and encourage an open dialogue between government and citizens; and
- adopting an "open by default" model for the release of government information and data sets moving forward (Scassa 2014, 399).

As part of their national action plan, for example, the US government has committed to the modernization of its government records and records management system and is currently working toward declassifying its national security information. The Swedish government is working to improve its Openaid.se platform and the full implementation of the International Aid Transparency Initiative standard.[10] The British government, deemed by advocates as being a notable leader in the open government movement, will release an inventory list of all of their data sets, both published and unpublished, and has launched a public sector information (PSI) directive that will transpose the reuse of PSI into UK law (Open Government Partnership 2014a, 2014b, 2014c). The examples demonstrate that a successful open government initiative must contain several different strategies and that a government's commitment to these should be consistently demonstrated and routinely assessed, evaluated, and updated as needed. The next section of the article will explore how several Canadian governments are approaching the open government challenge.

Open government initiatives in Canada

There are several government-lead open government initiatives currently underway in Canada at the federal, provincial/territorial and municipal levels.[11] These initiatives exist at different degrees of maturity in different jurisdictions. This section provides an overview of some of these initiatives based on readily available, publicly disclosed information sources.[12] This overview is intended to provide a high-level, bird's-eye view of several Canadian open government initiatives, representing a starting point in the research of records management issues in these contexts. Some preliminary observations of records management issues based on the examples in this overview will be provided in second section of this article.

Federal level

Canada's Action Plan on Open Government describes commitments made by the federal government to support open government. The two foundational commitments were a directive on open government (published October 2014) and an open government license (Canada 2012). The directive provides guidance to departments with regard to making more information and data available online (Canada 2014a). The license serves as one universal license, with the intent to reduce administrative overhead and restrictions on reusing published government information and data (Canada 2014g).

Additional activities in the action plan are organized within three activity streams:

1. *Open information*
- Activities include enabling easier access to ATI request summaries; creating a virtual library of published government documents; removing restrictions on historical records held by Library and Archives Canada wherever possible; advancing the government-wide development and implementation of its electronic document and records management solution, GCDOCS; and consolidating the government's web presence.

2. *Open data*
- The open data portal is a central catalogue of government open data sets of all kinds, including geospatial and statistical data sets.[13] Individual government departments and agencies remain responsible for hosting their own data sets, which are linked to the open data catalogue. The descriptions in the catalogue adhere to the open data portal metadata standard (Canada 2014b).[14]
- Additional online federal data resources include several geospatial sites such as Geogratis, GeoConnections Discovery Portal, and Atlas of Canada (Canada 2014d; Natural Resources Canada 2012, 2015).

3. *Open dialogue*
- Activities include a web 2.0 citizen engagement platform for use in public consultations and requirements for federal regulators to post their upcoming regulatory plans to give Canadians and businesses advance notice of upcoming changes.

Canada published an action plan in the fall of 2014, which was informed through consultations with Canadians and built upon original commitments. In the international arena, Canada officially joined the Open Government Partnership (OGP) in March 2012. The OGP is an international initiative working to support commitments by governments to improve transparency, citizen engagement, and ATI.[15] To participate, a country must devise a national action plan supporting open government principles and goals and be responsive to feedback from the OGP independent reporting mechanism, which carries out biannual reviews of the progress of participating OGP countries. Countries may also participate on OGP working groups, and Canada co-leads the Open Data Working Group. Canada also adopted the G8 Open Data Charter in 2013, which identifies five essential principles underpinning open data initiatives: data are open by default insofar as possible; data of high quality and quantity should be released; re-usable data should be released; data for improved governance should be released; and data for innovation should be released (Canada 2014c).

Provincial/territorial level

British Columbia

British Columbia's open government initiative is managed by the Ministry of Technology, Innovation and Citizen's Services. The Open Information and Open Data Policy provides direction and assigns responsibility for releasing

open government information under the control of government ministries (British Columbia, Office of the Chief Information Officer 2011). The open information component of this policy focuses on the proactive disclosure of responses to ATI requests under freedom of information legislation as well as designating other government information for routine release, while the open data component focuses on the means to assess, approve, and post open data for public use, adaptation, and distribution. The policy abides by the Freedom of Information and Protection of Privacy Act, and the release of data must be in accordance with the province's Open Government License (British Columbia 2014a).[16] Tools supplementing the policy include an open data assessment form and the open data physical format dataset standard. There are two principal websites:

1. *Open Information*. This website is where information required by the Open Information and Open Data Policy is posted by the BC government ministries (British Columbia 2014b). It primarily consists of information release summaries and travel expenses for ministers and deputy ministers.
2. *DataBC*. DataBC comprises several components and services: a warehouse or central repository in which data sets hosted by DataBC are kept; a catalogue of metadata describing all data sets hosted by, or linked from, DataBC; several tools for web syndication, data integration, data visualization and analysis, and search; and the DataBC website, which is the main mechanism for delivering DataBC services (British Columbia 2014c). Contributing ministries have the option of hosting their data sets themselves or having DataBC host them on their behalf.

Data management concepts, governance, and operations are well developed in this context. Managing government data is seen as an enterprise-level effort. Open data are recognized to be a subset of enterprise data, which is data that can be shared within the government, which itself is a subset of all operational data (some of which cannot be shared outside of the creating ministry). DataBC's processes help determine which enterprise-level data sets may be candidates for release as open data, if they lack legal, security, and privacy concerns.[17]

In July 2013, the Office of the Information and Privacy Commissioner for British Columbia produced an investigation report entitled *Evaluating the Government of British Columbia's Open Government Initiative* (Denham 2013). In the report, the commissioner provided concrete recommendations for building on the foundational open government program to better meet its core objectives, namely increasing government transparency and accountability:

- expanding the scope of government information released, potentially to include calendars, contracts, audit reports, and so on, which would be supported through ministries establishing consistent categories of records for proactive disclosure;
- expanding on the existing DataBC program to include more outreach and data literacy activities and to provide access to data sets identified as high-value; and

- ensuring sustainability of the program through adopting access-by-design principles to make government information open by default[18] and by modernizing the archives and records management statute.[19]

Alberta

Alberta's open government program was launched in 2012 as part of Service Alberta, the ministry responsible for helping deliver services to other government ministries and to Albertans. The following year, the government published an *Open Government Action Plan*, detailing how it was committed "to being more accountable and transparent to the citizens of Alberta" (Alberta 2013c, 4). The main components of its approach are identified as releasing more government information, improving interactions with citizens, and encouraging and facilitating greater public engagement with the government. The action plan identifies activities in relation to three themes:

1. *Enabling change*. It devises a strategic plan and vision; develops a governance framework, to be led by Service Alberta; devises a government of Alberta agreement confirming the commitment of different ministries to support the open government program; and develops a cultural change plan for public servants to clarify their roles and responsibilities to open government.
2. *Informing Albertans*. It launches open data and open information portals to increase public access, with the Alberta government library leading the latter; engages the development community in innovation competitions to work with government open data; and establishes a routine information disclosure program.
3. *Better conversations*. It develops a communication plan with the help of the Public Affairs Bureau and communications teams across the government of Alberta; develops a public education toolkit; and devises a citizen participation plan.

Since the release of this plan, Alberta has launched an open data portal, which consists of metadata descriptions of data sets contributed by Alberta ministries with links to external data sets (Alberta 2012a). Data are released in accordance with an open government license (Alberta 2012c). The *Open Data Set Publishing Guidelines* identifies a set of three documents geared toward assisting ministries with contributing to the portal, specifically: a value framework to help with the evaluation of data sets with an accompanying evaluation form; and an assessment checklist form which verifies that all release criteria have been satisfied and must be signed by the appropriate authority before posting the data set on the portal (Alberta 2014b).[20]

Alberta has also developed an Open Data and Open Information Policy, the purpose of which is "to provide direction and assign responsibility for a single approach to providing Government of Alberta information and data for public use, adaptation and distribution under the Open Government Licence" (Alberta 2012b).[21] The policy's guiding principles are open by design, innova-

tion from quality data, and improved governance. The policy also includes instructions pertaining to proactive disclosure of government information and considerations for balancing access and privacy in accordance with freedom of information legislation.[22] Supplementing this policy are open data standards that specify: publication criteria and obligations by government ministries; required data characteristics; the mandatory usage of the Open Data Metadata Application Profile; allowable formats; and assessment metrics (Alberta n.d.).

Ontario

Ontario is actively developing its open government program. Its government has established an open government web presence and provides an open data service (Ontario 2014a, 2014c). The Ontario open data portal contains both descriptions of the data sets and the data sets themselves. The data sets originate from different government ministries and pertain to different subject areas, including environment, education, travel, taxes, business, arts, health, employment, and others.

Ontario appointed an open government engagement team, the mandate of which was to provide advice to the minister of government services on the development and implementation of an open government initiative in Ontario. Incorporating the results of public consultations, this team produced a report, *Open by Default: A New Way Forward for Ontario*, which summarized recommendations for future directions, after which the team disbanded at the end of March 2014 (Ontario 2014b). The recommendations of this report were organized in four topical areas:

1. *Working together:* to focus on engaging the public in government and supporting their participation;
2. *Opening up government information*: focusing on improving the freedom of information framework; publishing inventories of personal information collected by the government not to be released for privacy, security, or other legal reasons; proactively publishing government planning documents; and publishing the results of the legislative process in an open format;
3. *Making data a public asset*: focusing on implementing an open-by-default data policy; supporting by-design principles by ensuring that newly procured IT systems support open data; integrating Ontario's open data portal with the broader IT enterprise infrastructure to support more holistic data management; developing partnerships to foster innovative reuse of public data; and developing new data sets to support key social and economic needs at the local, regional and provincial level; and
4. *Implementation and sustainability*: focusing on assigning responsibility for the open government portfolio to a senior minister within Cabinet; adding two permanent groups to the open government secretariat, namely a public engagement unit and a digital centre of expertise; and developing metrics to assess government progress on its open government initiative.

Newfoundland and Labrador

Newfoundland and Labrador launched an open government initiative in March 2014. Its stated objectives are to "improve access to government information and data; enhance government's overall engagement of citizens and stakeholders; and strengthen collaboration between and among all sectors including government" (Newfoundland and Labrador 2014f). The initiative is led by the Office of Public Engagement and is built on the following four pillars, for each of which a website was launched in March:

1. *Open information*. The proactive release of government information is provided, including responses to access to information requests, orders in council, ministerial expense claims, member accountability and disclosure reports, and awarded tenders. Additional information from various departments is available, organized by topic (Newfoundland and Labrador 2014g).
2. *Open data*. Tabular and spatial data sets are provided by government departments and agencies and are processed by the Newfoundland and Labrador Statistics Agency before publication on the website, accompanied by a metadata description. Some data visualization applications are also available (Newfoundland and Labrador 2014e).[23] The open government license is applied (Newfoundland and Labrador, 2014d).
3. *Dialogue*. This website contains materials pertaining to the public engagement plan, including results from consultations and feedback from presentations. The Office of Public Engagement is tasked with collaboratively building Newfoundland and Labrador's first open government action plan, and this website is part of the collaboration mechanism (Newfoundland and Labrador 2014c).
4. *Collaboration*. This website provides details on collaborative arrangements between the government of Newfoundland and Labrador with other governments, organizations, and communities (Newfoundland and Labrador 2014a).

Other provincial/territorial jurisdictions

Open government initiatives in other provincial/territorial jurisdictions are arguably not as developed as those already discussed. For instance, no province-led initiatives in Saskatchewan have been initiated. Further, in many provinces, government data initiatives have focused solely on geospatial data.[24] Quebec has declared its intent to work toward becoming an open government (Québec 2014a), and additional plans were proposed in a report commissioned by the Quebec government, one of which was an open data portal (Gautrin 2012). The portal, since it was launched, contains both metadata about data sets as well as the data sets themselves (Quebec 2014b). The data sets are released in accordance with an open government license, which was recently replaced by a creative commons license adopted jointly by the cities of Quebec, Montreal, Gatineau, and Sherbrooke (Quebec 2014c, 2014d).

Municipal level

Many Canadian municipalities have developed open government initiatives—in particular, open data components.[25] The informal partnership formed in 2011 known as G4—comprised of Toronto, Ontario; Vancouver, British Columbia; Edmonton, Alberta; and Ottawa, Ontario—is working collaboratively to improve open data standards and practices (Giggey 2012). Further, the Ontario provincial government has formed a public sector open data working group comprised of representatives from various municipalities and other partners (for example, the MaRS Discovery District), tasked with developing common processes and formats.[26] It is beyond the scope of this article to address all municipal open government and open data initiatives, so the following three municipalities will serve as examples.

City of Toronto

Toronto issued its *Open Data Policy* in 2012, outlining the principles, roles, and responsibilities related to the city's open data program, which supports its commitment to open government and is focused on making data publicly available in reusable formats (Toronto 2012). The program is implemented as a component of the city's enterprise information management initiative, codified in the *Information Management Framework* (Toronto 2013a). This framework embodies a standards-based approach to information management based on four principles: accountability, openness, lifecycle management, and trust and reliability. By aligning with the *Information Management Framework*, Toronto has situated open government data as one element within enterprise information management, thus taking a holistic or integrated approach.

The *Open Data Policy* also aligns with the access by design and privacy by design principles, developed by the Office of the Information and Privacy Commissioner of Ontario (2014a, 2014b). Data sets released comply with rights of privacy, security, and confidentiality as identified in the Municipal Freedom of Information and Protection of Privacy Act and other applicable legislation and are released in accordance with the Open Government License via Toronto's open data catalogue (Toronto 2013b, 2014a).[27]

In the policy, executives are responsible for identifying data sets for release, along with related planning and assessment exercises, working with Toronto's Open Data Team on the data set publication process. Toronto's Open Government Committee is tasked with providing governance and oversight for the open data program. The Open Data Team includes staff from the City Clerk's office and the Information and Technology Division and is mandated "to assess, prioritize, release and monitor datasets in accordance with this [open data] policy" (Toronto 2012).

Toronto's twenty-six strategic objectives for 2013–18 include Strategic Objective 13 on Open Government by Design, the focus of which is a shift in organizational culture to support more accountable, open and transparent government (Toronto 2013c, 2013d). Like the *Open Data Policy*, this strategic action is framed as being explicitly in alignment with the *Information Manage-*

ment Framework and by-design principles, specifying the need for integration of technology and information management at all levels. Training and awareness efforts are identified as an important and necessary part of fulfilling this objective, along with devising assessment mechanisms and metrics to evaluate progress. By situating open government as an information management initiative, information management is clearly recognized as the foundation and enabler of open government.

City of Vancouver

In 2009, Vancouver City Council passed a motion supporting the principles of open and accessible data, open standards, and open source software (Vancouver 2009). The city resolved to actively pursue an open data program, including developing plans to release archival data as open data. The city's open data website, launched in September 2009, includes both metadata descriptions of data sets as well as the data sets themselves (Vancouver 2014b). Also included is information about available data formats and the licenses under which the data sets are released.

Vancouver developed a *Digital Strategy* in 2013, which included not only mention of open government data but also e-government, citizen engagement, and digital infrastructure to support socio-economic development (Vancouver 2013). The high-priority initiatives identified in this strategy included expanding on the open data program, improving digital service delivery, and establishing digital services governance. These initiatives were all intended to support the four pillars and goals of the strategy: engagement and access; economy; infrastructure and assets; and organizational digital maturity. The strategy makes clear reference to its support of the "open government ecosystem" (Vancouver 2013, 3; Vancouver Chief Librarian 2013). This strategy is cited as a priority in Vancouver's 2014 *Corporate Business Plan* (Vancouver 2014a).

Vancouver released an online public engagement tool, "Talk Vancouver," in 2013 (Vancouver 2014c). The tool enables registered users, who must be residents of Vancouver over the age of fifteen, to participate in discussions concerning municipal affairs. It aligns with the open government objective of citizen participation and engagement in government.

City of Regina

Regina's open government initiative is based on the following three pillars:

1. *Open data*. Regina's open data catalogue includes both metadata about the data sets as well as the data sets themselves (Regina 2014b). Its technical platform, OGDI DataLab, is an open-source open data catalogue that is targeted for Microsoft's Windows Azure cloud-computing platform, and it includes functionality supporting both human and machine interaction with the data (Regina 2014a). Use of the data is subject to the city's licensing terms (Regina 2015).

2. *Open Information*. This is an online repository of government documents (for example, reports, expenses, surveys, and so on) proactively disclosed by the city or requested under the Local Authority Freedom of Information and Protection of Privacy Act (Regina 2014d).[28] The stated intention is to maintain the information disclosed on the open information page for a minimum of two years post-release.
3. *Citizen interaction*. Regina's open government website also includes mention of applications built from the city's open data and an invitation to follow city communications on various social media platforms (for example, YouTube,™ Facebook,™ and Twitter™) (Regina 2014c).

This open government initiative, launched in 2012, originated from an internal corporate information assessment that emphasized the need for external stakeholders to access city information. It is managed by a cross-organizational team within the city, including two open government committees, the Strategy Committee and the Operational Committee, which provide direction and governance for the program. The city has also worked in collaboration with open data citizen-lead groups, Open Data Saskatchewan and HackRegina (Currie 2013, 119–24).[29]

Common components of open government initiatives in Canada

Based on this overview, the following common components are identifiable in these open government initiatives:

- *Open government plans*. Some jurisdictions have well-developed and comprehensive planning documents for open government initiatives. In others, commissioned or investigative reports have provided advice, feedback, and direction for further developing the initiatives.
- *Legislation, policy, procedures, and guidelines*. In addition to freedom of information and privacy legislation, some jurisdictions have comprehensive policy and procedural documents specific to open government (including open information and open data), while others do not or are still in the process of drafting them.
- *Open data portal or catalogue*. Many jurisdictions offer some kind of open government data service, for example, a metadata registry (catalogue) and/or a repository of data sets. The types of data sets offered vary per initiative; in some jurisdictions, only geospatial data are available.
- *Open information*. Some jurisdictions have websites dedicated to providing access to open information resources, separate from their open data portals.
- *Open dialogue or citizen engagement*. Some jurisdictions have specific program lines dedicated to open dialogue or citizen engagement. This may take the form of public consultations, websites intended to collate citizen input on issues, projects or initiatives, or other mechanisms.

While some jurisdictions perceive open government information services as components of their information management infrastructure and responsibilities, this is not yet a common approach. Issues with this and other records

management implications of open government initiatives will be explored in the following section.

Open government within a records management framework

Records management plays a deciding role in creating reliable and trustworthy open government information. When records are well managed, they serve as instruments of accountability and as authoritative and trusted sources of information about government activities (Thurston 2012b). Underpinning an open government initiative with an already strong and reliable records management framework is therefore likely to have a considerable impact on the success of the initiative. Since an open government initiative introduces new processes, procedures, records, and, thus, a new series of records management challenges for government, these must be identified and the proper controls must then be introduced, first, within the framework of the initiative and then in relation to the organization's goals and strategies. As this section of the article will explore, this can be accomplished by conducting a work process analysis of the initiative, which can then help inform the records management requirements for records creation, capture, and control.

Open government as a business process

As described in the International Standardization Organization's *Technical Report: Information and Documentation—Work Process Analysis for Records* (2008), a business process analysis is an efficient way of identifying the systematic work processes, transactions, and records that comprise a business process or, in this case, an open government initiative. Such an approach includes both an analysis of functions being performed in a business context (functional analysis) and an analysis of the constituent steps within the business process, in which records are generated at the transaction level (sequential analysis). The former includes an assessment of the broader context of the business process, which includes its mandate and its regulatory environment. A functional analysis helps identify reasons why or for what purpose the work supported by the process is undertaken. Thus, performing a functional analysis of an open government initiative would first place it within its respective organizational context (that is, the governing body and its jurisdiction) and then identify the ultimate goals of the initiative (that is, achieving transparency and accountability by opening up government information). This type of analysis helps inform higher-level information management and records control issues—for example, information classification and retention/disposition. In another instance, a sequential analysis focuses on how the work is undertaken, illuminating records creation requirements at the transaction level as well as the roles, responsibilities, and dependencies between related processes (for example, when the output of one process is required as input for another). Considering the roles of various participants as part of this analysis further helps demarcate the specific sequence of steps (for example, providing guidance; providing approval or authorization; undertaking processing; undertaking evaluation or audit; and so on). Understanding these roles not only highlights

what types of records are generated through the process but also who is responsible for them.

This article divides and summarizes an open government business process into three categories—that is, three stages of implementation: initiation; identification and distribution; and promotion and evaluation. Each stage consists of tasks and transactions, summarized as components, that are key to the initiative's success. Each work process produces a series of information objects (that is, records and documents) that must be considered both for their purpose as evidence supporting the workflow as well as for their value as attributes of accountability in relation to the open government initiative. Only then can records management requirements for records creation and control be properly informed.

Initiation

Policy

A government's commitment to transparency and accountability must be demonstrated, first and foremost, at an enterprise level. The objectives and principles underlying the open government initiative must be accurately reflected in an open government policy. In general, an open government policy should help guide decision-making and outline the desired outcomes, deliverables, and courses of actions. With regard to the publication of open government information, a policy can serve to expand on the types of government information that can and should be distributed under the policy's intent as well as guide the development and dissemination of these outputs (for example, by specifying in what format and through which platforms the information is distributed). In turn, this can increase data accessibility and encourage the information's reuse. While a commitment to open government comprises much more than a policy, it is nevertheless a key document that demonstrates a first step in guiding a new initiative's implementation and oversight. It should state the intentions, duties, and proper authorities that will be given to those designated as being responsible for the initiative as well as outline the metrics that will be used to evaluate the program's progress once implemented. Most importantly, a policy should be a statement of accountability; a government is accountable to its citizens, but it must also be accountable to itself.

An open government policy should be designed in line with an organization's existing policies, legislation, and regulations, specifically those supporting the organization's accountability framework and records management practices. If properly designed, an open government initiative will likely influence (and hopefully strengthen) a government's existing accountability structures and, within these, its records management policies. With regard to the latter, an open government policy should include components that address records management, if not at least a cross-reference to an existing records management policy. If neither is the case, guidance should at least be provided by those responsible for the policy's oversight on how the two policies could be brought

into alignment. It is important to define a clear relationship between open government and records management practices and, as a consequence, between the records, the open government initiative, and its processes. Citizens must be able to base their trust on a government's ability to comply with both old and new standards of practice. Compliance with the provisions outlined in these new policies should not be realized at the expense of existing official standards (McDonald 2012, 13). If such a policy is to carry the same weight and authority as other enterprise-wide policies generated by the organization, then its development would follow the same drafting, editing, review, and approval processes that were used for these other policies.

Action plans, guidelines, and standards of practice are examples of information objects that are likely to accompany the creation of an open government policy. These, along with other similar documentation, will help create consistency in government actions and uniformity in outcomes. Creating procedures on the release of open government data sets, for example, would help standardize the format in which these data are released, creating consistency in output and enabling seamless interoperability between various data sets. Standard templates, which would capture different types of contextual information in the form of metadata,[30] is one example of an information object that could be required as part of the release of this information. It would help to ensure not only that information is properly contextualized but also that data are traceable and linkable, which are "core elements for dataset authenticity and reliability" (Thurston 2012b). Above all, the planning, processes, procedures, and document trail that underlie the drafting of this policy and its supporting documentation are as important to the open government initiative as the policy itself: "Each public body should open up as much data about the preparation and execution of policy-making as possible, in an accessible and understandable way" (Zuiderwijk, Janssen, Cheonni, Meijer 2014, 3).

Players

It is easy to state that a successful open government initiative requires a champion with the right attitude for the cause (and not those who implements a policy of openness simply because "everyone else is doing it"). While there often exists a significant gap between the ambitions of the politicians and the reality of the task at hand that must then be managed by the public servants, the two nevertheless play equally significant roles in the implementation and maintenance of a successful open government initiative. Widespread collaboration across government departments and within hierarchal ranks is not only important but also necessary. If government is going to establish an open dialogue with the public, it must first ensure that there is an open dialogue on such a topic within its own environment.

Assigning the responsibility of proper oversight and management of an open government program can present a significant challenge. The choice of the group or department that will oversee the implementation and oversight of the policy could have a significant impact on how new strategies are welcomed

and complied with within an organization. Creating general awareness and soliciting support for a new policy throughout government (for example, distribution of memos, employee training, and so on) is an important step in the process. The initiative's success depends, first and foremost, on an employee's ability to comply with and abide by the new framework (Gavelin, Burall, and Wilson 2009, 17). All relationships, both inside and outside of government, must therefore be considered when assigning responsibility and thus accountability.

The International Records Management Trust (IRMT) refers to the group or body designated with this responsibility to oversee the implementation of an open government policy as the "records authority."[31] Aptly named, the records authority should emulate excellence in records management. They are expected to provide the public with evidence of government decisions and actions, information that must be pulled from official government records. This group should therefore be in a position to advise government on records management policy, set standards, define and apply quality control metrics, and enforce compliance (International Records Management Trust 2013, 2–12).

Government records and archives professionals are in a position to significantly contribute to the success of these initiatives. In this regard, the IRMT argue further in favour of the role of a national archives body[32] as the unit within government that is (or should be) responsible for facilitating the creation, distribution, and preservation of authentic and reliable records. A national archives already has it within its mandate to protect "the documentary evidence that shows that a government is following the rule of law, [document] its actions in a transparent fashion, [maintain] evidence of its operations and so [remain] accountable to its citizens" (International Council on Archives 2005, 7; Thurston 2012a, 2). The National Archives and Records Administration (NARA) of the United States and the National Archives of the United Kingdom, for example, already share in the responsibility of their respective country's open government initiatives. NARA, has developed an independent open government plan,[33] which outlines their agency's unique responsibilities with regard to the country's open government initiative (US National Archives and Records Administration 2014). The National Archives oversee the development of the UK Government Licensing Framework, which includes an open government license (National Archives 2014, 2015). Unfortunately, not all government archives, especially lower-level jurisdictions, are in a position to adopt such a role (Thurston 2012b). However, records and archival units can still play a significant role within larger governing administrations by creating partnerships with ATI offices, regulatory bodies, and auditing authorities. Regardless of whether or not an archives is in a position of oversight or authority, the unit's knowledge and expertise can support open government objectives of transparency and accountability by providing the public with reliable evidence of government activities. In sum, the records authority should adopt a leadership role within government in ensuring that the information that is provided to the public is complete and reliable. In the same way, the records authority should be reliable and therefore

in a position to be trusted by both the government and the public to fulfil its responsibilities (International Records Management Trust 2013, 2–12).

Identification and distribution

Once a policy has been approved and roles and responsibilities assigned, the information that will be distributed in support of the policy must be identified. This next phase comprises two key components: records (the *what*) and technology (the *how*).

Records

Government records, created, maintained, and preserved as evidence of government activities, are the primary sources from which open government information must derive. The release of this information must begin by identifying what government information can and will be distributed.[34] Using the organization's classification scheme as a reference point, an inventory[35] or map of the organization's existing records can be created to help identify what information is eligible for publication, either in its current format (for example, government publications, internal communications, and so on) or following strict reformatting requirements (for example, parsing content from certain documents, anonymizing records, and so on).[36]

The data and information that is selected for publication and distribution under a new initiative derives from original government source material that has been repurposed and given value within a new open government context. The value of this information thus depends on the reliability of the source material as well as the accuracy with which this is then communicated with users. Since this material was created as a result of separate business functions and processes, information concerning the work processes, the system and the environment that guided the creation, capture, and control of the source material must therefore be considered as relevant to the publication of the open government information—together, they provide a complete and accurate picture of the information's context and meaning and, therefore, a basis for trust in the information.

Once the records have been identified, the information that will be distributed as open government information is collected and prepared for release by data custodians, technicians, and the like. This preparation entails:

- *Information modification*. Personal and confidential information must be anonymized; open government information that has not been properly vetted for personal and confidential information risks exposing citizens and government clients to privacy breaches.
- *Information reconfiguration or reformatting*. This may be required for some of the selected source materials based on their intended purpose and the platform on which they will be distributed (for example, parsing various data to then distribute it as a structured data set).

- *Identification or creation of relevant contextual metadata.* This may include information describing the original records, including: details about custody/ownership; the production context; omitted information; records management specifications, including retention and classification; legal concerns; and so on. This metadata should be released alongside the open government information so as to provide context, enhance discovery and accessibility by improving searchability, and allow for seamless linking (that is, interoperability) between different data sets. Without context, the information risks being misused and misunderstood, its utility undermined, and its value compromised (Thurston 2012a, 5, 7).[37]

Throughout these steps, documentation depicting how the information was identified, collected, and transformed should be amassed and provided alongside the final disseminated information. As this content provides context and allows for data traceability (that is, being able to trace the information back to the source material), it speaks to the accuracy and reliability of the final information product (Thurston 2012b). Final steps in the workflow include having the final products reviewed by authoritative bodies before being approved by designated employees, often executives, and then published via a web-based platform (for example, open data portal, open information website, and so on).

Technology

Each stage of the workflow outlined earlier requires the support and intervention of various technologies. While the identification, collection, and preparation of the open government information is likely to be achieved by information technicians using existing tools and platforms (for example, electronic document and records management systems; databases; search engines; and so on), the publication of the information may require the creation of new technologies, including websites, portals, and repositories.[38] These web-based technologies represent the front-end design of an open government initiative. The interface with which users will interact will effectively represent the face—albeit virtual—of the government's commitment to transparency and accountability. Open government platforms must be designed with open data principles in mind: access to the open government information cannot be hindered or altogether prevented by the imposition of fees or control barriers (for example, copyright, unwarranted licences, and so on) and information discovery, download, viewing, and repurposing should be made to be seamless and effortless to the user. The quality of the technological platforms will have a direct impact on how users view the quality of the information and the quality of the open government initiative as a whole. Information or tools that are judged to be of poor quality (for example, poor searchability; challenging user interface; data sets are difficult to understand or distributed in non-machine-readable formats or "un-mashable" formats; and so on) may discourage user participation and the reuse of this information, thus jeopardizing the initiative.

While front-end accessibility is important for encouraging citizen engagement and information reuse, it cannot be to the detriment of other system features. Whether designing an open data portal, an open information database, or a service delivery application, records management standards and requirements must be accounted for in the systems' features and design (McDonald 2012, 18). Following by-design principles, records management controls should be embedded within the design of these technologies. These platforms must be able to manage the full lifecycle of the open government information they support (for example, retention, disposition, preservation, and access), as well as the information objects they produce (for example, usage statistics, user feedback, and so on). The necessary controls can be identified by the results of the work process analysis as well as by an assessment of existing technologies and infrastructures, and later adjusted as needed. In this regard, proper guidance by the records authority is important, but seamless collaboration between open government, information technology, and information management professionals is essential not only for the design of the final tools and platforms but also for the compilation of information needed to identify the requirements. Meeting these requirements is essential to building a technical environment that enables a trusted records environment—if the environment can be trusted, so too can the information it manages and produces.

Promotion and evaluation

An open government policy should be designed with the future users of the information in mind, as value creation is strictly dependent on citizens' reuse of open government information. While the release of open government information sets the stage for enhanced transparency and accountability, the latter can only truly be achieved if citizens are made aware of the information's availability and of how it might be used and repurposed for their benefit. As the open government program matures, focus will shift from identifying government records for publication to creating an open dialogue with citizens, further enabling government to identify ways in which the initiative can be improved. Therefore, the strategies and systems that are adopted to support citizen engagement (for example, feedback forms, opportunities to request specific open government information, and so on) become constituent steps, final extensions, of the overall business process.

Citizen engagement and open dialogue

As previously mentioned, citizen engagement is a key pillar in achieving a successful open government initiative. Therefore, the effort of soliciting the public's participation in these types of projects must extend far beyond simply making information available for use. An open government initiative must be promoted and effectively communicated to the public through the use of promotional activities, programs, and tools. Creating incentives for users to make use of the information (for example, hack-a-thons) or lowering the threshold for new users, for example, are just some of the ways of how this can be achieved (Janssen, Charalabidis, and Zuiderwijk 2012, 265).

Communication is also an important component of citizen engagement. Open government information users must be able to effectively and efficiently relay their feedback, concerns, or requests back to those responsible for overseeing this component of the initiative. Common communication mechanisms include open dialogue or feedback forums and web platforms with social media functions, similar to those that may be used also to disseminate information about government activities. Non-web-based interaction could include public in-person consultations, from which feedback could be documented in other ways (for example, forms and audio recordings).

These promotional tactics and communication mechanisms as individual work processes and transactions produce a series of new information objects in the form of outputs (for example, media releases regarding government consultations or collaborations; media communications via social media platforms; and so on) and inputs (for example, completed feedback forms; requests for additional data sets; surveys; communications via social media platforms; and so on). These inputs, once submitted by users, are received and processed by information technicians and aggregated as reports or statistical analysis for reuse. They can be used to interpret patterns and trends in the public's use of the information, helping to inform the organization on how the open government program could improve and should evolve moving forward. This exercise could help government prioritize information releases, streamline internal procedures and activities, lighten workloads, and eliminate redundancies (Janssen, Charalabidis, and Zuiderwijk 2012, 260). By responding to these identified trends in a proactive and efficient manner, government could even influence a reduction in the volume (and therefore the cost) of ATI requests. As such, appropriate records control metrics must be put in place to manage the inputs as well as the outputs. Furthermore, having the means to attest to the veracity of the information that is received from citizens is another consideration that should be incorporated within the overall records management framework of an open government initiative. In the same way that steps must be taken to ensure the reliability and trustworthiness of information disseminated by the government, governments need some assurance that the feedback they receive is also trustworthy and reliable.

Programme evaluation

During the course of this research, there was no evidence of international standards in place for measuring the success of an open government initiative. For the time being, governments that fail to meet their open government goals may be subject to public scrutiny or criticism from advocacy groups such as the OGP, at most. However, that is not to say that internal metrics should not be devised as part of the overall performance and evaluation approach normally taken within an organization. Proper documentation of this evaluative process (for example, how assessment metrics are devised and applied) would be important components of the documentary trail that would also have to be preserved along with the records to which they are linked. Similar to other information

objects previously listed in this article, this information becomes a key piece of evidence of a transparent government and equally as important to a successful open government endeavour. Overall, these metrics would help ensure that the evaluation of an open government initiative would be done in a consistent way, allowing for ongoing tracking of program components over time.

Open government as a business process: a summary

In summarizing what has been highlighted earlier, a central task (or transaction or "process") and, thus, a primary output of an open government initiative is the publication of open government information on a web-based platform. The constituent processes and transactions that support the initiation, realization, and eventual evaluation of this business process each form a part of the overall process and, as a result, generate additional information objects. The series of different information object types that have been identified throughout the work process analysis above can be summarized in six categories:

1. Information objects that are generated to initiate, frame, and underpin the functioning of the program (for example, action plans, policy instruments, procedures, guidelines, assessment forms, checklists, approval forms, open data licenses, and so on).
2. Information objects that serve as inputs within the release process (for example, candidate data sets, inventories or other information objects to be assessed for release).
3. Information objects or, in this case, the open government information that serve as outputs of the release process, to be hosted on open data/open information platforms (for example, access to information request summaries, expense reports, awarded tenders or contracts, government decisions, such as orders in council; various data sets, including geospatial, statistical, financial, and so on). Some of these outputs may be transformations of their original source material, if any anonymizing, severing, or other modification has taken place.
4. Web-based platforms (websites and portals that host and/or enable access to open information objects) that typically include metadata descriptions of released information objects along with links to locally hosted or externally hosted data sets or information objects. For data sets, some of the metadata may describe the structure of the data (for example, fields in tables) as well as the nature of the content.
5. Information inputs from citizens, which typically include feedback from the public on the open government initiative and may also comprise material generated as the result of public consultations (for example, completed feedback forms, audio recordings, etc.) or communications received through web platforms (for example, social media posts, requests for additional data sets, and so on).
6. Other government information objects (inputs and outputs) associated with the open government initiative, including reports, public presentations, or

other information generated to describe activities, progress made, and the results of consultations or other citizen engagement activities. This category also includes communications by the government regarding the initiative, including those made via social media and other platforms.

Adopting an enterprise-level lens when conducting a business process analysis of an open government initiative can help assess the potential business value of these types of information. Assessing their purpose and value within a particular business context will establish whether or not the resulting objects are records or whether they are supporting documents and, also, whether they have transitory, short-term, or enduring value. Decisions can then be made regarding the characteristics of the information identified:

- Has a new data set (record) effectively been created through the release process (that is, was it anonymized or otherwise transformed to accommodate its release)?
- Was any information severed from a document before releasing it?
- Were these changes disclosed upon release? Is there a requirement to keep the released information object in sync with its original and/or official version?
- If standard corporate records management controls are used to manage the original/official version of the information object in a corporate repository, what controls or curation activities are required for managing the released copy (which may be conceptualized as a service or access copy) in the online data or information portal?
- Are different records controls (for example, retention and disposition) required for managing the released versions?

These are all examples of questions that need to be addressed as part of incorporating records management requirements into a new open government initiative within an organization's existing structures and procedures. Understanding the records creation environment of a new business process at both the functional and transactional levels will help identify and map all of the information objects being produced as a result of various processes. Furthermore, it will help clarify and establish proper records control metrics that should be introduced as a way to manage this information.

The documentary trail produced as a result of these processes and transactions is comprised of records, documents, transitory information, and data alike, and it becomes a key piece of evidentiary documentation with regard to the open government initiative. This documentation serves as a reflection of the relationship between the records and the processes as well as the processes and their role within the organization's operations. However, to be considered "complete," this documentation must also account for information concerning the source material from which the open government information stems, including the policies, practices, and systems in which they were created, captured, and maintained. This allows users and government employees alike to accurately trace open government information back to the original source material (Thurston

2012a, 6). A comprehensive documentary trail will define the overall workflow process that initiates, drafts, approves, and posts open government information, from original source to final product. It provides an authoritative, complete, and accurate source in ensuring the integrity and overall reliability of the records that will be distributed as open government information (McDonald 2012, 4). Ensuring the integrity and continued availability of this documentary trail therefore becomes a key records management concern that must be incorporated into the policies and practices of managing open government information. Users are more likely to consider the final product (that is, open government information) to be trustworthy if they are presented with the "big picture" (the contextual information) along with the smallest details (the data).

In sum, an effective enterprise records management framework should allow for the incorporation of the business processes and information objects that support open government initiatives. These new processes should be aligned with existing business structures, which in turn serve as reference points for identifying records management needs, implications, and possible configurations that will be needed to accommodate the distinct requirements of the new initiative. The new organizational records, created as part of a new open government business process, also need to be viewed within the broader organizational context with respect to the records management framework. As such, they should be considered as an extension of the lifecycle of the source material and therefore managed in a way that reflects existing records management practices within the organization.

Records management framework issues with open government initiatives in Canada: preliminary observations

As described in the first section of this article, open government initiatives within various Canadian jurisdictions exist at different stages of development. Preliminary observations of these organizations' policies, action plans, and open government-oriented projects highlight that not all of them completely include the components of the enterprise-level records management framework just explored. For instance, only two of the seven jurisdictions self-describe as containing an aspect of enterprise information management:

1. With DataBC, data management is approached from an enterprise level, recognizing that open data are a subset extracted from a larger pool of operational data. Governance structures and work processes are well defined in its *Concept of Operations* document (British Columbia, Ministry of Labour, Citizens' Services and Open Government 2012).
2. The city of Toronto's open government initiative is conceived as part of its overall *Information Management Framework*, with by-design principles clearly situating open government operational lines as components of broader information management operations, which are themselves integrated with information technology.

With respect to policy and planning, several of the jurisdictions have specific open government policies (British Columbia; Alberta; Toronto), while

others have policies in development. The same situation exists in relation to action plans or statements of strategic objectives: the Government of Canada, Alberta, and Toronto have explicit planning documents devoted to open government initiatives, with other jurisdictions either in the process of developing them (for example, Newfoundland and Labrador) or having less-explicit components of other plans (for example, the city of Vancouver's *Digital Strategy*) referring in some way to open government objectives. Furthermore, only some jurisdictions (British Columbia; Alberta) have publicly provided detailed information about workflows through which government information is released, including documentary tools that demarcate each transaction in the process (for example, assessment forms, checklists, approval forms, and so on).

With respect to open government information, including documents, data, records, and metadata, most jurisdictions have begun by focusing on structured open government data and related portals, including many portals exclusively devoted to geospatial data sets. Further, some jurisdictions (Government of Canada; British Columbia; Newfoundland and Labrador; Regina) have developed or are developing "open information" portals for unstructured government information (for example, information release summaries; reports; expense claims). Also, only some jurisdictions have dedicated websites for open dialogue or citizen interaction (for example, Newfoundland and Labrador; Vancouver; Regina).

Toronto has developed assessment indicators specific to its open government initiative. The city's strategic action plan, which describes components of implementing "open government by design," identifies an assessment framework for each implementation step, including details about indicators, baselines, targets, and enablers (Toronto 2013c). As such, this planning document would be one of the tools by which to coordinate and undertake an evaluation of the progress of the initiative.

These observations emphasize that there are many possible approaches to designing and implementing an open government initiative, with some demonstrated consensus around particularly crucial components (for example, policy, planning, and web-based dissemination platforms for open government information). What could be surmised from this selection of examples are the following points:

- the need for a more comprehensive alignment of enterprise-level open government and records management initiatives within a jurisdiction;
- the need to develop more citizen engagement mechanisms as well as foster citizen-initiated engagement mechanisms (for example, the earlier-mentioned Open Data Saskatchewan) to support both the objectives and principles of open government as well as to obtain necessary feedback on current components to feed future program improvements; and
- the need to devise and disclose (aligning with the spirit of open government) plans, procedures, strategies, and specific assessment metrics for evaluating and tracking the progress of the initiative, demonstrating both a commitment to the initiative as well as the means by which citizens have helped fuel and shape program evolution and improvement.

Next steps

The next phase of this project will seek to expand its analysis of open government initiatives within various Canadian jurisdictions by sharpening its focus on the issues that these groups may be facing with regard to the management of their new organizational records. The project team plans to begin conducting a series of interviews with some of the jurisdictions mentioned in this article to perform a more detailed analysis of the issues, strategies, and projects adopted by these organizations. In this regard, the team will also seek to understand how these jurisdictions may be addressing questions of records capture, retention and disposition, and preservation with regard to their respective open government initiatives. In the same way that it has been approached in this article, the next phase will seek to evaluate the applicability of conducting an enterprise-level analysis of organizational business processes (in this case, of those concerning open government initiatives) as a way to identify and manage the challenges that arise with regard to the creation, capture, and control of these new records. In line with ITrust's long-term objectives, this project aims to develop an enterprise-level framework that would help guide organizations in addressing similar issues.

Conclusion

Open government has become a key political strategy in establishing a trust relationship between a governing body and its citizens. As such, it is an objective that is no longer satisfied by the availability of an open data portal. Despite the various ways in which different government bodies may choose to showcase their support for this type of initiative, the objectives of open government remain the same for all levels of government: enhance transparency, create an environment for greater democratic accountability, and encourage citizen engagement and participation. As this article has explored, a successful open government initiative must be supported by strong policies and key internal players as well as by the dissemination of accurate data and information, stemming from reliable government source records and distributed through innovated platforms that encourage data use and citizen engagement. Such an initiative relies heavily not only on the information it chooses to distribute but also on the nature of that information, including the source material and work processes that were key in that information's creation and eventual release. A reliable enterprise-level recordkeeping framework that fosters an environment of accountability as well as the creation, capture, and control of complete, accurate, and reliable records, therefore, centres at the heart of a successful open government initiative and thus a trust relationship between government and citizens.

Open government may not be a new concept, but it is a changing one. A combination of modern technologies, social media outlets and whistleblower-like tendencies have not only had a significant impact on the sometimes unintended ways in which government information is distributed into the public sphere, but it has also shaped the way both government and the public have approached questions of transparency and accountability. With the spotlight shining brighter than ever before, governments are under increasing public

pressure to demonstrate their commitment to a more open form of governance. While standardized metrics for measuring the success or failure of these initiatives may not yet exist, public response and feedback can shed a significant amount of light on how these initiatives are currently being received.

Acknowledgements

This project was realized through the support of the Social Science and Humanities Research Council-funded InterPARES Trust project based at the School of Library, Archival and Information Studies at the University of British Columbia and headed by Project Director Luciana Duranti. The authors are grateful for the comments and editorial suggestions made by Jim Suderman, records director at the City of Toronto, John McDonald, a retired consultant specializing in records and information management, and Grant Hurley, a recent graduate of the School of Library, Archival and Information Studies at the University of British Columbia, all of whom are fellow researchers on InterPARES Project 08 "The Implications of Open Government, Open Data, and Big Data on the Management of Digital Records in an Online Environment." Please note that any opinions expressed herein are solely those of the authors.

Notes

1. Archival aspects and concerns within a broader records management framework, while important, have not been explicitly explored within the scope of this particular article. The project team hopes to include this lens in our future studies.
2. Some jurisdictions pursued freedom of press movements much earlier. For example, Sweden introduced its Freedom of the Press Act in 1766.
3. The United States introduced their Freedom of Information Act in 1966 following the end of the Cold War. It was employed as a countermeasure to the tactics of secrecy and information restrictions that were often employed prior as a way to protect the country from foreign spies (Yu and Robinson 2012, 184–85).
4. Access to Information Act, RSC 1985, c A-1.
5. Privacy Act, RSC 1985, c P-21.
6. "Trust" and, thus, a "trust relationship" between government and citizens is, in and of itself, a complex concept. It is outside the scope of this article to explore the underlying issues of what may or may not be required to establish a "trust relationship" between governments and citizens. However, this article, recognizing that "open information" is only a subset of government information, does not advocate "complete access to information" as a "be all, end all" solution to establishing trust in government but, rather, an aspect to consider in this endeavour.
7. This list of characteristics and thus the definition of open data provided in this article has been made based on the author's (Léveillé) interpretation of a number of different sources (Janssen, Charalabidis, and Zuiderwijk 2012; Open Knowledge Foundation 2011, 2012; Thurston 2012b; James 2013; McDonald and Léveillé 2014; Ubaldi 2013; Zuiderwijk et al. 2014).
8. While "universal participation" is the ultimate goal, there is always a risk that opening up government information may "further contribute to the digital divide," as only certain groups may be able to learn how to use and fully benefit from data. (Janssen, Charalabidis, and Zuiderwijk 2012, 263).

9. It is important to note that the benefits of using open data, including the value of the data itself, are subject to change over time. While governments may not always be able to anticipate these changes, they may be required to adapt quickly in order to keep the public engaged (Janssen, Charalabidis, and Zuiderwijk 2012, 260).
10. The International Aid Transparency Initiative is a "voluntary, multi-stakeholder initiative that seeks to improve the transparency of aid, development and humanitarian resources in order to increase their effectiveness in tackling poverty" (International Aid Transparency Initiative 2014).
11. The analysis presented in this portion of the article was drafted as a result of a literature review of the resources that have been made publicly available online under the assumption that these resources would provide a fair representation of the open government initiatives currently underway within these individual Canadian jurisdictions. These examples and the analysis presented as a result thus offer an initial starting point for the complete-picture analysis of the Canadian landscape with regard to open government initiatives—to be accomplished as a future objective of this research—as it is acknowledged that there may exist certain limitations to a literature review (for example, not all pertinent resources may be available online; many of these documents do not present feedback or results with regards to the perceived level of success of an initiative; and so on) Addressing citizen-lead open government or open data initiatives is beyond the scope of this article.
12. Generally, jurisdictions with more mature programs have been covered in more detail in this overview. A potential exception to this is the selection of municipal examples, which were chosen more arbitrarily but with the goal to demonstrate different approaches.
13. Previously available at data.gc.ca, open data sets are now accessible, along with open information resources, on one common portal at http://open.canada.ca/en (Canada 2014f).
14. While the FAQs note the use of the Open Data Portal Metadata Standard, it does not clarify that departments individually host their own data. Still, this fact is evident when checking any URL on the site to download individual datasets.
15. While there are many other notable open government and open data initiatives underway in other countries, it is beyond the scope of this article to discuss them in any detail.
16. Freedom of Information and Protection of Privacy Act, RSO 1990, c F-31.
17. Section 5.3 of DataBC's *Concept of Operations* document provides detailed workflow diagrams demonstrating various steps in the process toward releasing datasets including assessment, analysis, notification, prioritization, publication, hosting, validation, and so on (British Columbia, Ministry of Labour, Citizens' Services and Open Government 2012, 33–38).
18. Access by design (AbD) and privacy by design (PbD) are approaches to enabling access or protecting privacy by embedding mechanisms into design specifications of technologies, business practices and physical infrastructures. Both are based on foundational sets of principles. In relation to open government, by accommodating AbD and/or PbD into information systems at the outset, the resulting systems inherently could support open information and open data requirements (Information and Privacy Commissioner of Ontario 2014a, 2014b).
19. The need for legislative reform to support modern records and archives management was further addressed in *A Failure to Archive: Recommendations to Modernize Government Information Management*, also produced by Elizabeth Denham (2014) in the Office of the Information and Privacy Commissioner for British Columbia.

20. For more information, see Open Data Program, Data Value Framework (Alberta 2013b); Dataset Evaluation Form (Alberta 2013a); and Open Data Assessment Checklist (Alberta 2014a).
21. The Alberta "Open Government Program—frequently asked questions" document states that Alberta is one of the first sub-national jurisdictions to develop a policy consistent with the G8 Open Data Charter (Alberta n.d., 1).
22. See also Freedom of Information and Protection of Privacy Act, RSA 2000, c F-25.
23. Pre-dating the current open government initiative, the Community Accounts website has been providing public access to different sources of socio-economic government data along with explanatory reference material and supporting tools since 1996 (Newfoundland and Labrador 2014b).
24. See *Manitoba Land Initiative* (Manitoba 2014); GeoNOVA, *Geographic Gateway to Nova Scotia* (Nova Scotia 2014); *GIS Data Layers* (Prince Edward Island 2014); *Geomatics Yukon* (Yukon 2011); *Centre for Geomatics* (Northwest Territories 2014a); *Geoscience Office – Research, Analysis, Information* (Northwest Territories 2014b); *Canada-Nunavut Geoscience Office* (Nunavut 2014).
25. Several lists of municipal initiatives have been compiled and are available at *Open Data in Canada* (Canada 2014e) and at *Open Data* (Datalibre.ca 2014). Further, Liam James Currie (2013) completed a Master's thesis entitled "The Role of Canadian Municipal Open Data Initiatives: A Multi-city Evaluation," which examines existing municipal open data initiatives to assess the role they play in relation to open government, including a detailed assessment of the type of data released and challenges facing the municipal programs. This thesis includes detailed case studies of ten Canadian municipalities: Toronto, Ontario; Edmonton, Alberta; Ottawa, Ontario; Montreal, Quebec; District of North Vancouver, British Columbia; Mississauga, Ontario; Regina, Saskatchewan; Guelph, Ontario; Fredericton, New Brunswick; and Hamilton, Ontario.
26. Little information is available on this working group, though it is referenced on some web resources (Toronto 2014a; Regional Municipality of York 2013; MARS Discovery District 2014).
27. The catalogue includes both metadata descriptions of the datasets as well as the datasets themselves. Municipal Freedom of Information and Protection of Privacy Act, RSO 1990, c M-56.
28. Local Authority Freedom of Information and Protection of Privacy Act, SS 1990, c L-27.1.
29. See also OpenDataSK.ca (2014). HackRegina is an annual hackathon event held in Regina, Saskatchewan. The fifth hackathon was in spring 2014 (McCallum 2014).
30. The United States' open data portal (data.gov) uses a standardized metadata template (Thurston 2012b).
31. In drafting their tool for measuring *Open Government and Trustworthy Records*, the International Records Management Trust (IRMT) (2013, 2) outline a series of benchmarks for implementing a successful open government initiative, including the ideal role that should be played by the records authority. The IRMT define this role in the following way: "A records authority (a national records/ archives body or state/ local body with equivalent authority) is empowered to advise government on policy, set standards and define quality controls for the management of public records in all formats."
32. The IRMT's *Open Government and Trustworthy Records: Institutional/ Regulatory Framework and Capacity Benchmarking Tool* (2013) only explores the role of the national archives in their framework. Unfortunately, the tool does not explore alternative options for provincial/state or municipal governments.

33. On 30 May 2014, National Archives and Records Administration (2014) published its third *Open Government Plan*.
34. From a general perspective, this may be outlined in the policy (for example, categories of information types for potential release). Further, government information requests are never static in nature and the choice to distribute what information and when is likely to be subject to frequent change.
35. Mapping the information sources of an organization has been a strategy adopted by the UK government as well as by the Girona City Council when implementing its open data project (Open Government Partnership 2014b; Casselas 2013, 2).
36. The inventory not only becomes a by-product of the activity but also a useful management tool for government (that is, it could help prioritize future open government information releases and streamline workflows) and a significant piece of documentation that can be used to further support government transparency.
37. While the importance of metadata is not disputed, the current problem lies in that there is currently no consistency in different metadata models and/or formats that are used across different open data types. Inputting metadata can be a time-consuming endeavour and perhaps overlooked in part by organizations lacking the necessary staff and resources to accomplish the task. Furthermore, there is also a risk of metadata containing assumptions for the use of the data and pointing to certain choices and interpretations to be made, which could create biases that would otherwise exclude certain ways in which the data could be reused (Zuiderwijk, Jeffrey, and Jannsen 2012, 232).
38. It should be acknowledged that the introduction of new technological platforms could also result in the creation of new partnerships for government, for example, outsourcing the design, hosting and/or the management of these tools to third-party companies, which could present equal benefits and risks for organizations. It is beyond the scope of this article to explore this dimension.

References

Alberta. N.d. *Open Data Standards*. Edmonton, AB. http://data.alberta.ca/sites/default/files/Open%20Data%20Standards.pdf.

——. 2012a. *Alberta Open Data*. http://data.alberta.ca/.

——. 2012b. *Open Information and Open Data Policy*. http://data.alberta.ca/content/government-alberta-open-information-and-open-data-policy.

——. 2012c. *Open Government Licence: Alberta*. http://data.alberta.ca/licence.

——. 2013a. *Dataset Evaluation Form*. Edmonton, AB. http://data.alberta.ca/sites/default/files/02%20%20Dataset%20Evaluation%20Form.pdf.

——. 2013b. *Open Data Program, Data Value Framework*. Edmonton, AB. http://data.alberta.ca/sites/default/files/01%20%20Open%20Data%20Value%20Framework.pdf.

——. Open Government Office. 2013c. *Open Government Action Plan*. Edmonton, AB. http://data.alberta.ca/sites/default/files/Alberta%20Open%20Government%20Action%20Plan%20%20v5.7.pdf.

——. Open Government Program. 2014a. *Open Data Assessment Checklist*. Edmonton, AB. http://data.alberta.ca/sites/default/files/ADMINSA0040.pdf.

——. 2014b. *Open Data Dataset Publishing Guidelines*. Edmonton, AB. http://data.alberta.ca/sites/default/files/Open%20Data%20Dataset%20Publishing%20Guidelines.pdf.

British Columbia. 2014a. *Open Government License: British Columbia*. http://www.data.gov.bc.ca/local/dbc/docs/license/OGL-vbc2.0.pdf.

——. 2014b. *Open Information*. http://www.openinfo.gov.bc.ca/ibc/index.page.
——. 2014c. DataBC. http://www.data.gov.bc.ca.
British Columbia. Ministry of Labour, Citizens' Services and Open Government. 2012. *DataBC: Concept of Operations*, vol. 1. Victoria, BC. http://www.data.gov.bc.ca/local/dbc/docs/databc/DataBC_Concept_of_Operations_-_V1.0.pdf.
——. Office of the Chief Information Officer. Knowledge and Information Services Branch. Ministry of Labour, Citizens' Services and Open Government. 2011. *Open Information and Open Data Policy*, version 1.0. Victoria, BC. http://www.cio.gov.bc.ca/local/cio/kis/pdfs/open_data.pdf.
Canada. 2012. *Canada's Action Plan on Open Government 2014-2016*. http://open.canada.ca/en/canadas-action-plan-open-government-2014-16.
Canada. 2014a. *Directive on Open Government*. http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=28108.!
——. 2014b. *Frequently Asked Questions*. http://open.canada.ca/en/frequently-asked-questions.
——. 2014c. *G8 Open Data Charter: Canada's Action Plan*. http://open.canada.ca/en/g8-open-data-charter-canadas-action-plan.
——. 2014d. *GeoConnections: Discovery Portal*. http://geodiscover.cgdi.ca/web/guest/home.
——. 2014e. *Open Data in Canada*. http://open.canada.ca/en/maps/open-data-canada.
——. 2014f. *Open Government*. http://open.canada.ca/en.
——. 2014g. *Open Government License: Canada*. http://open.canada.ca/en/open-government-licence-canada.
Casselas, L.-E. 2013. *Mapping, Selecting and Opening Data: The Records Management Contribution to the Open Data Project in the City Council of Girona*. Paper delivered at the ICA Annual Conference, Brussels, Belgium, 23–24 November. http://www.ica.org/?lid=14819&bid=1134.
Currie, Liam James. 2013. "The Role of Canadian Municipal Open Data Initiatives: A Multi-City Evaluation." MA thesis, Queen's University. http://qspace.library.queensu.ca/bitstream/1974/8159/1/Currie_Liam_J_201308_MA.pdf.
Datalibre.ca. 2014. *Open Data*. http://datalibre.ca/links-resources/.
Denham, Elizabeth. 2013. *Evaluating the Government of British Columbia's Open Government Initiative*, Investigation Report F13–03: Information and Privacy Commissioner for British Columbia. Victoria, BC. https://www.oipc.bc.ca/investigation-reports/1553.
——. 2014. *A Failure to Archive: Recommendations to Modernize Government Information Management*, Special Report: Information and Privacy Commissioner for British Columbia. Victoria, BC. https://www.oipc.bc.ca/special-reports/1664.
European Commission. 2011. *Open Data: An Engine for Innovation, Growth and Transparent Governance* Europa.eu Digital Agenda for Europe COM(2011) 822 final. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:EN:PDF
Francoli, Mary. 2011. "What Makes Government 'Open'?" *eJournal of eDemocracy and Open Government* 3 (2): 152–65. http://www.jedem.org/article/view/65/85.
Gautrin, Henri-François. 2012. *Gouverner ensemble: Comment le Web 2.0 améliorera-t-il les services aux citoyens?* Québec, QC. http://www.mce.gouv.qc.ca/publications/rapport-gautrin-web-2-2012-03-06.pdf.

Gavelin, Karin, Simon Burall, and Richard Wilson. 2009. *Open Government: Beyond Static Measure*, A paper produced by Involve for the OECD. http://www.oecd.org/gov/46560184.pdf.
Giggey, Robert. 2012. *The G4: Setting the City Data Free*, Canadian Government Executive 17 (8). http://www.canadiangovernmentexecutive.ca/category/item/152-the-g4-setting-city-data-free.html.
International Council on Archives. Committee on Current Records in an Electronic Format. 2005. *Electronic Records: A Workbook for Archivists*, ICA Study 16. Paris, France. http://www.ica.org/download.php?id=1612.
Information and Privacy Commission of Ontario. 2010. *Access by Design: The Seven Fundamental Principles*. https://www.ipc.on.ca/images/Resources/accessbydesign_7fundamentalprinciples.pdf.
International Aid Transparency Initiative. 2014. *About IATI*. http://www.aidtransparency.net/about.
International Records Management Trust. 2013. *Open Government and Trustworthy Records Institutional/Regulatory Framework and Capacity Benchmarking Tool*. http://www.irmt.org/portfolio/open-government-trustworthy-records/attachment/benchmarks-for-open-government-and-trustworthy-records-final-2.
InterPARES 2 Project. N.d. *Data: Terminology Database*. http://www.interpares.org/ip2/ip2_terminology_db.cfm
InterPARES Trust. 2014. *Home*. https://interparestrust.org/.
International Standardization Organization (ISO). 2008. *Technical Report: Information and Documentation – Work Process Analysis for Records*, ISO/TR 26122:2008(E). Switzerland: ISO.
James, L. 2013. *Defining Open Data: Open Knowledge Foundation Blog*. http://blog.okfn.org/2013/10/03/defining-open-data/.
Janssen, Marijn, Yannis Charalabidis, and Anneke Zuiderwijk. 2012. ''Benefits, Adoption Barriers and Myths of Open Data and Open Government.'' *Information Systems Management* 29 (4): 258–68. http://dx.doi.org/10.1080/10580530.2012.716740.
Manitoba. 2014. *Manitoba Land Initiative*. http://mli2.gov.mb.ca/.
MARS Discovery District. 2014. *What We Do – Open Data*. http://www.marsdd.com/systems-change/data-catalyst/what-we-do/.
McCallum, Chad. 2014. *HackREGINA Spring 2014 Winners*! *HackDays [blog]*. http://hackdays.ca/.
McDonald, John. 2012. ''Managing Electronic Records: The Importance of Standards.'' Paper delivered at the SARBICA Conference on Electronic Records, Bangkok, Thailand, June.
McDonald, John, and Valerie Léveillé. 2014. ''Whither the Retention Schedule in the Era of Big Data and Open Data?'' *Records Management Journal* 24 (2): 99–121. http://dx.doi.org/10.1108/RMJ-01-2014-0010.
National Archives. 2014. *Open Government License for Public Sector Information*. http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2/.
——. 2015. *Licensing for Re-use*. http://www.nationalarchives.gov.uk/information-management/re-using-public-sector-information/licensing-for-re-use/.
National Archives and Records Administration. 2014. *Open Government Plan: 2014–2016*. http://www.archives.gov/open/open-government-plan-3.0.pdf.
Natural Resources Canada. 2012. *The Atlas of Canada*. http://atlas.nrcan.gc.ca/site/english/index.html.
——. 2015. *Free Data: GeoGratis*. https://www.nrcan.gc.ca/earth-sciences/geography/topographic-information/free-data-geogratis/11042.

Newfoundland and Labrador. 2014a. *Collaboration*. http://open.gov.nl.ca/collaboration/default.html.
——. 2014b. *Community Accounts – Data, Information, Knowledge*. http://nl.communityaccounts.ca/default.asp.
——. 2014c. *Dialogue*. http://open.gov.nl.ca/dialogue/default.html.
——. 2014d. *License*. http://opendata.gov.nl.ca/public/opendata/page/?page-id=licence.
——. 2014e. *Open Data*. http://opendata.gov.nl.ca/.
——. 2014f. *Open Government*. http://open.gov.nl.ca/what.html.
——. 2014g. *Open Information*. http://open.gov.nl.ca/information/default.html.
Northwest Territories. 2014a. *Centre for Geomatics*. http://www.geomatics.gov.nt.ca/.
——. 2014b. *Geoscience Office: Research, Analysis, Information*. http://www.nwtgeoscience.ca/.
Nova Scotia. 2014. *GeoNOVA, Geographic Gateway to Nova Scotia*. http://www.novascotia.ca/geonova/home/default.asp.
Nunavut. 2014. *Canada-Nunavut Geoscience Office*. http://cngo.ca/.
O'Hara, Kieron. 2012. "Transparency, Open Data and Trust in Government: Shaping the Infosphere." WebSci '12 Proceedings of the 4th Annual ACM Web Science Conference, 223–32. http://dx.doi.org/10.1145/2380718.2380747.
Ontario. 2014a. *Open Data*. http://www.ontario.ca/government/ontario-open-data.
——. Open Government Engagement Team. 2014b. *Open by Default—A New Way Forward for Ontario*. Toronto, ON. https://dr6j45jk9xcmk.cloudfront.net/documents/2428/open-by-default-2.pdf.
——. 2014c. *Open Government*. http://www.ontario.ca/government/open-government.
OpenDataSK. 2014. *OpenDataSK.ca*. http://opendatask.ca/.
Open Government Partnership. 2014a. *Sweden*. http://www.opengovpartnership.org/country/sweden.
——. 2014b. *United Kingdom*. http://www.opengovpartnership.org/country/united-kingdom.
——. 2014c. *United States*. http://www.opengovpartnership.org/country/united-states.
Open Knowledge. 2014. *What Is Open?* https://okfn.org/opendata/.
Open Knowledge Foundation. 2011. *Open Definition*. http://opendefinition.org/od/.
——. 2012. *Open Data Handbook*, release 1.0.0. http://opendatahandbook.org/pdf/OpenDataHandbook.pdf.
Prince Edward Island. 2014. *GIS Data Layers*. http://www.gov.pe.ca/gis/.
Quebec. 2014a. *Déclaration du gouvernement du Québec*. http://www.défis.gouv.qc.ca/?node=/declaration.
——. 2014b. *Données.gouv.qc.ca BETA*. http://www.défis.gouv.qc.ca/.
——. 2014c. *Données ouvertes: Le gouvernement du Québec et les Villes de Québec, Montréal, Gatineau et Sherbrooke adoptent une licence commune d'utilisation*. http://www.fil-information.gouv.qc.ca/Pages/Article.aspx?aiguillage=ajd&idMenuItem=1&idArticle=2202195608.
——. 2014d. *Licence*. http://www.défis.gouv.qc.ca/?node=/licence.
Regina. 2014a. *About OGDI DataLab*. http://openregina.cloudapp.net/Home/About.
——. 2014b. *City of Regina Datasets*. http://openregina.cloudapp.net/.
——. 2014c. *Open Government*. http://www.regina.ca/residents/open-government/.
——. 2014d. *Open Info*. https://www.regina.ca/residents/open-government/open-information/.

——. 2015. *Open Government Licence – City of Regina*. https://www.regina.ca/residents/open-government/open-government-licence/.
Regional Municipality of York. Committee of the Whole, Planning and Economic Development. 2013. *Open Data for York Region—Moving Forward*. Regional Municipality of York, ON. http://archives.york.ca/councilcommitteearchives/pdf/sep%2012%20ped%20open.pdf.
Scassa, Teresa. 2014. "Privacy and Open Government." *Future Internet* 6 (2): 397–413. http://dx.doi.org/10.3390/fi6020397.
Thurston, Anne. 2012a. "Public Records: Evidence for Openness." Paper delivered at the Institute of Commonwealth Studies Secrecy and Disclosure: Freedom of Information and the Commonwealth Conference, London, England, 14 June. http://blogs.estadao.com.br/publicos/files/2012/08/Public-Records-as-Evidence-for-Openness-FINAL.doc_.pdf
——. Anne. 2012b. "Trustworthy Records and Open Data." *Journal of Community Informatics* 8 (2). http://ci-journal.net/index.php/ciej/article/view/951/952.
Toronto. Corporate Information Management Services. 2012. *Open Data Policy*. http://www1.toronto.ca/wps/portal/contentonly?vgnextoid=7e27e03bb8d1e310VgnVCM10000071d60f89RCRD.
——. City Clerk's Office. 2013a. *Information Management Framework*. Toronto, ON. https://www1.toronto.ca/City%20Of%20Toronto/City%20Clerks/Corporate%20Information%20Management%20Services/Files/pdf/I/IMFrameworkToronto.pdf.
——. 2013b. *Open Government Licence: Toronto*. http://www1.toronto.ca/wps/portal/contentonly?vgnextoid=4a37e03bb8d1e310VgnVCM10000071d60f89RCRD&appInstanceName=default.
——. 2013c. *Strategic Action no. 13: Open Government by Design*. Toronto, ON. http://www1.toronto.ca/City%20Of%20Toronto/City%20Clerks/Corporate%20Information%20Management%20Services/Teaser/Strat%20Action%2013%20Open%20Gov%20by%20Design%20Implementation%20Steps%20FINALN_1.pdf.
——. 2013d. *Strategic Actions, 2013–2018*. Toronto, ON. http://www1.toronto.ca/City%20Of%20Toronto/City%20Manager%27s%20Office/Files/StratActionsBklt_Tags.pdf.
——. 2014a. *Open Data*. http://www1.toronto.ca/wps/portal/contentonly?vgnextoid=9e56e03bb8d1e310VgnVCM10000071d60f89RCRD.
——. 2014b. *Open Data – Data Catalogue*. http://www1.toronto.ca/wps/portal/contentonly?vgnextoid=1a66e03bb8d1e310VgnVCM10000071d60f89RCRD.
Ubaldi, Barbara. 2013. *Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives*. OECD Working Papers on Public Governance no. 22. OECD Publishing. doi: 10.1787/5k46bj4f03s7-en.
US National Archives and Records Administration. 2014. *Open Government at the National Archives*. http://www.archives.gov/open/.
Vancouver. 2013. *Digital Strategy*. Vancouver, BC. http://vancouver.ca/files/cov/City_of_Vancouver_Digital_Strategy.pdf.
——. 2014a. *Corporate Business Plan*. Vancouver, BC. http://vancouver.ca/files/cov/corporate-business-plan.pdf.
——. 2014b. *Open Data Catalogue*. http://vancouver.ca/your-government/open-data-catalogue.aspx.
——. 2014c. *Talk Vancouver*. https://www.talkvancouver.com/Portal/default.aspx.
Vancouver Chief Librarian. 2013. *Digital Strategy*, RR2 Administrative Report. Vancouver, BC. http://former.vancouver.ca/ctyclerk/cclerk/20130409/documents/rr2.pdf.

Veljković, Nataša, Sanja Bogdanović-Dinić, and Leonid Stoimenov. 2014. "Benchmarking Open Government: An Open Data Perspective." *Government Information Quarterly* 31 (2): 278–90. http://dx.doi.org/10.1016/j.giq.2013.10.011.

Yu, Harlan, and David G. Robinson. 2012. "The New Ambiguity of 'Open Government'." *UCLA Law Review Disclosure* 59: 178–208. doi: 10.2139/ssrn.2012489.

Yukon. 2011. *Geomatics Yukon*. http://www.geomaticsyukon.ca/.

Zuiderwijk, Anneke, Keith Jeffery, and Marijn Janssen. 2012. "The Potential of Metadata for Linked Open Data and its Value for Users and Publishers." *eJournal of eDemocracy and Open Government* 4 (2): 222–44.

Zuiderwijk, Anneke, and Marijn Janssen. 2014. "Open Data Policies, Their Implementation and Impact: A Framework for Comparison." *Government Information Quarterly* 31 (1): 17–29. http://dx.doi.org/10.1016/j.giq.2013.04.003.

Zuiderwijk, Anneke, Marijn Janssen, Sunil Choenni, and Ronald Meijer. 2014. "Design Principles for Improving the Process of Publishing Open Data." *Transforming Government: People, Process and Policy* 8 (2): 185–204. http://dx.doi.org/10.1108/TG-07-2013-0024.

New Technologies, New Challenges: Records Retention and Disposition in a Cloud Environment

Nouvelles technologies, nouveaux défis : Conservation et déclassement des documents dans un environnement de nuage informatique

Patricia. C. Franks
School of Information, San José State University
patricia.franks@sjsu.edu

Abstract: This article describes core records retention and disposition functional requirements extrapolated from relevant standards and guidelines and from responses to a questionnaire developed to gather information about retention and disposition functionality built into cloud services. Results of a survey completed by 168 records and information professionals are analysed to identify cloud records retention and disposition challenges. Gaps between the functionality required and that provided in selected cloud environments are identified, and recommendations to mitigate the challenges posed are offered. Future research plans are shared.

Keywords: records, retention, disposition, cloud services, functional requirements

Résumé : Cet article décrit les exigences fonctionnelles de base pour la conservation et le déclassement des documents, extrapolées à partir des normes et des lignes directrices pertinentes ainsi que des réponses à un questionnaire élaboré dans le but de recueillir des informations sur la conservation et le déclassement en tant que fonctionnalités intégrées dans les services informatiques en nuage. Nous avons analysé les résultats d'une enquête réalisée auprès de 168 professionnels de la gestion documentaire et de l'information afin d'identifier les défis de conservation et de déclassement des documents. Nous avons identifié des écarts entre les fonctionnalités requises et celles qui sont prévues dans les environnements informatiques en nuage sélectionnés. Nous formulons des recommandations visant à atténuer les défis posés et nous partageons nos projets de recherche à venir.

Mots-clés : documents, conservation, déclassement, services informatiques en nuage, exigences fonctionnelles

Introduction

Businesses and government agencies employ cloud services to take advantage of the benefits offered, such as increased operational efficiencies, accessibility, collaboration, security, reliability, and opportunities for innovation. Individuals within organizations charged with managing the records and information residing in the cloud understand, however, the myriad challenges presented when control

is relinquished to a third-party provider. One way to minimize risks associated with content stored in the cloud is to employ only an enterprise-hosted private cloud to communicate, collaborate, and conduct transactions. To do so, though, would sacrifice the opportunity to engage the public. Organizations, therefore, are increasingly turning to hybrid solutions.

By 2017, nearly one-half of all large enterprises are expected to be engaged in hybrid (that is, public/private) cloud computing (Babcock 2013). A hybrid cloud is an integrated cloud service using both private and public clouds to perform distinct functions within the same organization. A 2014 survey of 1,068 technical professionals revealed that "hybrid and multi-cloud implementations continue to be the end goal for the enterprise: 74 percent of enterprise respondents have a multi-cloud strategy, and 48 percent are planning for hybrid clouds" (Weins 2014).

Regardless of the implementation model, it is essential that organizations are able to "trust" that their records residing in the cloud can be retained and disposed of in accordance with the same requirements that govern the retention and disposition of records stored within the enterprise (Franks and Doyle 2014, 52).

Methodology

In June 2013, InterPARES Trust, a multi-national, interdisciplinary research project funded by a five-year partnership grant (2013–18) from the Social Sciences and Humanities Research Council of Canada was launched to explore issues concerning digital records and data entrusted to the Internet. Several studies were approved to investigate diverse facets of the larger research agenda, among them retention and disposition in a cloud environment research project. The project committee determined that a qualitative approach should be taken to answer two research questions:

- How does the use of cloud services affect our ability to retain and dispose of records in accordance with the law and other applicable guidelines?
- What can be done to mitigate any risks arising from the gaps between our ability to apply retention and disposition actions to manage records residing within the enterprise and those residing in the cloud?

Three data collection methods were employed to achieve these objectives: (1) content analysis of national and international standards and guidelines to identify functional requirements for retention and disposition of records stored in electronic systems; (2) analysis of data gathered through publicly available cloud service information and interviews with cloud service representatives to understand the retention and disposition functionality offered; and (3) an analysis of the responses to an online survey of records and information management professionals to understand the retention and disposition challenges faced when records and information reside in a cloud environment.

Literature review

Five themes that emerged from a review of the literature are risk analysis and risk management, legal issues, information governance, new approaches to manage retention and disposition in the cloud and records classification.

Risk analysis and risk management

Allison Grounds and Benjamin Cheesbro (2013) cite eDiscovery risks due to the mismanagement of retention policies and the inability to implement legal holds successfully in the cloud environment. Peter Géczy, Noriaki Izumi, and Köiti Hasida (2013) purport that two major risks posed by a hybrid cloud are inherited from their public cloud segment: data security and loss of control. Of the top ten cloud computing risks identified by Amab Dutta, Chao Alex Peng, and Alok Choudhary (2013, 44), two impact the organization's ability to adequately govern content residing in the cloud: difficulties in changing cloud vendors in the event of service dissatisfaction (vendor lock-in) and enterprise data re-migration difficulties at the end of the cloud contract.

Legal issues

When managing information in the cloud environment, retention and disposition no longer entail local storage but, rather, global and cross-border storage locations with multiple jurisdictional laws, especially related to data privacy. To respond to this dilemma, some cloud providers locate physical data centres in various geographic regions. According to Masooda Bashir and Jay Kesan (2011), contracts and terms of service agreements do not protect customer data from misuse of data or disclosure of data to third parties by cloud service providers. Iulia Ion and his colleagues (2011) observe that the expectation of privacy is not typically written into cloud provider service agreements. Cloud users potentially do not even know if and when their data are accessed by other users. For instance, on 23 July 2013, the Supreme Court of the State of New York (2013) ordered the execution of 381 search warrants directed at subscribers of Facebook, authorizing the district attorney and its investigators to search and seize digital information uploaded by hundreds of individual account users and stored within Facebook's servers. As a result of the fungible nature of digital information, the ability of a user to delete information instantly, and other possible consequences of disclosure, the court ordered the search warrants sealed and Facebook not to disclose the search and seizure to its users. This decision has implications for customers of hosted cloud services—such as Google Docs and Amazon's Cloud—in that the court found that, as a mere "landlord" or custodian of the customers' records, Facebook had no "legitimate expectation of privacy" in the customer or client's records.

Information governance and records in the cloud

In 2009, ARMA International identified eight Generally Accepted Recordkeeping Principles® that could be applied to records residing in the cloud. To do so, organizations must address, among other issues, a persistent preservation strategy

and disposition practices that ensure removal of both data and metadata (Hoke 2011). The Information Governance Maturity Model, which was also developed by ARMA International (2010), describes transformational retention programs as those that, among meeting other criteria, apply retention to all information in an organization, not just official records, and a transformational disposition process as one that covers all records and information in all media. Disposition is assisted by technology and is integrated into all applications, data warehouses, and repositories.

To maintain effective information governance for records residing in a public cloud, "preservation of metadata" and "enforcement of retention periods" should be included as two key components of service agreements and contracts (Blair 2010). A private cloud can offer retention and disposition capabilities that public clouds do not. For example, Hewlett Packard (HP) Autonomy's private cloud utilizes a cloud-based suite of meaning-based governance solutions that enable the organization to enforce defensible governance in archiving, eDiscovery, compliance, data protection, and records management (HP Autonomy 2013).

Emergence of new approaches to handle research and development in the cloud

As a result of the variety of cloud models, products, services, and vendors, new approaches to retention and disposition challenges will take a variety of forms. Currently, product documentation reflects that the data centres of most cloud vendors are designed to be compliant with physical and network security, but very few of those investigated for this study offered more than limited retention and disposition functionality. Yang Tang and his colleagues (2010) propose file assured deletion (FADE) encryption technology to implement and execute retention and disposition policies. This technology will also facilitate complete data withdrawal when switching vendors. Hitachi Data Systems explains that Hitachi Content Platform (HCP) ensures retention and disposition in the cloud environment, enables litigation hold or release, and provides assurances for data segregation in a multi-tenancy environment (Ratner 2013).

Few of the cloud products or services reviewed are designed to provide long-term retention. Jan Askhoj, Shigeo Sugimoto, and Mitsuharu Nagamori (2011) suggest remodelling the Open Archival Information System (OAIS) with a platform-as-a-service (PaaS) layer, a software-as-a-service (SaaS) layer, a preservation layer, and an interaction layer to preserve records in the cloud (2011). One vendor reviewed, Preservica, offers active preservation solutions based on the OAIS model that are available in cloud-hosted and on-premise editions. Preservica supports workflows to automate bulk ingest of exported DSpace, CONTENTdm, SharePoint, and Outlook packages, advanced website harvesting, and the bulk ingest of digitized content (http://preservica.com/).

Records classification

Many electronic records systems identify the disposition status and retention period of the record at the point of capture and registration, a process that can

be linked to business activity-based classification. The classification terms are applied to the aggregation (that is, a file or container); individual records contained in the aggregation inherit the classification terms. When the classification scheme is mapped to retention requirements, inherited classification facilitates the retention and disposal of aggregations of records.

International Organization for Standardization (ISO) 16175, Module 2, 3.6.1 on Disposition Authorities specifies that an "electronic records management system must by default ensure that every record in an aggregation is governed by the disposal authority(s) associated with that aggregation." It further states that the electronic records management system must, for each aggregation, automatically track retention periods that have been allocated to the aggregation; and initiate the disposition process." According to ISO 16175, more than one disposal authority may be associated with an aggregation. If so, all retention periods specified in these disposal authorities must be automatically tracked and the disposal process initiated only after the last of all of the retention dates have been reached.

By contrast, a system that adheres to MoReq2010®'s (2010) principle that "classification determines destiny," closely associates classification with retention and disposal. Following this principle, each class has an associated disposal schedule and each record inherits its disposal schedule by default, from its class. A record is subject to no more than one disposal schedule at a time, but the default disposal schedule inherited from its class can be overridden. Each record within an aggregation may have a classification different from other records and, therefore, be due for disposal at different times. Following the principle of "bottom-up destruction," an aggregation can only be disposed of when all of its contents have been destroyed and the aggregation is closed. Aggregations need not have disposal schedules; only one disposal schedule is required—the one associated with the record (27).

In practice, probably no more than 5 percent of all digital records created or received by organizations ends up in classified aggregations in record-keeping systems. The rest are stored, unclassified, on network drives, in email folders and, increasingly, in the cloud (Warland and Mokhtar 2012). Organizations seeking to extract knowledge from big data and legal firms seeking to locate relevant documents during a review process are investigating new technologies to make the process of sorting information less taxing; one such methodology is predictive coding. Predictive coding is

> the use of machine learning technologies to categorize an entire collection of documents as responsive or non-responsive, based on human review of only a subset of the document collection. These technologies typically rank the documents from most to least likely to be responsive to a specific information request. This ranking can then be used to "cut" or partition the documents into one or more categories, such as potentially responsive or not, in need of further review or not, etc. (Austin 2010)

The goal of predictive technology in eDiscovery remains the same as described in 2010, but technology and the view of the courts have evolved in recent years.

In a 2014 eDiscovery decision in the case of *In re Domestic Drywall Antitrust Litigation*, US District Judge Michael Baylson emphasized that counsel should use predictive coding and other "computer based programs" to help prepare their cases for trial (Favro 2014).[1] This same technology is increasingly used to automate electronic records management processes (Skamser 2013).

Records management standards and guidelines for electronic systems

The first step in understanding the challenges posed to retention and disposition in a cloud environment is to identify the functional requirements systems should possess to control retention and disposition of records hosted within the enterprise. Retention and disposition functional requirements for electronic records management were extrapolated from the following standards documents: ISO 15489 on Information and Documentation—Records Management (parts 1 and 2); ISO 23081 on Information and Documentation—Records Management Processes—Metadata for Records (parts 1, 2, and 3); ISO 16175 on the Principles and Functional Requirements for Records in Electronic Office Environments (parts 1, 2 and 3); Department of Defence (DOD) Electronic Records 5015.2 on Management Application Design Criteria Standard; and MoReq 2010®.

Records systems are designed specifically to manage records, either by hosting them in a dedicated repository or by controlling records residing in another repository. According to ISO 15489–1,

> record systems should be capable of facilitating and implementing decisions on the retention or disposition of records. It should be possible for these decisions to be made at any time in the existence of records, including during the design stage of records systems. It should also be possible, where appropriate, for disposition to be activated automatically. Systems should provide audit trails or other methods to track completed disposition actions.

The term retention in relation to electronically stored information (ESI) is the act of storing electronic information for a specified, predetermined period based on its value. The retention period is based on several factors, including the organization's operational needs; governing statues, laws, and regulations; legal issues such as the duty to preserve records for current or future audits; and historical or research needs. The organization's official policy for retention is expressed in the form of a records retention schedule and supporting procedures.

According to ISO 15489–2, any records created or captured need to have a retention period assigned so it is clear how long they should be maintained. All records within a records system should be covered by some form of disposition authority, from records of the smallest transactions to the documentation of the system's policies and procedures. Retention periods should be stated clearly and disposition triggers clearly identified. For example, "destroy x years after audit" or "transfer to the archives x years after last transaction completed." As ISO 15489–1 specifies,

> records systems should be designed so that records will remain authentic, reliable, and useable through any kind of system change, including format conversion, migration between hardware and operating systems or specific software applications, for the entire period of their retention . . .
>
> When a records system is discontinued or decommissioned, no further records may be added to the system, although they should continue to be accessible. Records may be removed from the system in accordance with retention and disposition guidelines in force, or with conversion and migration strategies. The process of discontinuing systems should be documented, as such detail will be required to maintain the authenticity, reliability, usability and integrity of records still held within that system, including conversion plans or data mapping.

Not all records reside in dedicated records systems. Some reside in electronic document and records management systems, enterprise content management systems, email systems, systems specific to the organization's business, and in a variety of cloud hosted services such as social media, cloud storage, and business applications. According to ISO 16175, Module 3, business systems must either prevent the destruction or deletion of electronic records and associated metadata, alone or in conjunction with other systems, except when records are legally authorized for disposition. Business systems must also support the disposition of records in compliance with disposition authorization regimes, which includes the following:

- allowing the definition of disposition classes, which can be applied to electronic records, either through the internal functionality of the business system software or via an automatic or manual external mechanism;
- ensuring the definition of each class includes a disposition trigger, a retention period, and a disposition action;
- supporting the following disposition actions: review, export, transfer, and destruction; and
- allowing retention periods to be defined from one day to an indefinite length of time.

Additional business system functional requirements specified in ISO 16175 include allowing disposition classes to be applied to records and associated metadata and where applicable to aggregations of electronic records; recording all disposition actions in a metadata profile; allowing a disposition freeze to be placed on the electronic record, aggregations of records, and associated metadata; preventing the deletion or destruction of records subject to a disposition freeze; and providing the ability to remove a disposition freeze to a system administrator or other authorized user.

In addition, business systems must alone, or in conjunction with other systems, allow for a review of the records before the application of a disposition action. Disposition metadata can be used to trigger the automated processes and should be retained for electronic records that have been transferred or destroyed. Finally, according to ISO 16175, the system should be able to produce a report

detailing the disposition activity, identifying records that were disposed of and those that were not successfully destroyed.

Cloud service retention and disposition functional capabilities

The functional requirements described in the previous section were analysed and then categorized according to actions related to the disposition authorities. A questionnaire (shown in Table 1) was devised to evaluate cloud services. This questionnaire assumes that the system under review contains records and that both a classification scheme and disposition authority are in place.

The task of determining the existence of retention and disposition functional requirements in cloud systems is complicated by the variety of cloud service models (for example, infrastructure as a service (IaaS), PaaS, and SaaS), cloud deployment models (for example, private cloud, public cloud, and hybrid cloud), and cloud vendors (for example, IBM, AMAZON, and Rackspace). To complicate matters further, participation in public social media results in content stored in social networks in the cloud.

A 2013 Forrester consulting survey of 154 US IT decision makers at 500+ employee companies asked the question: "What best describes your organization's current use/implementation of cloud services?" (Forrester 2013). SaaS was selected by 78 percent of the respondents, storage/backup as a service by 75 percent, and disaster recovery as a service by 70 percent of the respondents. Intelligence/analytics as a service was selected by 67 percent of the respondents, and business process as a service by 62 percent. One of the services missing from the responses to the survey was records management in the cloud. When employed in a private cloud, even one provided by a third party, this is the solution that provides the greatest degree of control over an organization's records.

More than twenty cloud services, shown in Table 2, were investigated to determine the retention and disposition functionality of each.

One of the products, HP TRIM (now HP Records Manager) was deployed as a "solution-as-a-service" to make the management of government records by the Oregon secretary of state's office more transparent. The product is designed to the international records management standard, ISO 15489:2001, and to elements of ISO 16175: *Principles and Functional Requirements for Records in Electronic Office Environments* and is DOD 5015.2 certified. Housing this solution in a private government cloud hosted by Synergy Data Center and Services in Oregon required the services of a technology integrator, Arikkan Incorporated (http://www.autonomy.com/products/hp-records-manager).

Retention and disposition functionality integrated into the offerings of the remaining cloud providers is less robust, as would be expected. Extensive examination of publicly available information and personal contacts with several cloud vendors revealed answers to some of the questions on the research and development functional requirements survey. For example,

Table 1: Questionnaire of retention and disposition functional requirements (for use when evaluating specific cloud products/services)

No.	Questions	Yes	No	Do not know
	Privacy and Security Considerations			
1	Does the vendor allow independent audits of systems and processes?	☐	☐	☐
2	Is the content encrypted when in transit to the cloud?	☐	☐	☐
3	Is the content encrypted when at rest in the cloud?	☐	☐	☐
4	Are the physical servers located within a jurisdiction approved for your organization?	☐	☐	☐
5	Are the backup servers located within a jurisdiction approved for your organization?	☐	☐	☐
	Establishing Disposition Authorities			
6	What indexing capability is supported (can it accommodate customers' taxonomy for indexing)?	☐	☐	☐
7	Can retention periods be applied?	☐	☐	☐
8	Can destruction be automated?	☐	☐	☐
	Applying Disposition Authorities			
9	Can a disposition authority (retention and disposition specifications) be applied to aggregations of records?	☐	☐	☐
10	Can records be locked down for viewing only?	☐	☐	☐
11	Can records be retained indefinitely?	☐	☐	☐
12	Can records not in an aggregation be destroyed at a future date?	☐	☐	☐
13	Can records not in an aggregation be transferred at a future date?	☐	☐	☐
	Executing Disposition Authorities			
14	Can records be deleted according to the retention/disposition schedule?	☐	☐	☐
15	Can backups be deleted according to the retention/disposition schedule?	☐	☐	☐
16	Are users alerted to conflicts related to links from records to be deleted to other records aggregations that have different records disposition requirements?	☐	☐	☐
17	If more than one disposal authority is associated with an aggregation of records, can these multiple retention requirements be tracked to allow the manual or automatic lock or freeze on the process (for example, freeze for litigation or freedom of information request)?	☐	☐	☐
	Documenting Disposal Actions			
18	Are disposal actions documented in process metadata?	☐	☐	☐
19	Can all disposal actions be automatically recorded and reported to the administrator?	☐	☐	☐
	Reviewing Disposition			
20	Are electronic aggregations presented for review along with their records management metadata and disposal authority information so both content and records management metadata can be reviewed?	☐	☐	☐
21	Can records be marked for destruction, transfer, and further review?	☐	☐	☐
22	Are all decisions made during review stored in metadata?	☐	☐	☐
23	Can the system generate reports on the disposition process?	☐	☐	☐
24	Is the ability to interface with workflow facility to support scheduling, review, and export transfer processes provided or supported?	☐	☐	☐
	Integration			
25	Is the metadata schema compatible with other systems, such as enterprise content management or records management systems?	☐	☐	☐

Table 2: Cloud services investigated for this study

Cloud Storage	RM Software and add-ons	IaaS	Litigation Support and E-Discovery
Dropbox for Business	Collabware	Century Link (Tier3)	NextPoint
Egnyte	Gimmal	GoGrid	CloudNine
One Drive for Business	HP Trim	Rackspace	
Archiving Solutions	**Collaboration / Content Mgmt**	**Long-term Digital Preservation**	**Backup and Data Protection**
ArchiveSocial	SharePoint Online	Archivematica	CrashPlan
Google Vault (email and chats)	Office 365/ Exchange/Linc Online	Preservica	HP Autonomy Live Vault
Symantec Enterprise Vault			
Smarsh			

- Rackspace (http://www.rackspace.com/) provides cloud-enabled managed hosting of the public and private clouds it designs, builds, and runs for clients. Some of the information gathered that applies to retention and disposition relates to its email hosting solution. Nine copies of each message are held across multiple data centres. Users have access to archived email without having to ask the IT team. Users can locate and recover deleted emails.
- Smarsh (http://www.smarsh.com/) provides archiving and compliance solutions to archive email, social media content from public and private social platforms, internal communications, mobile communications, and web content. Smarsh supports e-discovery searches or more advanced supervision workflow, can automate and implement legal holds and retention policies, and enforces internal governance policies for recordkeeping, supervision, and data protection. Rules can be created and then configured to take automatic action (for example, flag, classify, delegate, and apply a legal hold or a retention policy) on messages that match the criteria.
- CrashPlanPro (http://www.code42.com/business/) provides backup services for business. CrashPlan works across platforms and operating systems to backup data automatically. Users can restore data to any device on their own. Data are secured from end to end, encrypted at the source, during transit and at rest (in storage). The data can be stored in a private cloud within the enterprise or hosted in a managed private cloud, public cloud, or hybrid cloud. Administrators can enforce data retention policies, implement legal holds, specify backup scheduling, and adjust security settings.

Summaries of all of the cloud services evaluated will be published in a separate report by the InterPARES Trust project team in late 2015.

Analysis of data gathered related to cloud services

Information gathered using the questionnaire designed to evaluate cloud services revealed the following information. Questions 1–5 relate to vendor services.

Approximately 57 percent of the cloud services encrypt content residing in the cloud, and 71 percent provide encryption for content while in transit. Approximately 50 percent allow independent audits of systems. Approximately 38 percent have physical servers located within a jurisdiction approved for the client, and approximately 33 percent have backup servers located within an approved jurisdiction.

Questions 6–8 relate to establishing disposition authorities. The cloud services explored did not refer to disposition authorities, but 71.4 percent allow retention periods to be applied. Destruction is automated in 47.6 percent of the services. Indexing capability is present in 61.9 percent of the cloud services studied.

Questions 9–13 relate to applying disposition authorities and locking down records for view only. Less than half (47.6 percent) of the services allow a disposition authority (retention and disposition specifications) to be applied to aggregations of records. Only 52.4 percent of cloud services allow records that are not in an aggregation (individual records) to be destroyed (42.8 percent) or transferred (42.8 percent) at a future date.

Questions 14–17 relate to executing disposition authorities; the dedicated records management solution as a service (HP Trim) and add-ons for SharePoint (Gimmal and Collabware) meet all of these requirements. Those that provide e-discovery or compliance services allow for the deletion of records and backups according to a retention/disposition schedule (disposition authority) and allow legal holds to be imposed. The responses in this section reveal that 76.1 percent allow records to be deleted according to a retention/disposition schedule, and 57.1 percent allow backups to be deleted according to the retention and disposition schedule. However, only 9.5 percent of the services alert users to conflicts related to links from records to be deleted to other records aggregations that have different retention requirements, and 33.3 percent allow multiple retention requirements to be tracked to allow the manual or automatic lock or freeze on the disposition process if more than one disposal authority is associated with an aggregation of records.

Questions 18–19 relate to documenting disposal actions. This functionality is rarely mentioned since the model of most cloud providers focuses on retention of content of their clients and not disposition. However, 57.1 percent of the services document disposal actions in process metadata, and 57.1 percent automatically record disposal actions and report them to the administrator. In some cases, the metadata exported is descriptive and does not include operational metadata added while in the custody of the cloud provider.

Questions 20–24 relate to reviewing disposition. Dedicated records management solutions will possess the functionality that allows electronic aggregations, their records management metadata, and disposal authority to be reviewed and records to be marked for destruction, transfer, or further review; they will also store decisions in metadata. Most other systems will generate reports, and a few can also interface with a workflow facility. Only 19 percent of the solutions reviewed preset electronic aggregations, their metadata, and disposal authority information to be reviewed; 28.6 percent allow records to be marked for

destruction, transfer, or further review; 23.8 percent store all decisions made during the review in metadata; 61.9 percent provide system-generated reports on the disposition process; and 38 percent provide the ability to interface with a workflow facility to support scheduling, review, and export transfer processes.

Question 25 is related to integration. At least one of the vendors expressed frustration when discussing the metadata schema used in their products since no one industry standard exists. Only 33.3 percent of the services reviewed indicated they use a metadata scheme compatible with other systems, such as enterprise content management systems or records management systems. In some instances, third-party providers develop connectors that allow integration of cloud services with other products. One example is a connector called Vega Unity available from Vega, a consulting firm, to merge Salesforce Cloudbase with ECM repositories, file systems, databases, and workflow systems. In another case, Preservica includes multiple connectors to allow content to be ingested from ContentDM, DSpace, Outlook, Lotus Notes, and SharePoint.

Records and information professional user survey

The third and final data gathering exercise involved creating a web-based survey and inviting records and information managers to participate. The survey was opened from 9 February to 29 March 2015, and 168 useable responses were collected.

General participant information

Records managers comprised 60.84 percent of the respondents, followed by information governance professionals at 10.24 percent. Business executives, archivists, and information technology specialists each made up 2.41 percent of the total respondents, followed by information officers at 1.81 percent and legal professionals at 0.60 percent.

The majority of respondents, 37.13 percent, work in the government sector followed by professional and technical services at 8.98 percent and finance and industry at 8.38 percent. Additional industries represented are education (5.39 percent); mining, quarrying, and oil and gas extraction (5.39 percent); construction and manufacturing (4.19 percent); healthcare (2.99 percent); wholesale trade/retail trade (1.80 percent); and media arts and entertainment (0.60 percent). Organizations with more than 5,000 employees made up 26.67 percent of the respondents followed closely by those with 1,000 to 5,000 at 24.24 percent. The remaining 49.09 percent of respondents were employed in organizations with less than 1,000 employees.

Responses of current cloud service users

Of the 168 respondents who participated in the survey, ninety-seven (57.74 percent) said their organization used cloud services, forty (23.81) said it did not, twelve (7.14 percent) admitted they did not know, one (0.60 percent) declined to answer, and eighteen (10.71 percent) selected the "other" option. Other responses included statements that such use is not intentional, the cloud

is used on a limited basis, and the organization is currently considering cloud services.

The use of cloud products and services is a recent phenomenon according to the respondents, 56.82 percent of which indicated their organization had engaged in such use between one and three years, followed by 13.64 percent who have been employing cloud products and services less than one year. Only 25 percent have employed cloud services more than three years.

The types of cloud models in use vary. A private cloud is used by the majority of respondents (36.14 percent), followed by a hybrid model comprised of a private third-party hosted cloud and a public cloud (19.28 percent), a hybrid model comprised of a private enterprise hosted cloud and a public cloud (18.07 percent), a public cloud (12.05 percent), a government cloud (8.43 percent), and a community cloud (2.41 percent). The responses to questions related to general retention and disposition issues are illustrated in Figures 1–4.

When asked if the organization had performed any dispositions on its content in the cloud, the majority (53.75 percent) responded no, 27.5 percent did not know, 1.25 percent declined to answer, and only 17.5 percent stated yes. One reason that disposition may be problematic for these respondents is the fact that 49.37 percent did not include retention and disposition considerations in the initial decision to use specific cloud services and another 20.25 percent did not know if such considerations were made.

Figure 1: Percentage of respondents indicating their organization employs cloud services that have a retention and disposition policy in place

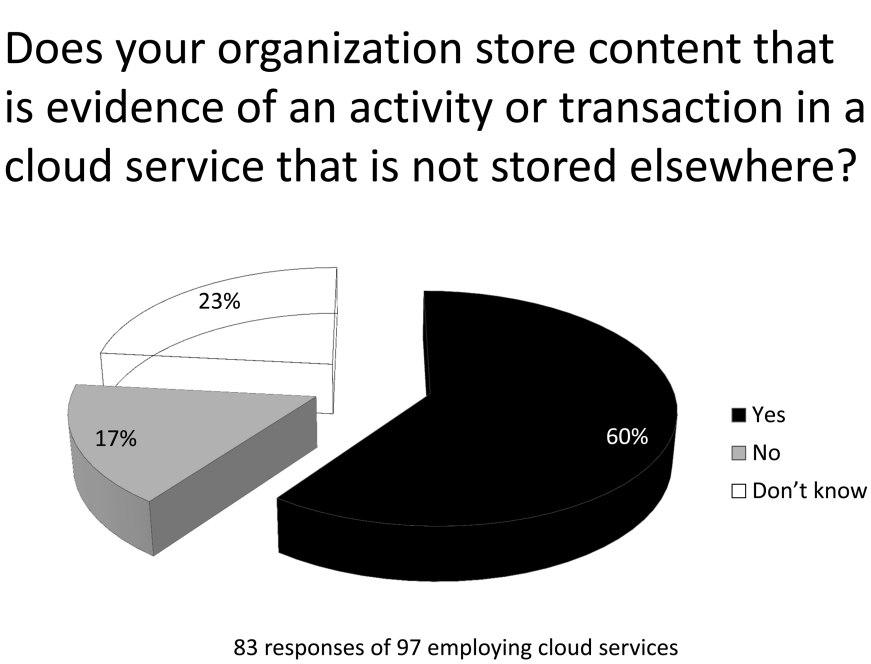

Figure 2: Percentage of respondents indicating their organization employs cloud services that store records in the cloud not stored elsewhere

Of the eighty individuals responding to the question whether vendor terms and conditions were consistent with their organization's goals and objectives for retention and disposition, 31.25 percent answered yes, 17.50 percent answered no, and 51.25 percent stated they did not know or declined to answer. The comments of two respondents indicate a position that could be taken to mitigate risk related to retention and disposition. Vendors that do not support retention and disposition are not considered.

Many of the survey questions related to disposition authorities and actions were replicas of those asked of cloud vendors on the questionnaire shown in Table 1. Responses to the questions are summarized in the following list.

- *Privacy and security considerations.* Almost 40 percent of the cloud vendors allow independent audits of their systems and processes. Over 49 percent state content is encrypted when in transit to the cloud, and over 40 percent state that content is encrypted when at rest in the cloud. Physical servers are located within a jurisdiction approved by the client for more than 53 percent of the organizations, and backup servers are located within an approved jurisdiction for more than 50 percent.
- *Establishing disposition authorities.* Respondents indicated the following indexing capabilities supported by the cloud services they employ: metadata schema (50 percent), document naming conventions (45.16 percent), classification codes (38.81 percent), taxonomies (29.03 percent), and retention periods (24.19 percent).

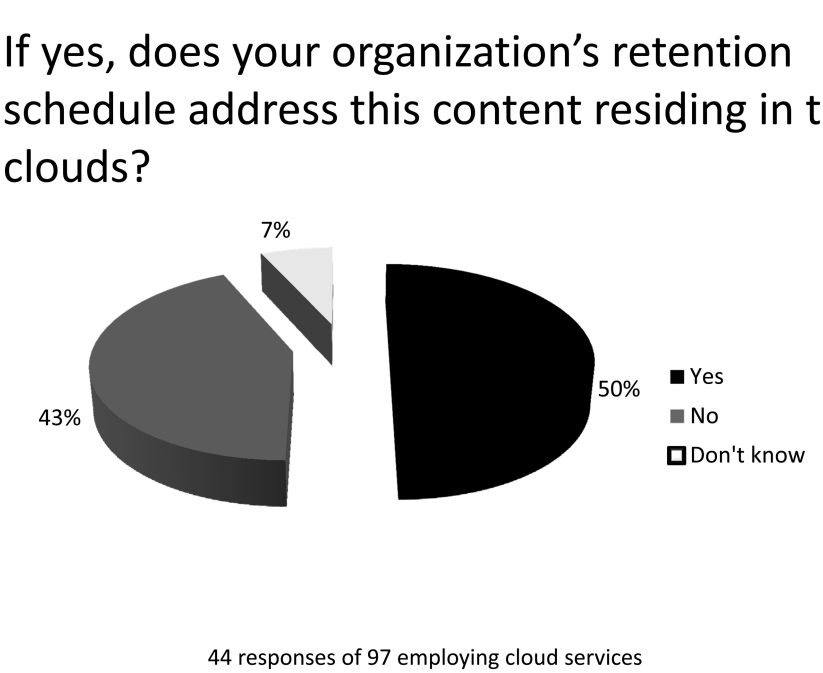

Figure 3: Percentage of respondents indicating their organization employs cloud services that address records in the cloud within their retention schedules

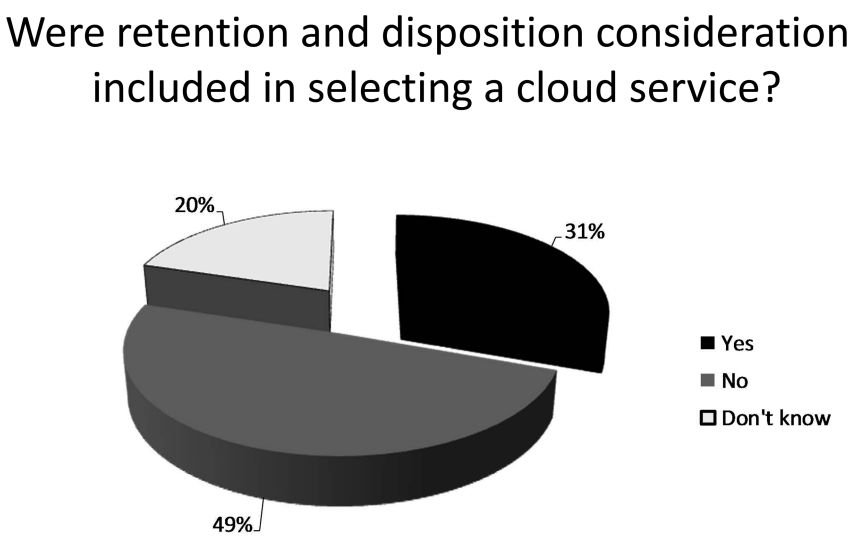

Figure 4: Percentage of respondents indicating their organization employs cloud services that include retention and disposition considerations when selecting cloud services

- *Applying disposition authorities*. Retention and disposition specifications can be applied to aggregations of records according to 25 percent of the respondents to this question, and records can be locked down for viewing according to 34.78 percent of the respondents. Records can be transferred to other systems from 56.16 percent of the cloud services in use.
- *Executing disposition authorities*. More than 45 percent of respondents indicated that records can be deleted according to a retention/disposition schedule, but only 24.29 percent indicated that backups could be deleted according to the same schedule. Twenty-eight percent stated that the administrator could change/override the disposition action, but another 59.70 percent did not know if this was possible for their cloud service provider.
- *Documenting disposal actions*. Almost 15 percent of respondents stated that disposal actions are documented in processes metadata; however, slightly over 66 percent did not know if this was the case in their organization. Disposal actions could be automatically recorded and reported to the administrator according to 21.74 percent of the respondents, but almost 58 percent did not know if this was the case for their organization.
- *Reviewing disposition*. When asked if disposition notifications are presented to the administrator to allow a review of content and records management metadata before disposition, 22 percent said yes, but 55.88 percent did not know. Almost 12 percent of the respondents stated that all decisions made during review are stored in metadata, but 73.91 percent did not know if this was the case within their organization. The system could generate reports on the disposition process for 18.84 percent of the respondents, but 60.87 percent of the respondents did not know if this was possible within the cloud services employed by their organization.
- *Integration*. When asked if the metadata schema was compatible with other systems, 29.17 percent said yes, but 57.75 percent did not know. In response to a similar question, 26.39 percent of respondents stated it was possible to integrate the cloud provider's system with other systems, such as an ECM or a records management system, but 63.89 percent did not know if this was possible for their organizations.

Conclusion

When considering the implications of the use of cloud services on the organization's retention and disposition process, the best approach is a strategic one. Begin the way you would if all records and information were stored on premise. Understand the business goals that can be achieved by using cloud services. Then consider the records and information generated. Develop a method to appraise the value of all information. Determine a process to classify information to assign retention periods and develop a disposition authority (retention and disposition schedule).

Once you understand the business goals and cloud services to be used (or already in use), investigate each of the cloud options using a questionnaire

similar to that introduced in Table 1. Analyse the data gathered. Consider carefully the potential risks and then decide if you will accept them, mitigate them, or avoid them completely. One way to mitigate risk presented by posting content to social media networks, for example, is to employ the services of a cloud-archiving solution to capture and manage the social media content in a way that enables your organization to comply with governing laws and regulations.

Records residing in a cloud environment must be captured, managed, preserved, and made accessible according to the organization's records management policy for "all" records. Functional requirements presented in de jure and de facto standards such as ISO 15489 and DOD 5015.2 apply to systems used to manage records whether they reside within the enterprise or in the clouds. However, cloud vendors may not meet all of the organization's retention and disposition requirements.

Even when records are under the control of a cloud vendor, the organization is ultimately responsible and accountable for managing its own records. Guidance for managing records is available from various sources, including professional associations. It is the client's responsibility to determine if a specific cloud provider meets their needs. This task is complicated by the services offered (for example, SaaS, IaaS, and PaaS), cloud environments (for example, public, private, and hybrid), and vendors (for example, IBM, Amazon, Microsoft, and Rackspace). When surrendering control of records and information to a cloud service provided by a third party, due diligence must be paid to identifying the appropriate mix of cloud services and providers.

Further research

The InterPARES Trust retention and disposition in a cloud environment research project is ongoing. Three separate research methods were employed to gather data related to retention and disposition functionality from existing standards and guidelines, from cloud vendor publications and interviews, and through a survey of records and information professionals. This study provides a valuable glimpse into the current landscape related to retention and disposition functional requirements offered within various cloud services. However, there are limitations to this study. The major issue is that the gap analysis is inconclusive for these reasons:

- *Cloud vendor questionnaire.* As a result of the reluctance or inability of cloud vendors to provide answers to questions about retention and disposition functionality within their offerings, capabilities may exist that have not yet been identified. In addition, since we first initiated this study, the vocabulary of many cloud vendors has broadened to include records management terms, retention and disposition features are now being offered by some cloud vendors, and technologies to integrate some cloud solutions with enterprise content management and/or records management systems are being developed. A similar study may reveal a very different landscape in the near future.

- *Records and information management user survey.* Another concern is the number of "don't know" responses to many of the questions in the records and information management user survey. For example, although 28.17 percent of the seventy-one respondents to the question about the existence of metadata schema compatible with other systems stated yes, 57.75 percent did not know. When asked if destruction can be automated, only 19.44 percent said yes, but 62.5 percent said they did not know. And when asked if backups could be deleted according to the retention/disposition schedule, 24.29 percent said yes, but 64.29 percent did not know. Several of the responses indicated a passive stance by records managers in that they stated they were not invited or did not have a seat at the table. With the current emphasis on information governance and the role records managers can play within the information governance process, a similar study in another year or two may provide fewer "don't know" responses.

Investigation into functionality built into cloud services is ongoing as new products and services are introduced and existing products and services are enhanced. In addition, case studies will be developed to describe successful cloud implementations models that enable organizations to retain and dispose of cloud-based records according to retention and disposition functional requirements.

Note

1. *In re Domestic Drywall Antitrust Litigation*, MDL No. 2437, 13-MD-2437 (E.D. Pa. 2014), http://www.paed.uscourts.gov/documents/opinions/14d0375p.pdf.

References

ARMA International. 2010. *ARMA International Maturity Model for Information Governance*. http://www.arma.org/r2/generally-accepted-br-recordkeeping-principles/metrics/metrics-retention.

Askhoj, Jan, Shigeo Sugimoto, and Mitsuharu Nagamori. 2011. "Preserving Records in the Cloud." *Records Management Journal* 21 (3): 175–87. http://dx.doi.org/10.1108/09565691111186858.

Austin, Doug. 2010. "eDiscovery Trends: What the Heck Is "Predictive Coding"?" eDiscovery Daily Blog, 15 December. http://www.ediscoverydaily.com/2010/12/ediscovery-trends-what-the-heck-is-predictive-coding.html.

Babcock, Charles. 2013. "Gartner: 50% of Enterprises Use Hybrid Cloud by 2017." Information Week, 1 October. http://www.networkcomputing.com/cloud-infrastructure/gartner-50–of-enterprises-use-hybrid-cloud-by-2017/d/d-id/1111769?

Bashir, Masooda N., and Jay P. Kesan. 2011. "Privacy in the Cloud: Going Beyond the Contractarian Paradigm." Annual Computer Security Applications Conference, Orlando, FL. https://acsac.org/2011/workshops/gtip/Bashir.pdf.

Blair, Barclay T. 2010. "Governance for Protecting Information in the Cloud." In *Making the Jump to the Cloud?: How to Manage Information Governance Challenges*, 1–4. Overland Park, KS: ARMA International. http://www.arma.org/docs/hot-topic/makingthejump.pdf.

Dutta, Amab, Chao Alex Peng Guo, and Alok Choudhary. 2013. "Risks in Enterprise Cloud Computing: The Perspective of IT Experts." *Journal of Computer Information Systems* 53 (4): 39–48.

Favro, Philip. 2014. *Breaking News: Court Touts the Importance of Predictive Coding in Preparing for Trial*. Recommind, 20 May. http://www.recommind.com/blog/2014/05/20/breaking-news-court-touts-importance-predictive-coding-preparing-trial.

Forrester Consulting. 2013. *Building for the Future: What the New World of Cloud IT Means for the Network*. Cambridge, MA: Forrester Consulting.

Franks, Patricia C., and Alan Doyle. 2014. "Retention and Disposition in the Cloud: Do You Really Have Control?" *Proceedings of International Conference on Cloud Security Management ICCSM-2014*, The Cedars, University of Reading, Reading, UK, 52.

Géczy, Peter, Noriaki Izumi, and Kōiti Hasida. 2013. "Hybrid Cloud Management: Foundations and Strategies." *Review of Business and Finance Studies* 4 (1): 37–50.

Grounds, Alison A., and Benjamin W. Cheesbro. 2013. "Cloud Control: eDiscovery and Litigation Concerns with Cloud Computing." *Computer and Internet Lawyer* 30 (9): 23–31.

Hoke, Gordon E. J. 2011. *Challenges to Governing Remote Information*. Baseline, 4 October. http://www.baselinemag.com/c/a/IT-Management/Challenges-to-Governing-Remote-Information-709978/.

HP Autonomy. 2013. *Best Practices for Cloud-Based Information Governance*. HP Autonomy. http://www.informationweek.com/whitepaper/Infrastructure/Network-Systems-Management/making-the-move-to-the-cloud-best-practices-adv-wp1347981072?articleID = 191705703.

Ion, Iulia, Niharika Sachdeva, Ponnurangam Kumaraguru, and Srdjan Čapkun. 2011. "Home is Safer Than the Cloud!: Privacy Concerns for Consumer Cloud Storage." Symposium on Usable Privacy and Security, Pittsburgh, PA, July 2011. https://www.vs.inf.ethz.ch/publ/papers/iion-cloud-2011.pdf. http://dx.doi.org/10.1145/2078827.2078845.

MoReq2010®. 2010. *Modular Requirements for Records Systems*, versioner 1.1., DLM Forum Foundation. http://moreq2010.eu/pdf/moreq2010_vol1_v1_1_en.pdf.

Ratner, Michael. 2013. *Introduction to Object Storage and Hitachi Content Platform*. Santa Clara, CA: Hitachi Data Systems. http://www.hds.com/assets/pdf/hitachi-white-paper-introduction-to-object-storage-and-hcp.pdf.

Skamser, Charles. 2013. *Predictive Coding is Expanding to Records Management and Information Governance*. http://ediscoverytimes.com/predictive-coding-is-expanding-to-records-management-and-information-governance/.

Supreme Court of the State of New York. 2013. *In Re Search Warrants Directed to Facebook, Inc*. http://s3.documentcloud.org/documents/1209711/court-order-on-facebook-search-warrants.pdf.

Tang, Yang, Patrick P. C. Lee, John C. S. Lui, and Radia Perlman. 2010. "FADE: Secure Overlay Cloud Storage with File Assured Deletion." *Security and Privacy in Communication Networks* 50: 380–97. http://dx.doi.org/10.1007/978-3-642-16161-2_22.

Warland, Andrew, and Umi Asmá Mokhtar. 2012 "Can Technology Classify Records Better Than a Human?" Image and Data Manager, 19 December. http://idm.net.au/article/009392-can-technology-classify-records-better-human.

Weins, Kim. 2014. "Cloud Computing Trends: 2014 State of the Cloud Survey." Rightscale, 2 April. http://www.rightscale.com/blog/cloud-industry-insights/cloud-computing-trends-2014-state-cloud-survey#Hybrid-Cloud-Is-the-Strategy-of-Choice.

Archival Cloud Services: Portability, Continuity, and Sustainability Aspects of Long-term Preservation of Electronically Signed Records

Les services d'archivage dans un nuage informatique : Portabilité, continuité et durabilité: Aspects de la conservation à long terme des documents signés électroniquement

Hrvoje Stancic
Faculty of Humanities and Social Sciences, University of Zagreb
hstancic@ffzg.hr

Arian Rajh
Faculty of Humanities and Social Sciences, University of Zagreb
arian.rajh@halmed.hr

Hrvoje Brzica
Financial Agency, Croatia
Hrvoje.Brzica@fina.hr

Abstract: The authors discuss key processes needed to establish archival cloud services. This is done by examining long-term preservation mechanisms and their elements. In this context, the authors explore the electronic document safe concept and analyse two models of cloud-based digital archives. The authors propose a model of archival cloud services and discuss the portability, continuity, and sustainability aspects of long-term preservation of electronically signed records.

Keywords: archives in the cloud, electronic signatures, records portability, post-custodial paradigm, changing role of archives

Résumé : Les auteurs de cet article discutent des processus clés nécessaires à la mise en place de services d'archivage par nuage informatique. Pour ce faire, ils examinent les mécanismes de conservation à long terme et les éléments qui les composent. Dans ce contexte, ils explorent le concept de coffre de sûreté pour documents électroniques, et ils analysent deux modèles d'archives numériques basés sur un nuage informatique. Les auteurs proposent un modèle de services d'archivage par nuage informatique et discutent de la portabilité, de la continuité et de la durabilité en tant qu'aspects de la conservation à long terme des documents signés électroniquement.

Mots-clés : archives en nuage informatique, signatures électroniques, portabilité des documents, paradigme de post-surveillance, changement du rôle des archives

La Revue canadienne des sciences de l'information et de bibliothéconomie 39, no. 2 2015

Introduction

Digital records are regularly stored in digital archives—solutions that could be made fairly safe in terms of long-term preservation if made compliant with the relevant standards. However, the increasing momentum of cloud storage might downgrade the level of overall quality of digital records preservation because it mainly focuses on the accessibility of the stored records. Digital records are being increasingly timestamped or signed by (advanced) electronic signatures, and (qualified) certificates are being associated with them. This adds complexity to the preservation of their authenticity, integrity, reliability, usability, and non-repudiation. The matter of legal jurisdiction over the records stored in the cloud creates additional uncertainty. What used to be a relatively controllable environment for digital archives has become volatile with the use of cloud services. Cloud service providers (CSPs), as new players, are being introduced into the archival process that used to involve only two parties—archives and record creators. In this new landscape, the role of archives changes. While the post-custodial paradigm has shifted the role of the archives from recordkeepers to supervisors of record creators, the introduction of cloud services might push this paradigm one step further. Archives should try to influence cloud-service providers to develop services more attuned to the archival standards, while, at the same time, counsel record creators in how to approach cloud services and set up or require proper archival processes.

The aim of this article is to analyse the relevant standards, explain the elements of a public key infrastructure (PKI) in the context of long-term preservation, briefly present the concepts behind cloud solutions, discuss archiving in the cloud as a process, examine the concept of electronic document safe in the context of trusted cloud service, analyse two models of cloud-based digital archives, and, finally, building on all of this, propose a model of archival cloud services supporting the long-term preservation of electronically signed records.

Relevant standards

The International Organization for Standardization (ISO) 15489 is a basic records management standard related to the establishment of the environment for records management in public and business organizations, records management policies, internal practices, systems, training, and other mechanisms. It also defines the expected qualities of records (for electronic records and information stored in document management systems, it is expanded in ISO 15801). Standards related to ISO 15489 lead practitioners to analyse their business processes and comprehend the business context before designing records management environments and systems (ISO 26122), to design relevant metadata schemas (ISO 23081), to assess related risks (ISO 18128), and to work adequately with digitized (ISO 13028) and digital records (ISO 16175, ISO 13008) in automated system surroundings (ISO 14641). ISO 17068 states what requirements have to be met in the third party repository. In addition, there are several ISO standards concerning information security, ISO 27001 being one of them. ISO

16363 specifies practices for assessing digital repositories and systems' trustworthiness, and ISO 16919 specifies requirements for certification bodies according to ISO 17021 and ISO 16363 criteria. ISO 14721 defines the reference model of archives—that is, the system or digital repository capable of long-term preservation of information, records, and digital objects. It defines a contemporary archival environment that consists of producers, archives (system, digital repository), management, and consumers. Furthermore, it defines the basic functional model of such a system with functional entities such as ingest, archival storage, data management, system administration, preservation planning, and access. Finally, it defines logical information models of information object and information packages (the submission information package delivered by the producer or client to the archives, the archival information package preserved in the system, and the dissemination information package prepared for further usage by the designated community).

PKI elements for long-term preservation of electronically signed records

The concept of long-term preservation of digital records requires a complex digital solution (Brzica, Herceg, and Stancic 2013). The terms electronic signature, digital certificate, non-repudiation, trusted archives service, timestamp and trusted digital timestamping will be explained. However, first, for a better understanding of these concepts, the concept of PKI needs to be explained.

PKI

A PKI represents a complex information infrastructure, which is used to manage electronic identities. PKI relies primarily on asymmetric encryption. Asymmetric encryption actually relies on a mathematically related pair of keys, one called the public key and another called a private key, which are generated to be used together. The private key is kept secret and used only by its owner, while the public key is made available to anyone who wants to use it (Jacobs et al. 2003, 330–31). Modern systems can easily use keys with a length of 2,048 characters, which are impossible to break even by today's supercomputers.

Electronic signature and advanced electronic signature

There are two types of electronic signatures—basic (usually referred to just as "electronic signature") and advanced (Brzica, Herceg, and Stancic 2013). The European Telecommunications Standards Institute defines electronic signature as,

> essentially the equivalent of a hand-written signature, with data in electronic form being attached to other electronic subject data (invoice, payment slip, contract, etc.) as a means of authentication. Electronic signature is not just a "picture" of the hand written signature. It is a digital signature that uses a cryptographic transformation of the data to allow the recipient of the data to prove the origin and integrity of the subject data. (*Electronic Signature* n.d.)

European legislation in EC Directive 1999/93 on a Community Framework for Electronic Signatures states that an electronic signature needs to meet the following requirements to become an advanced electronic signature:

a) it is uniquely linked to the signatory;
b) it is capable of identifying the signatory;
c) it is created using means that the signatory can maintain under his sole control; and
d) it is linked to the data to which it relates in such a manner that any subsequent change of the data are detectable.[1]

Digital certificate and qualified digital certificate

Digital certificates are digital records used to confirm the identity of a person, an organization or a machine. A digital certificate is valid for a certain period and contains several additional elements. EC Directive 1999/93 allows issuing of the so called qualified certificate, which is based on the RFC 3039 standard (Santesson 2001) and implements the concept of non-repudiation. Annex I of the Directive sets the requirements for the qualified certificate. It must in particular include:

1. an indication that the certificate is issued as a qualified certificate;
2. the identification of the certification-service-provider and the State in which it is established;
3. the name of the signatory or a pseudonym, which shall be identified as such;
4. provision for a specific attribute of the signatory to be included if relevant, depending on the purpose for which the certificate is intended;
5. signature-verification data which correspond to signature-creation data under the control of the signatory;
6. an indication of the beginning and end of the period of validity of the certificate;
7. the identity code of the certificate;
8. the advanced electronic signature of the certification-service-provider issuing it;
9. limitations on the scope of use of the certificate, if applicable; and
10. limits on the value of transactions for which the certificate can be used, if applicable.

Timestamp and trusted digital timestamping

According to Carl Wallace, Ulrich Pordesch, and Ralf Brandner (2007, 5), a digital timestamp is an attestation generated by a time stamping authority (TSA)—a trusted service—that a data item existed at a certain time. Jasmin Ćosić and Miroslav Bača (2010, 1227–28) explain that

> time stamps are typically used for logging events, in which case each event in a log is marked with a time stamp. In file systems, time stamp may refer to the stored date/time of the file creation or modification. Trusted time stamping is the process of securely keeping track of the creation and modification time of a document. . . . Trusted TSA can be used to prove the consistency and integrity of digital evidence in every stage of its existence.

Formats of electronic signatures

In the earlier discussion, the technologies and concepts supporting trust in electronic records were explained. It was shown that the concept of an electronic signature can be viewed as the basis for developing all other technologies. Further, electronic signatures can be realized through several formats of electronic signatures—XML Digital Signature, XML Advanced Electronic Signature, Cryptographic Message Syntax Advanced Electronic Signature, and PDF Advanced Electronic Signature (Brzica, Herceg, and Stancic 2013).

Cloud services

Cloud solutions, in general, can be deployed in various forms, such as the public cloud, which can used by anyone, the private cloud for private use by a single organization, the community cloud for a group of users, or the hybrid cloud. Users of cloud solutions are offered software or applications, platform or entire environments, as well as infrastructure or entire virtual datacentres (Stancic, Rajh, and Milosevic 2012). Whether to develop a cloud-hosted system or a system on premise depends on the complexity of an organization's information technology (IT) environment, needed functionality, size of the organization, data volume, legal regulations, IT skills of the in-house experts, resources, as well as operational costs (*Cloud: On-Premise or Hybrid?* n.d.): "Applications and services need to be run where they are most efficient and not just because cost is the most attractive option. In the long-term, falling into the 'all Cloud' solution trap can prove to be more expensive, time-consuming and problematic" (*On-Premise versus Cloud-based Solutions* 2010).

Archiving in the cloud as a process

As Sue McKemmish (2013, 19) has explained,

> cloud computing offers attractive benefits including significant cost savings, efficiencies, flexibility and scalability, as well as opportunities for the innovative development and delivery of new services. It also carries significant risks associated with the security, privacy, integrity, authenticity, accessibility and digital continuity of data and records in the cloud. There are also issues relating to commercial continuity and the lack of transparency of cloud services that impact on recordkeeping and archiving.

All this raises a question of entrusting records to the cloud. To minimise risks and maximise benefits a standardised process of archiving in the cloud should be used.

A generic business process of archiving in a cloud environment should include a creator's part and a CSP's part (see Figure 1). The creator creates documents in sustainable formats and signs them electronically (1). Verification data should be preserved for later use (2). Before submitting documents to the chosen service provider, documents and signatures should be checked (3), and documents should be declared as records (4). Then the service provider ingests records (5) and processes them (6). All documents could be verified with qualified eStamp mechanism (8). Archiving should be done by creating a copy in a

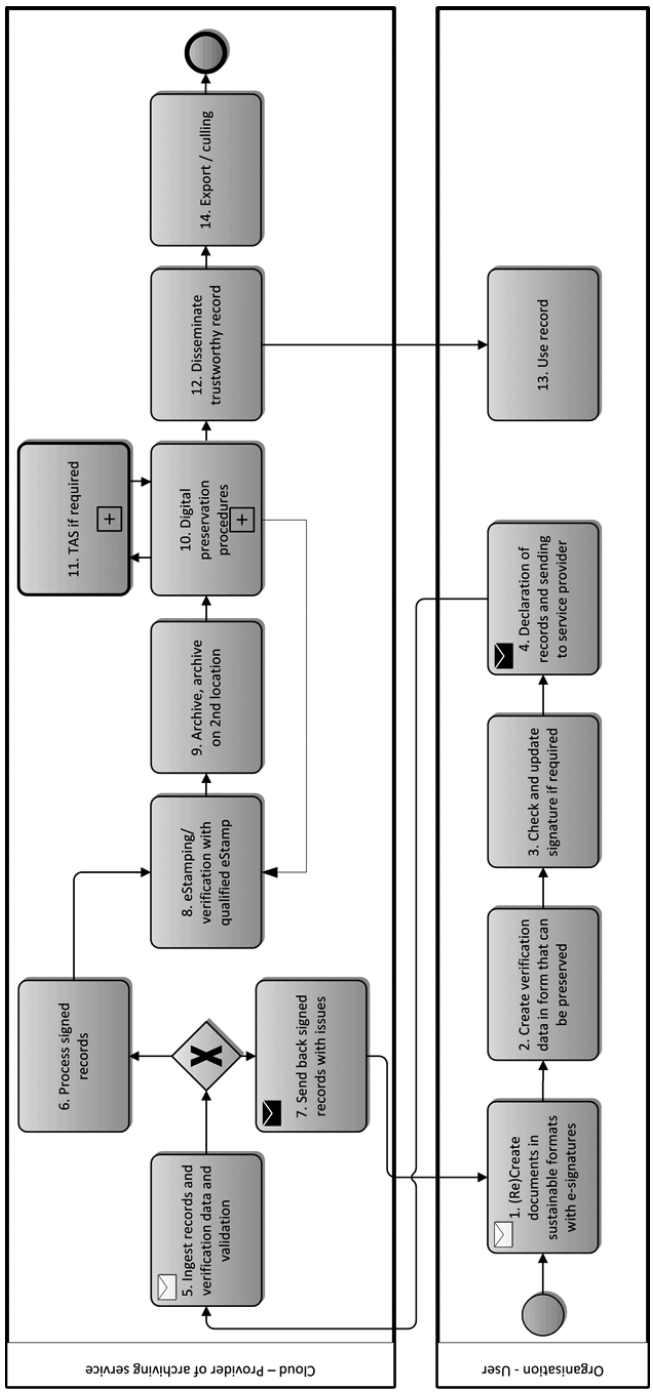

Figure 1: Generic business process of archiving in cloud environment

secondary location (9). Digital preservation procedures (10) should include file format migration, if needed. The service provider should have implemented a procedure of regular validation of signed records with a trusted archives service (TAS) emulator (11). The TAS process is treated as specific sub-process with its own mechanisms (for example, eStamping). A record is provided (12) on demand to the creator who uses it (13). The service provider should actively promote long-term preservation measures to the creator and communicate with the creator when verification/validation criteria are not met.

Jos Dumortier and Sofie Van den Eynde (n.d., 7) explain that

> a TAS must guarantee that it will still be possible to validate an archived document years after the initial archival date, even if the applications that have been used at signature creation time are no longer in use. In other words, the TAS should maintain a set of applications (viewers as well as signature validation applications) together with the corresponding platforms (hardware, operating systems) or at least an emulator of such applications and/or environment in order to guarantee that the signature of the document can still be validated years later.

Trusted cloud service: underlying concept

To better understand the complexity of building a trusted cloud service, the underlying concept of the electronic document safe (EDS) needs to be analysed. An explanation of the concept in the context of secure storage of officially issued governmental documents will show how the authenticity of digitally signed documents and their trustworthiness can be preserved over the long term.

EDS

In the context of cloud services, Peter Deussen and his colleagues (2012) define the concept of the EDS as a secure storage for official documents (for example, used as a part of citizen support services). The EDS functionalities are secure long-term storage of official electronic documents and established electronic workflows among administration, enterprises, and citizens. Employees of government agencies can store documents in the EDS of any user. On their part, users can access only their own EDS, and, for doing so, they need a special application providing secure encrypted communication and authentication. The Fraunhofer Institute has developed a version of the EDS called eSafe. It shares with the EDS the idea that documents should be stored in a trustworthy, secure way within a cloud infrastructure, but it goes further by describing a mechanism to fragment and distribute such documents among several cloud storage providers, thus making it very difficult for an unauthorized person to retrieve the original document (Breitenstrom, Brunzel, and Klessmann 2008).

Regarding the feasibility of developing EDS solutions, there are at least two hypothetical issues that need to be addressed: privacy protection and long-term service availability. Christian Breitenstrom, Marco Brunzel, and Jens Klessmann discuss the data protection issue in the context of privacy protection by analysing the requirements of storing personal data in public cloud infrastructures and using those data to interact with public sector authorities. Significantly, when an EDS

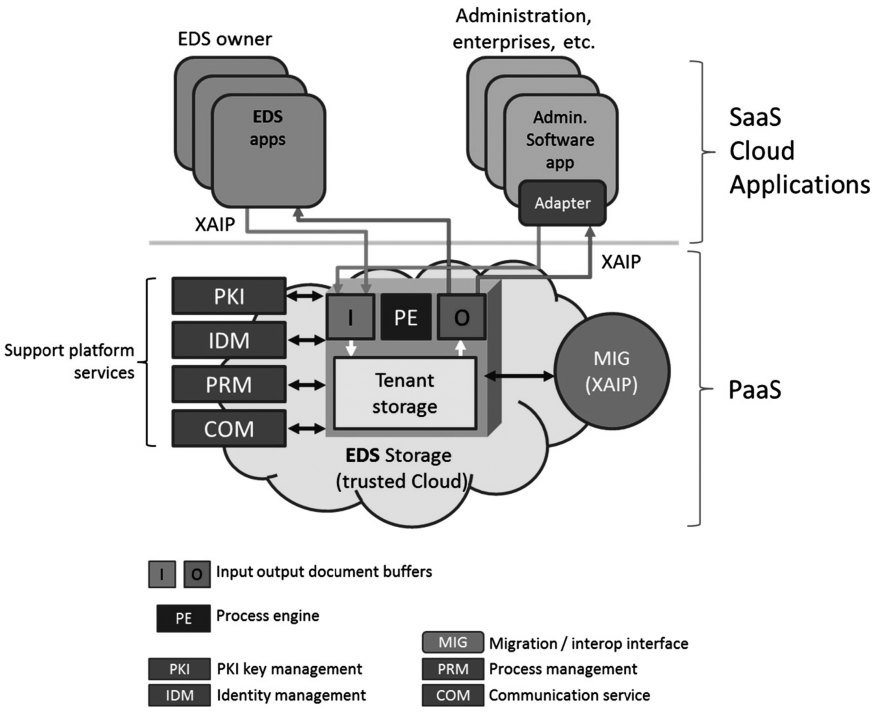

Figure 2: EDS principle architecture (Deussen et al. 2012, 45)

is established, citizens can approve that personal data are stored and processed electronically. Hence, data privacy applies only insofar as citizens have to trust the EDS provider. Since data in the EDS are encrypted and thus not visible to the provider, the "data protection barrier" can be considered to be low and the principle applied can be qualified as data anonymization.

Second, if we look at the EDS as a technical solution for supporting administrative processes, then the government agencies should guarantee and take responsibility to ensure the continuous availability of the service. For some types of records, it could be more than 100 years. Therefore, the most important question is how a private sector provider is capable of giving sufficient guarantees on its own continued existence, future business orientation, and so on. A possible solution is to maintain a governmental cloud provider as a fall-back (or business continuity) solution, while allowing public sector providers to participate in the emerging market of electronic document and records storage (Deussen et al. 2012, 43–44). Of course, to achieve this, political and budget-related decisions need to be made first. The infrastructure that needs to be developed to establish an EDS solution in the cloud is shown in Figure 2.

The main component of the EDS system is the EDS storage—that is, a cloud-based infrastructure providing storage, access, and management functions. It comprises a tenant storage containing the documents of the EDS user and

several components to handle access, authorization, and usage. Input (I) and output (O) components are responsible for managing storage of the documents requiring authorization to be ingested into an EDS or delivered to the users, while a process engine (PE) is responsible for the coordination of the actual usage. The document representation format used in the EDS is the XML-formatted Archival Information Package (XAIP). Long-term preservation data formats such as XAIP or the universal object format (UOF) usually contain a combination of data and metadata. For example, XAIP was designed for an archives system, and its structure is based on the technical directive of the Federal Office for Information Security (Preservation of Evidence of Cryptographically Signed Documents 2011).

The archival information package is an XML file that contains the data and the corresponding metadata, while the UOF stores the data and metadata, based on the Metadata Encoding and Transmission Standard, in two separate files (Potthoff, Walk, and Rieger 2013, 28). Concretely, XAIP is structured in four parts: (1) archival package header that contains information about XAIP's logical structure; (2) meta information with the description of the transactional and archiving context of the content data; (3) content data that contains encrypted documents; and (4) certificate section containing digital signatures, digital certificates, information needed to verify digital signatures, as well as digital timestamps. Thus, the last part of the XAIP structure provides the relevant information on authenticity, integrity, and trustworthiness of the archived data objects.

Document migration

In the case that, during long-term preservation, a CSP providing EDS goes out of business, the records should be transferred to another CSP providing EDS. That is why the EDS architecture has the migration interface. Since the preserved materials contain records to be used as evidence and, thus, need to be signed to ensure authenticity, migration from one EDS provider to another has to be compliant with the laws and regulations governing the authorities who have initially issued these documents and are responsible for further processing (Deussen et al. 2012, 82). The process of migration of archival objects identified by their identifier (that is, the archive object identifier) from EDS 1's tenant storage (CSP A) to EDS 2's tenant storage (CSP B) is shown in Figure 3. To achieve this, certification service, signature validation service, and encryption service are needed. In addition, any transmission protocol supporting data encryption can be used for the migration of XAIPs.

Two models of cloud-based digital archives

A model of archival cloud services supporting long-term preservation proposed later on in this article is based on two models developed in Germany and Lithuania.

Germany: Federal Office for Information Security

The German Federal Office for Information Security (Bundesamt für Sicherheit in der Informationstechnik [BSI]) developed a model of long-term preservation

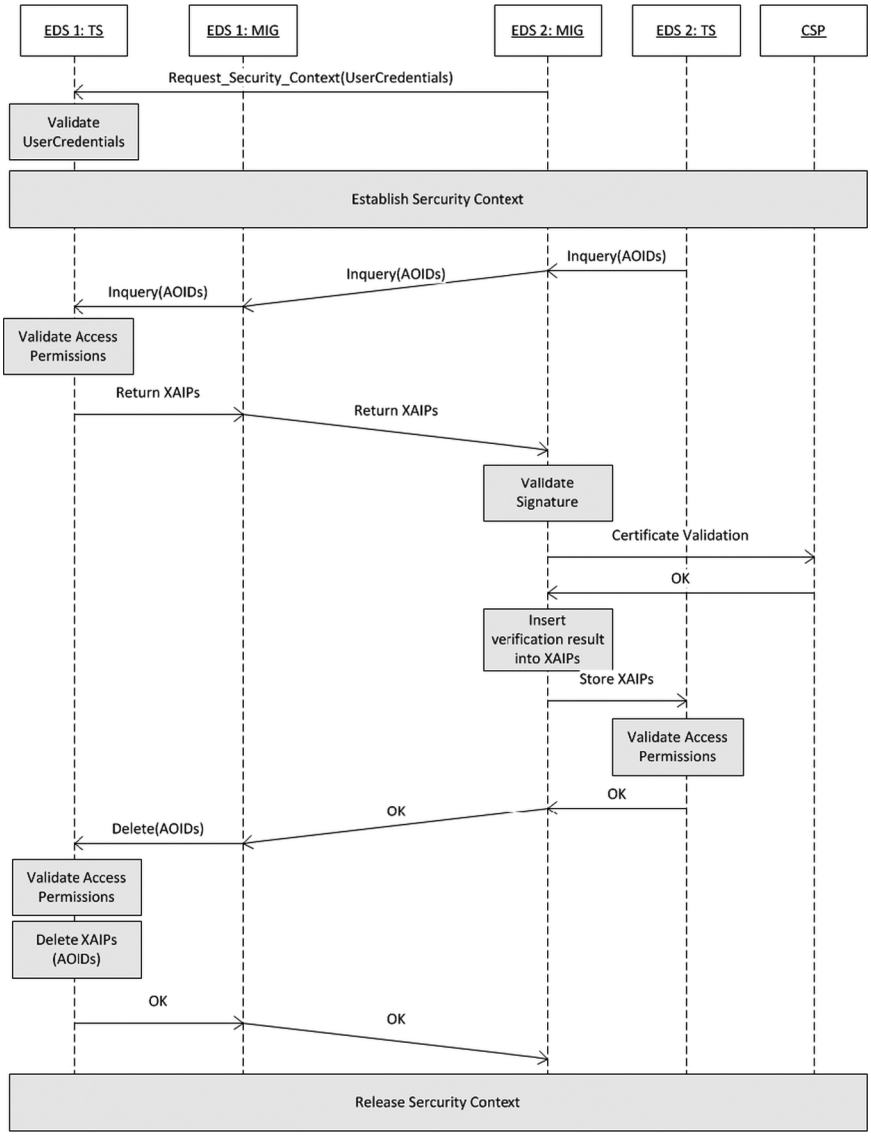

Figure 3: Document migration from EDS 1 (Provider A) to EDS 2 (Provider B) (Deussen et al. 2012, 84)

of digitally signed documents based on several ISO standards and the German Federal Archiving Act (Bundesarchivgesetz).[2] BSI has published technical guidelines explaining the architecture of the system responsible for long-term preservation—BSI Technical Guideline 03125 on the Preservation of Evidence of Cryptographically Signed Documents (BSI Guidelines) (Federal Office for Information Security 2011).

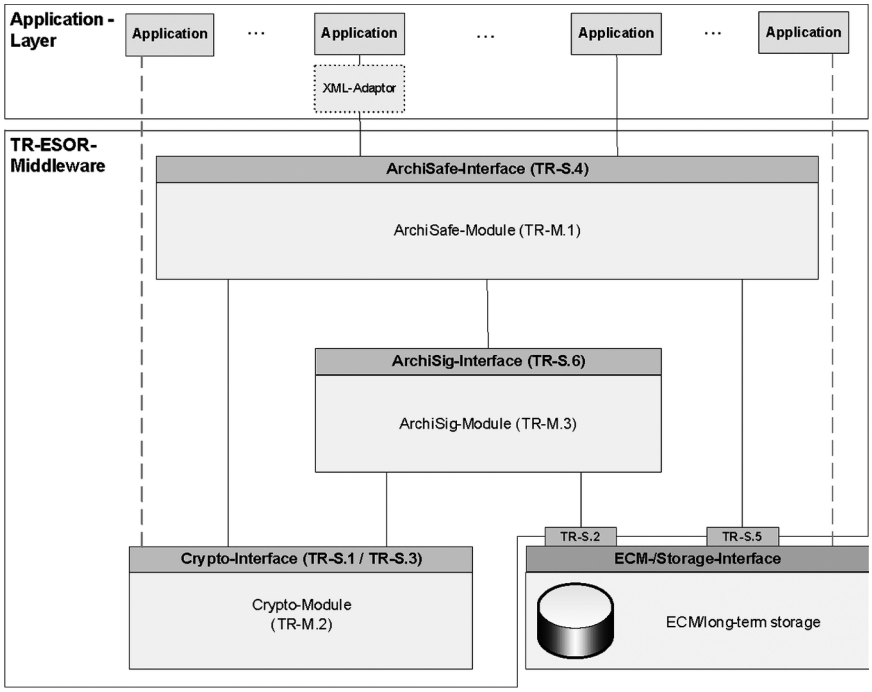

Figure 4: BSI referent architecture (Federal Office for Information Security 2011, 41)

BSI architecture comprises two main parts: (1) IT infrastructure for long-term storage and (2) IT applications archiving data and documents or working with them. The IT infrastructure used for archiving typically consists of:

- an enterprise content management / long-term storage system that includes and manages various storage media used for archiving and that guarantees the reliable and secure access to the storage media for the deposit, retrieval, and deletion of archived documents and data and
- the middleware, including the cryptographic components that support the preservation of the elements required by the laws of evidence governing the archived documents (and data). (Federal Office for Information Security 2011, 15–16) (see Figure 4).

Lithuania: electronic archives information system

The first Lithuanian system for working with electronic signatures was the e-Servicing System of the Insurers (EDAS), launched in 2007 by the State Social Insurance Fund Board of Lithuania. The EDAS system uses XAdES format of electronic signatures to sign the documents. Although metadata could be signed separately, as a sub-tree within the XML metadata file, the basic principle of the Lithuanian approach is that the metadata are integral part of the electronic document.

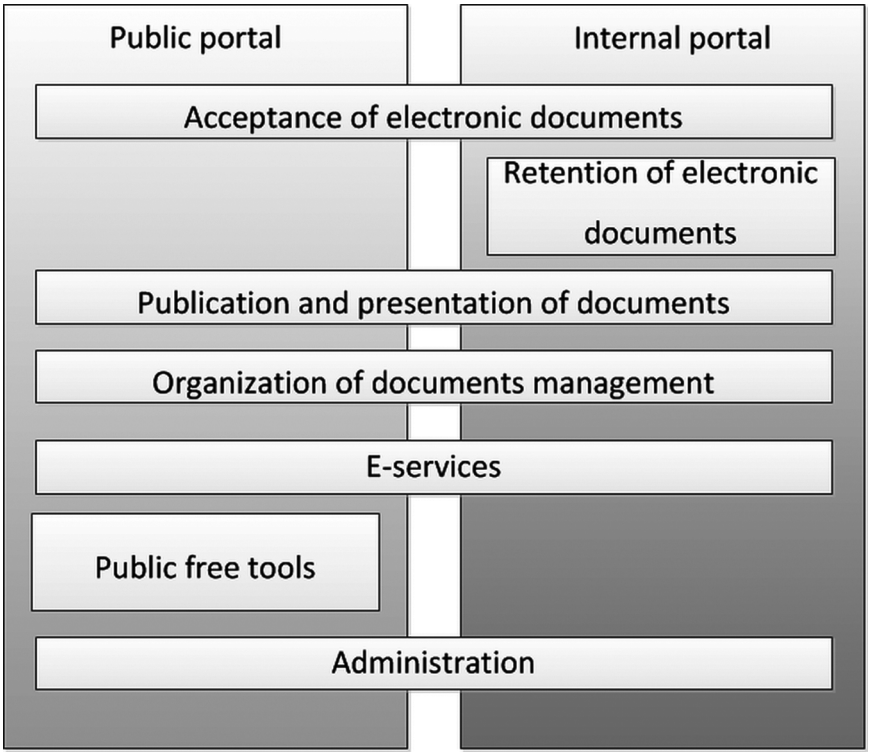

Figure 5: Modules of EAIS (Ragaisis et al. 2012)

A further step was to develop the Electronic Archives Information System (Elektroninio archyvo informacinė sistema [EAIS]) in 2011 as the final step in the Lithuanian government's preparation to fully work with e-documents (Elektroninio archyvo informacinė sistema 2011). EAIS enables archiving of the official e-documents signed with advanced electronic signature. The system ensures integrity, authenticity, non-repudiation, and usage during long-term preservation. The EAIS architecture consists of a public portal, an internal portal, and the storage of electronic documents. Storage is done in two geographically remote electronic archive data centres. Modules of public and internal portals are shown in Figure 5.

A model of archival cloud services supporting long-term preservation

The model of archival cloud services supporting long-term preservation proposed here builds upon the previously explained concept of electronic document safe and the two implementation models, taking into account the specifics of electronically signed records and archiving in the cloud environment as a business process.

System functionality

An archival cloud service supporting the long-term preservation of electronically signed records should preserve the integrity, authenticity, and confidentiality of stored records. It should make records available and usable—that is, maintain their readability. Data protection and system security should be enforced. At the same time, an archival cloud service should be expected to provide functionalities such as document creation, e-signing, archiving of non-signed and signed documents and records, records publication (with visualization of electronic signature), indexing, search and retrieval of archived records, provision of proof of evidence, preservation procedures not influencing the evidential capacity of preserved records, deletion of records, and administration.

There could be at least two expected access points to such an archival cloud service: one internal, used by governmental bodies as the creators of documents and records, and one external, used by citizens accessing the records. The Lithuanian experience with the EAIS system shows that the users would accept such a system if free software tools were offered—for example, for document preparation, e-signing, viewing, and verification of official digital documents. Those tools could be offered either as desktop applications having connectors to the cloud service or as web applications (software-as-a-service [SaaS] approach).

Automated processes, such as those for the digitization of documents to be stored in PDF/A format or those for embedding electronic signatures in the PDF/A documents, could be added to the system. An example of such an addition can be found in the e-health domain when medical documentation, such as x-ray images, need to be scanned and archived.

Key processes

To implement the system of archival cloud services supporting long-term preservation, it is necessary for the following to happen:

1. Electronic signature and timestamp should be created, verified, renewed, and stored in a safe and trustworthy way that is compliant with the relevant legal framework.
2. Data needed for later verification of electronic signature should be obtained immediately after its creation and/or verification. The verification data should be ingested along with the records.
3. All verification steps and results of the verifications should be logged and stored in the format that would guarantee non-repudiation.
4. Electronic signatures should be renewed before the expiration of the protection measures used in cryptographic algorithms. The renewal should be done according to the legal regulations and by a (semi-)automatic and economic process.
5. As a result of technological advancements, which have caused cryptographic algorithms considered strong today to be weak in the future, digital timestamps should be added. This way, to resign electronically signed records, it would be sufficient to certify them with a qualified timestamp, which consists of at least one qualified electronic signature.

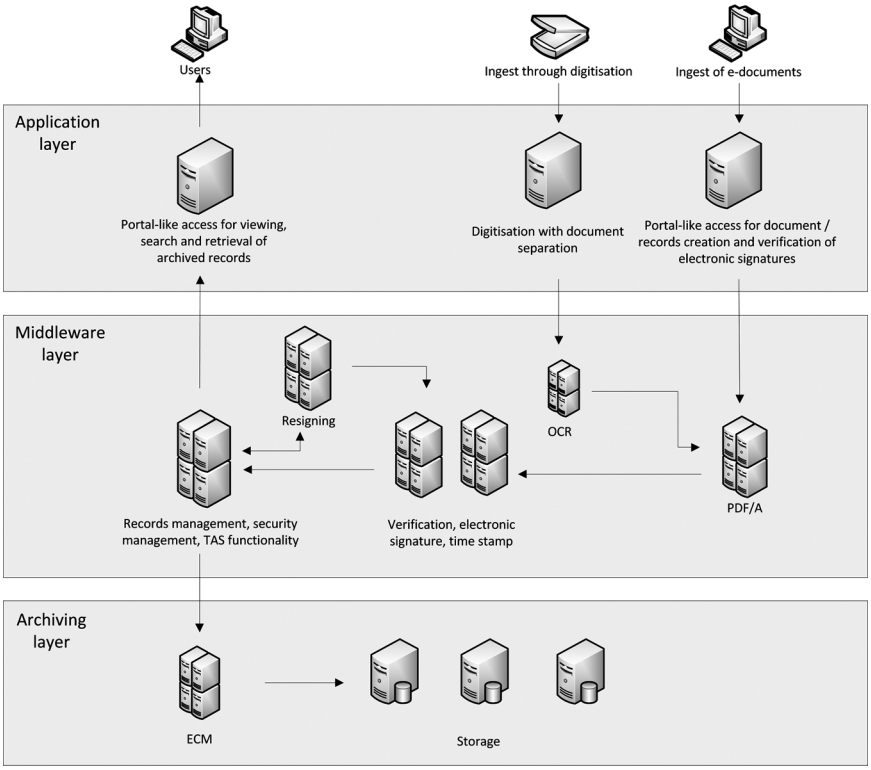

Figure 6: Architecture of archival cloud services supporting long-term preservation

6. The system components used for access to the content of the preserved records should be able to visualize electronic signatures, certificates, and the results of their verification.
7. The integrity of the non-signed records being submitted for long-term preservation need to be secured by cryptographic methods, such as hash values or electronic signatures and qualified timestamps.
8. The system should be, at a minimum, located at two geographically distant locations in case of data replication and disaster recovery situations.
9. The system should implement a trusted archives service) process—that is, ensure the processes of validation of archived, electronically signed records, years after ingest in the archives, in spite of the fact that the application solution for e-signatures and digital timestamping will be obsolete at that time.

System architecture

The architecture of the system of archival cloud services supporting long-term preservation should consist of at least three main layers: (1) application layer; (2) middleware layer; and (3) archiving layer (see Figure 6).

1. The Application layer accommodates web or portal as well as desktop applications. The applications should be used for creating documents that will be archived in the formats compliant to the standardized long-term preservation formats (for example, PDF/A of XML). They should also enable the functionality of electronic signature verification. To enable signature validation over the long-term preservation, it is suggested that the verification data of the electronic signature validity be archived—that is, proof of evidence, along with the signature—that is, with the e-signed document/record. The applications should also be used as trusted viewers. Trusted viewers are used for viewing electronically signed records, and they function as trustworthy components of the applications through which an advanced electronic signature is shown. An application layer should be used as interface, and it should provide functionalities for archiving records and enable search and retrieval of archived records and archived proof of evidence.
2. The middleware layer is a standardized and secure gateway that controls the access of applications to the archival storage. This layer logically separates applications in the application layer from long-term preservation storage. All actions, such as write, change, or delete, should be done through this layer. It is responsible for the cryptographic functions needed for preservation of evidence, such as creation of electronic signature, its verification, verification of electronically signed archival information packages, certificate validation, calculation of hash values, as well as request and verification of qualified timestamps. TAS functionality enables storage of software emulators used for verification and viewing of electronically signed records created by obsolete software. The functionality of resigning of electronically signed records can be used instead of the TAS functionality, since it enables resigning of the records using the latest signature formats. The process of resigning should be automated as much as possible.
3. The archiving layer is the layer used for long-term storage of archived records. It comprises main storage and remote storage. The technical principles of cloud storage physical organization apply here.

Discussion

The idea of an archival cloud service is very appealing due to the fact that creators of records are usually non-archival institutions that may lack infrastructure, technical capacities, and staff knowledge adequate for long-term preservation of electronically signed records. Although appealing, it is also challenging due to archival legislation, the protection of information, and the protection of national interests. As McKemmish states (2013), national legislation can be expanded to service providers regardless of the fact that the clients can be residents of other countries or in some cases even regardless of the location of data centres. It becomes more common today to use the data centres located in the creator's country so that the national legislation can be applied. Service providers may be asked to guarantee compliance with the corresponding creator's legal context.

There are four important aspects that should be considered when a creator decides to archive its holdings in the cloud: holding portability, digital continuity, environment sustainability, and warranty of compliance with corresponding legal context. Portability in this context has two meanings. First, it denotes an ability to transfer electronically signed records from a creator to the CSP's environment, during the ingest procedures, without losing reliability, authenticity, and trust in records. Suggested mechanisms include validation and verification procedures, checking file formats and so on. This aspect is not present if the creator is using CSP's SaaS concept for the creation of documents and records as an addition to, or part of, the archive-as-a-service concept because the records are already being created in the cloud.

Second, portability refers to the possibility of transferring creator's records from one CSP to another one—for example, if a CSP goes out of business, also without losing reliability, authenticity, and trust in records. Digital continuity refers to archiving electronically signed records and ensuring their usability in the time frame of their retention period and according to the business needs of the client organization (What Is Digital Continuity? n.d.; National Archives of Australia 2015). One of the available measures for ensuring digital continuity is the TAS function provided by the CSP. It should be kept in mind that the portability concept is inherent to the digital continuity concept.

Sustainability is a quality of the whole CSP's environment, and it can be ensured by technological and financial capacities and measures implemented by the CSP and assessed by the creator. For example, technology that is used for implementing a solution can be obsolescence resistant and based on well-known and robust solutions, the third-party provider can ensure adequate number of full-time employed technical support and administrative staff members, and the creator can check CSP's financial stability. Finally, for the protection of the information contained in records and their evidential capacity, the creator may consider asking CSP to store records within the creator's national boundaries.

Conclusion

The concepts of portability, continuity, and sustainability are preconditions for long-term preservation of electronically signed records by an archival cloud service at times when corresponding legal frameworks are not fully established everywhere. Achieving portability, continuity, and sustainability can stimulate a creator's trust in a CSP and, consequently, a creator's trust in its own archived records. Public trust in a records creator can ultimately depend on meeting these requirements. Portability is crucial for the successful transfer of files into a new CSP's environment, while sustainability ensures that this environment is stable. Continuity ensures maintenance of records between import and dissemination points. Stability of environment and service is twofold—it presumes technical excellence and response to changes in environment as well as financial stability of the CSP. Continuity and the technological part of the aspect of stability can be seen as further elaboration of the OAIS reference model's preservation planning function. They are conceptually similar to the preservation planning

function, but continuity is at the level of information and records and stability is above this level and expands on the capability of a system to preserve information and records according to the built-in mechanisms and methods. An example of meeting the continuity requirement is implementation of a proposal for conversion of a file format to an upgraded format submitted to the client by the CSP. An example of meeting the technological stability aspect can be switching from one conversion tool to another if the target file format is upgraded and the older conversion tool cannot do the job right. All of these criteria are not easy to define in the tendering procedures and contracts with CSPs, but information package prototypes testing in the first period of the execution of the contract, as well as periodical testing and testing after technological changes that may influence digital archival holding, can be foreseen and prescribed in such contracts. The models explained in this article, their underlying principles, and the proposed model of archival cloud services supporting long-term preservation could be used either as guidelines for choosing an archival CSP or for establishing such a cloud service.

Acknowledgements

This research was completed in the context of the international multidisciplinary research project InterPARES Trust, http://www.interparestrust.org.

Notes

1. EC Directive 1999/93 on a Community Framework for Electronic Signatures, [2000] OJ L13.
2. Bundesarchivgesetz, 1988, http://www.bundesarchiv.de/bundesarchiv/rechtsgrundlagen/bundesarchivgesetz/index.html.en.

References

Breitenstrom, Christian, Marco Brunzel, and Jens Klessmann. 2008. *White Paper: Elektronische Safes für Daten und Documente*. Berlin: Fraunhofer Institut für Offene Kommunikationssysteme. http://wwwold.fokus.fraunhofer.de/de/elan/_docs/_hpp-gruppe/esafe_white-paper_081219.pdf.

Brzica, Hrvoje, Boris Herceg, and Hrvoje Stancic. 2013. ''Long-term Preservation of Validity of Electronically Signed Records.'' In *INFuture2013: Information Governance. Zagreb: Deaprtment of Information and Communication Sciences*, ed. Anne Gilliland, Sue McKemmish, Hrvoje Stancic, Sanja Seljan, and Jadranka Lasic-Lazic, 147–58. Zagreb: University of Zagreb, Faculty of Humanities and Social Sciences.

Cloud: On-Premise or Hybrid? n.d. Bluesource Information Limited. https://d3759s1c6gf66q.cloudfront.net/u/_201211/52981741/59991441/dtwRJfTb/CloudOn-premiseorHybrid.pdf

Ćosić, Jasmin, and Miroslav Bača. 2010. ''(Im)Proving Chain of Custody and Digital Evidence Integ-rity with Time Stamp.'' MIPRO: *Proceedings of the 33rd International Convention*, 1226–30.

Deussen, Peter, Klaus-Peter Eckert, Linda Strick, and Dorota Witaszek. 2012. *Cloud Concepts for the Public Sector in Germany: Use Cases*. Berlin: FOKUS Fraunhofer Institute for Open Communication Systems.

Dumortier, Jos, and Sofie Van den Eynde. N.d. *Electronic Signatures and Trusted Archival Services.*

Electronic Signature. N.d. http://www.etsi.org/technologies-clusters/technologies/security/electronic-signature.

Elektroninio archyvo informacinė sistema. 2011. http://eais-pub.archyvai.lt/eais/.

Federal Office for Information Security. 2011. *BSI Technical Guideline 03125 on the Preservation of Evidence of Cryptographically Signed Documents*. Bonn: Federal Office for Information Security.

Jacobs, J., L. Clemmer, M. Dalton, R. Rogers, and J. Posluns. 2003. *SSCP Study Guide*. Sebastopol, CA: Syngress Publishing.

McKemmish, Sue. 2013. ''Recordkeeping and Archiving in the Cloud. Is There a Silver Lining?'' In *Information Governance*, ed. Anne Gilliland, Sue McKemmish, Hrvoje Stancic, Sanja Seljan, and Jadranka Lasic-Lazic, 17–29. Zagreb: University of Zagreb, Faculty of Humanities and Social Sciences, Department of Information Sciences.

National Archives of Australia. 2015. *What Is Digital Continuity?*

On-Premise versus Cloud-based Solutions, White Paper. 2010. GFI Software.

Potthoff, Jan, Marius Walk, and Sebastian Rieger. 2013. ''Data Management According to the Good Scientific Practice.'' In *The Fifth International Conference on Advances in Databases, Knowledge, and Data Applications*, 27–32.

Ragaisis, Saulius, Adomas Birstunas, Antanas Mitasiunas, and Arunas Stockus. 2012. ''Electronic Archive Information System.'' In *Databases and Information Systems: Tenth International Baltic Conference on Databases and Information Systems—Local Proceedings, Materials of Doctoral Consortium*, ed. Albertas Čaplinskas, Dzemyda Gintautas, Audronė Lupeikienė, and Olegas Vasilecas, 107–14. Vilnius: Zara.

Santesson, S., et al. 2001. *Public Key Infrastructure: Qualified Certificates Profile*, Internet X:509. Internet Society. http://www.internetsociety.org.

Stancic, Hrvoje, Arian Rajh, and Ivor Milosevic. 2012. ''Archiving-as-a-Service: Influence of Cloud Computing on the Archival Theory and Practic.'' In *The Memory of the World in the Digital Age: Digitization and Preservation*, ed. Luciana Duranti and Elizabeth Shaffer, 108–25. Vancouver: UNESCO.

Wallace, C., U. Pordesch, and R. Brandner. 2007. *Long-Term Archive Service Requirements*. IETF Trust. http://dx.doi.org/10.17487/rfc4810.

Public Cloud Archives: Dream or Reality?

Les archives publiques dans le nuage informatique : Rêve ou réalité ?

Anna Sobczak
Szczecin University
amsobczak@gmail.com

Abstract: The German Landesarchiv Baden-Württemberg (State Archives of Baden-Württemberg) developed its own software called Digitales Magazin (Digital Storeroom) to appraise, acquire, manage, describe, and provide access and long-term archiving of different kinds of electronic records. In an effort to develop the system further and to offer it to other state archives as well as small public or private ones, several new cooperation possibilities were formulated, involving the use of the software among other in a software as a service or developer model.

Keywords: Germany, state archives, small archives, cloud computing, services, non-commercial software, SaaS

Résumé : Les archives d'État du Land de Bade-Wurtemberg en Allemagne (Landesarchiv Baden-Württemberg) ont développé leur propre logiciel appelé Digitales Magazin (Entrepôt numérique) pour évaluer, acquérir, gérer, décrire, fournir un accès et l'archivage à long terme de différents types de documents électroniques. Dans un effort pour poursuivre le développement du système et de l'offrir à d'autres archives d'État ainsi qu'à d'autres archives publiques ou privées de petite taille, plusieurs possibilités nouvelles de coopération ont été formulées, impliquant l'utilisation du logiciel dans un modèle SaaS (software as a service) ou dans un modèle développeur.

Mots-clés : Allemagne, archives d'État, petites archives, nuage informatique, services, logiciels non commerciaux, SaaS

The aim of this article is to present models of partnerships in digital archiving and building digital archives designed by the Landesarchiv Baden-Württemberg (State Archives of Baden-Württemberg). In the Federal Republic of Germany, there are more than 2,300 archives. They can be divided into the following categories: national, municipal, city, academic and research institutions, mass media, church, political organizations, parliament, business, private, and family/house. This article focuses mainly on state archives, which form part of the national archives network. Based on the administrative organization of Germany, there are sixteen such institutions, each located in one state (*land*).[1]

La Revue canadienne des sciences de l'information et de bibliothéconomie 39, no. 2 2015

The Landesarchiv Baden-Württemberg is located in the southwest of Germany and is one of the most information technology (IT)-oriented archives in the country. It can be treated as one of the leaders in digital preservation, together with other archives, such as the Federal Archives (Bundesarchiv). It is responsible for records produced by state public agencies and the local community. The archives possess the oldest electronic holdings in the German Federal Republic (Naumann 2007, 53), which include data (statistics, pictures, and procedures) of the local census dated from 1961 to 1970.[2]

The problem of electronic records already began to emerge in the 1960s in both the Federal Republic of Germany and the German Democratic Republic. At that time, archivists in the German Democratic Republic were contemplating which storage medium would be best for data preservation,[3] while those in the Federal Republic of Germany started to consider how to preserve and store data created in a digital environment. Ideas ranged from printing everything out or reproducing it on microfilm,[4] but it was finally decided to preserve the material in the native form—that is, digitally. The Federal Archives put forward a proposal to acquire all digital records produced, but a strong federalist spirit made this idea impossible to realize. However, since the time that electronic records began to be transferred to archives, archivists started to develop electronic tools to manage them and protect them from obsolescence (Ullmann 1998, 599–600). Later, they also realized that archivists needed to be in contact with each other and with the administration when the records were being created and even during the design and implementation of systems to come to an agreement on technical issues, such as file format, depth of metadata description, and so on (Kluttig 2006, 56–58).

Some of the first repositories for electronic records[5] were created in Germany by the Landesarchiv Baden-Württemberg (State Archives of Baden-Württemberg) and the Bundesarchiv at the beginning of the twenty-first century. The design was based on the Open Archival Information System (OAIS) model, and it covered the following areas: ingest, storage, management, administration, access, and preservation of trusted data (Keitel and Lang 2010, 55; Keitel 2013b, 148–49). However, the beginning of digital repositories, in this author's opinion, can be traced back to the first databases in which the descriptive metadata of records were collected; these later developed into much more advanced data systems.

Generally speaking, there are three concepts of digital preservation known to German archivists: by the creator of the records, by a third party (service provider), or by public archives (National Library of Australia 2003, 58–59). The first idea was rejected at the very start of discussions and the last one has gone on to become the most common in Germany.

Digitales Magazin (DIMAG) was started in 2006 as a project called Concept of a Digital State Archives (Konzeption für ein digitales Landesarchiv),[6] shortly after the first electronic records were transferred to the archives, and some reflections on digital archiving were made in 2002 by Christian Keitel.[7] At that point, there were no preservation systems for archives, and archivists only heard of electronic repositories in libraries. Thus, these were used as a basis

for further discussion on the matter, despite differences existing on the preservation of digital material between the two types of institutions. In the following years, archivists were trying to define how to preserve a new kind of material, while maintaining its connection to the analogue-based primary holdings, and they considered which metadata standards should be used (for example, OAIS, METS, Nestor manuals, ISAD (G), EAD, PREMIS, or the National Library of New Zealand Metadata Model). The initial step was research, which led to the identification of all kinds of records created by the local public administration. After this very important stage, the production of digital repositories went on track. At the beginning, DIMAG was equipped with two modules: ingest and registration of records. The access point module was added later, and further file formats were recognized and supported. The final version featured the following sections: appraisal, description, access, storage, and preservation. The hierarchical structure of the storeroom allowed searches based on its tectonics: fonds, units, series, objects, and their properties defined in technical metadata.

This software has been in use since 2006 and presently can operate with materials of different file formats (limited) and functionalities such as records, intranet sites, websites, audio-visual, and data extracted from databases (for example, land register files).[8] Materials stored in the repository can be digitally born, therefore created natively in an electronic environment or digitized from analogue material. Apart from establishing the repository, several important documents were prepared, such as *Records Digital Preservation Metadata* (*Metadaten für die Archivierung digitaler Unterlagen*)[9] and other tools for automated batch identification of file formats and properties, called IngestList.[10] As a result of cooperation with other archives (Hessische Staatsarchive, die Staatliche Archive in Bayern), new modules, such as the automatic Submission Information Package for ingest were planned (Keitel 2013a, 54; Keitel 2013b, 147–51; Kemper and Naumann 2013, 86).

DIMAG was meant to be easily accessible via a standard Internet browser and configurable for different needs and users. It meets the purpose of acquiring archival material from several archives in one storage space, makes unauthorized access impossible, and can be offered as a final product for self-installation and use or can also be used as a service to other archival institutions, in particular smaller ones (Keitel 2013a, 56–60).[11]

Both concepts of sharing software or providing it as a service are not new but are still not very common. These concepts were developed in Germany in relation to the requirements of archival science, its methodology and the national law, which determines how the software needs to function. There are some examples, even worldwide, of archivists sharing software, such as open-source International Council on Archives (ICA) AtoM[12] web-based description software developed by Artifactual Systems in collaboration with the ICA and other international parties. The success of sharing this software is based on the application of ICA descriptive standards and interface multilingualism.

The idea of offering software as a service (SaaS) is an opportunity for institutions that are required to preserve digital records but have small budgets for

IT solutions and have no or poor know-how.[13] This service is based on principles of cloud computing.[14] It generally means that clients receive their virtual space that can be accessed remotely. This solution is of course connected with some regular costs, but it is cheaper than creating one's own infrastructure and providing for its own long-term maintenance (Keitel n.d., 1–2).

The many inquiries about digital preservation received by the Landesarchiv Baden-Württemberg demonstrate that not every archive can afford, or would want to create, its own digital repository in the future, mainly due to financial reasons. They also proved that German archivists need to agree on approaches on collaborative preservation and license the software widely in the German archival world. Not only would this give an opportunity to test DIMAG, but it would also help to find co-developers. First attempts to promote the software and find potential collaborators began in 2009. One year later, the software was used by the Hessisches Hauptstaatsarchiv Wiesbaden (Hessian State Archives Wiesbaden) and, in 2012, by the Staatliche Archive Bayern (Bavaria State Archives), including five more state archives from Bremen, Hamburg, Mecklenburg-Vorpommern, Niedersachsen, and Schleswig-Holstein in 2014.[15] At present, there are four kinds of partnerships: repository (*Magazinpartnerschaft*), development (*Entwicklungspartnerschaft*), support (*Supportpartnerschaft*), and provider (*Diensleistungspartnerschaft*) partnership. The first one is dedicated to archives that have no resources to maintain the software or IT infrastructure on their own but are able to pay a fee for space in the digital storerooms of other archives. The second one is for those who are ready to co-develop the software, for example, by programming new modules. The third kind of partnership offers software for installation on an archives' own IT infrastructure with a limited period of technical support (including installation and updates). The last one suits those archives that want to use digital preservation as a service delivered by external data processing centres. It is very similar to the first option.

The partnerships are planned as business models. Software is never given for free and always involves legal issues. Participating archives have to contribute to the costs of building DIMAG and at the very minimum accept the main technical solutions. They are also not allowed to decompile the source code, license it, pass it on to third parties, or do anything that could be perceived as competitive behaviour (Keitel 2013a, 55; Keitel 2013b, 152–55). As developers, the Hessian and Bavarian state archives create tools for ingest and access.

To put the idea of DIMAG into action as a service, cooperation with central and local data centres belonging to the Data Processing Network Baden-Württemberg (Datenverarbeitungsverbund Baden-Württemberg) was initiated, and it brought about some promising results. Data centres provided their IT infrastructure and technical support, and the State Archives of Baden-Württemberg delivered specialized support on archival issues related to the software. This idea is still in progress and so far has been presented, among others, to municipal archives at three conferences in 2012. Archivists expressed their interest, and the archives intended to finalize the cooperation with three local data centres. The first test project was started with the Municipal Data Processing Region

Stuttgart (Kommunale Datenverarbeitung Region Stuttgart) and university archives (Keitel 2013a, 56–57).[16]

To summarize, the archivists at the Landesarchiv Baden-Württemberg are of the opinion that all kinds of cooperation result in success, but, in practice, however, it can be difficult to find someone to work with, and to convince, a partner to cooperate (Keitel 2013a, 54). These archivists were, and still are, very lucky to have such partners on board as the Data Processing Network Baden-Württemberg and especially the Municipal Data Processing Region Stuttgart or the Karlsruhe Institute for Technology (Karlsruher Institut für Technologie) and seven out of sixteen state archives that want to work together on developing the software. This has developed into a phenomenon on a global scale. One might reflect on the fact that outsourcing solutions for archives had been considered as early as the 1990s, but they had not been implemented because of legal constraints related to the transfer, security, and access of data by third parties (Ullmann 1998, 601). Furthermore, as can be seen from the German example, public archives in the cloud are close to becoming a reality in many states, cities, and municipalities, with universities and churches also being interested in using DIMAG.

Notes

1. Federal Republic of Germany, as a federal parliamentary republic, consists of sixteen regions with limited autonomy. They are called *land*—that is, state.
2. More about these data (Keitel 2004).
3. This kind of the approach shows that the German Democratic Republic archivists treated archival material in a traditional way. More attention was paid to the data carrier than to information itself—the opposite of what it is like in modern digital preservation attempts.
4. However, this idea of preserving digital archival material on microfilm was successfully developed by, among others, the State Archives of Baden-Württemberg. More about this project can be found in ''ARCHE: Projekt mit Förderung durch—'Förderung von innovativen Netzwerken' des Bundesministeriums für Wirtschaft und Arbeit,'' Landesarchiv Baden-Württemberg, http://www.landesarchiv-bw.de/web/46253. ''Projekt Ausbelichtung von Farbdigitalisaten mit dem ARCHE-Laserbelichter. Erprobung des Echtbetriebs,'' Landesarchiv Baden-Württemberg, http://www.landesarchiv-bw.de/web/49137.
5. ''Electronic records'' is a term used to refer to all records in digital form, transferred to, and accessioned in, the archives. It does not matter if they are born digital or digitized (Keitel 2010).
6. ''Baden-Württemberg, Hessen und Bayern kooperieren bei der Archivierung digitaler Unterlagen. Bundesweit einmalige Drei-Länder-Kooperation zur Software-Fortentwicklung,'' http://www.landesarchiv-bw.de/highlight_hp3.php?hl_link=http://www.landesarchiv-bw.de/web/bundesweit_einmalige_drei-laender-kooperation_zur_software-fortentwicklung/53471&q=kooperation.
7. More about the Projekt 'Konzeption für ein digitales Landesarchiv'. Projekt mit Förderung durch das Land Baden-Württemberg,'' see Landesarchiv Baden-Württemberg. http://www.landesarchiv-bw.de/web/44346.
8. It is an archives for land register records, so there were some changes in the primary DIMAG done. For more about this issue, see Lang (2013).

9. Germany, Landesarchiv Baden-Württemberg (Keitel, Naumann, and Lang 2008).
10. See IngestList Beta, http://sourceforge.net/projects/ingestlist/.
11. More about the technical site of the project can be found in Keitel and Lang (2010).
12. See ICA AtOM, https://www.ica-atom.org/.
13. Definition of software as a service (SaaS) by the National Institute of Standards and Technology (NIST 2011, 2).
14. Definition of cloud computing by the National Institute of Standards and Technology (NIST 2011, 2–3).
15. Elektronische Archivierung im Digitalen Archiv Nord (DAN), Landesamt für Kultur und Denkmalpflege Mecklenburg-Vorpommern, http://www.kulturwerte-mv.de/cms2/LAKD1_prod/LAKD1/de/Landesarchiv/Elektronisches_Landesarchiv/Elektronische_Archivierung_im_Digitalen_Archiv_Nord/index.jsp.
16. Digitale Langzeitarchivierung, Kommunale Informationsverarbeitung Reutlingen-Ulm, http://www.rz-kiru.de/,Lde/Startseite/Service/DIMAG.html.

References

Keitel, Christian. N.d. *Zugänglichkeit contra Sicherheit? Digitale Archivalien zwischen Offline-Speicherung und Online Benutzung*. http://www.staatsarchiv.sg.ch/home/auds/06/_jcr_content/Par/downloadlist_3/DownloadListPar/download_2.ocFile/Text%20Keitel.pdf.

——. 2004. ''Erste Erfahrungen mit der Langzeitarchivierung von Datenbanken. Ein Werkstattbericht.'' In *Digitales Verwalten: Digitales Archivieren, Veröffentlichungen aus dem Staatsarchiv der Freien und Hansestadt*, Band 19, ed. Rainer Hering and Udo Schäfer, 71–81. Hamburg: Hamburg University Press. http://www.staatsarchiv.sg.ch/home/auds/08/_jcr_content/Par/downloadlist_2/DownloadListPar/download_9.ocFile/Text%20Keitel.pdf.

——. 2010. ''Digitale Archivierung beim Landesarchiv Baden-Württemberg.'' *Archivar* 1: 23.

——. 2013a. ''Dienstleisterpartnerschaften mit DIMAG.'' In *Das neue Handwerk. Digitales Arbeiten in kleinen und mittleren Archiven*. Vorträge des 72. Südwestdeutschen Archivtags am 22. und 23. Juni 2012 in Bad Bergzabern, ed. P. Müller and K. Naumann, 54–57. Stuttgart, Germany: W. Kohlhammer.

——. 2013b. ''DIMAG-Kooperationen.'' In *Digitale Archivierung in der Praxis*. 16. Tagung des Arbeitskreises ''Archivierung von Unterlagen aus digitalen Systemen'' und nestor-Workshop ''Koordinierungsstellen,'' Werkhefte der Staatlichen Archivverwaltung Baden-Württemberg Serie A, vol. 24. ed. Ch. Keitel and K. Naumann, 147–55. Stuttgart, Germany: W. Kohlhammer.

Keitel, Christian, and Rolf Lang. 2010. ''DIMAG und IngestList. Übernahme, Archivierung und Nutzung von digitalen Unterlagen im Landesarchiv Baden-Württemberg.'' In *Archivische Informationssysteme in der digitalen Welt, Werkhefte der Staatlichen Archivverwaltung Baden-Württemberg Serie A*, vol. 23. ed. G. Maier and T. Fritz, 55–60. Stuttgart, Germany: W. Kohlhammer.

Keitel, Christian, Kai Naumann, and Rolf Lang, eds. 2008. *Metadaten für die Archivierung digitaler Unterlagen*. http://www.landesarchiv-bw.de/sixcms/media.php/120/48392/konzeption_metadaten10.28354.pdf.

Kemper, Joachim, and Kai Naumann. 2013. ''Selbermachen! Praktische Tipps zur Archivierung digitaler Unterlagen, Digitalisierung und Öffentlichkeitsarbeit im Netz.'' In *Das neue Handwerk. Digitales Arbeiten in kleinen und mittleren Archiven,* ed. P. Müller and K. Naumann. Vorträge des 72. Südwestdeutschen Archivtags am 22. und 23. Juni 2012, Bad Bergzabern, 86. Stuttgart, Germany: W. Kohlhammer.

Kluttig, Thekla. 2006. ''Strategies of German State Archives for the Preservation of Electronic Records.'' In Archives in the New Age: The Strategic Problems of the Automatization of Archives Information. *Papers of the International Conference Warsaw*, September 28–29, 2001, Colloquia Jerzy Skowronek dedicata, ed. A. Biernat and W. Stępniak, 56–59. Warsaw: Naczelna Dyrekcja Archiwów Państwowych – Wydział Wydawnictw.

Lang, Rolf. 2013. ''Die elektronische Grundakte in G-DIMAG.'' In Digitale Archivierung in der Praxis. 16. Tagung des Arbeitskreises ,,Archivierung von Unterlagen aus digitalen Systemen'' und nestor-Workshop ,,Koordinierungsstellen, Werkhefte der Staatlichen Archivverwaltung Baden-Württemberg Serie A Heft 24, ed. Christian Keitel and Kai Naumann, 129–41. Stuttgart, Germany: W. Kohlhammer.

National Institute of Standards and Technology. 2011. *The NIST Definition of Cloud Computing: Recommendations of the National Institute of Standards and Technology*. Gaithersburg, MD: US Department of Commerce, National Institute of Standards and Technology. http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.

National Library of Australia. 2003. *Ochrona dziedzictwa cyfrowego. Zalecenia*. Warsaw: Naczelna Dyrekcja Archiwów Państwowych.

Naumann, Kai. 2007. ''Älteste digitale Archivquelle der Bundesrepublik gesichert: Daten der Volkszählung von 1961 für das Land Baden-Württemberg übernommen und aufbereitet.'' *Der Archivar* 60 (1): 53.

Ullmann, Angela. 1998. ''Der EDV-Ausschuß der Archivreferentenkonferenz des Bundes und der Länder 1972–1994.'' *Der Archivar* 51 (4): 587–607.

Archivematica As a Service: COPPUL's Shared Digital Preservation Platform

Le service Archivematica : La plateforme partagée de conservation de documents numériques du COPPUL

Bronwen Sprout
Digital Programs and Services, University of British Columbia Library
bronwen.sprout@ubc.ca

Mark Jordan
W.A.C. Bennett Library, Simon Fraser University
mjordan@sfu.ca

Abstract: The Council of Prairie and Pacific University Libraries (COPPUL) is piloting a cloud-based preservation service using the Archivematica digital preservation system. The service is offered to COPPUL member institutions that wish to preserve digital holdings but are unable or unwilling to install and manage local Archivematica instances. This service is a joint effort of COPPUL, Artefactual Systems (Archivematica lead developers), and University of British Columbia (UBC) Library (the cloud storage provider). COPPUL is responsible for promoting the service, signing up new institutions and seeding the one-time set-up costs. Artefactual Systems provides account administration, installation, server administration, and user technical support. UBC Library provides fee-based server hosting and digital object storage service. This article discusses COPPUL's Archivematica-as-a-service model generally and covers certain aspects of implementation in greater detail, concluding with a discussion of future directions.

Keywords: digital preservation, Archivematica, COPPUL, library consortia, cloud computing

Résumé : Le Conseil des bibliothèques universitaires des Prairies et du Pacifique (Council of Prairie and Pacific University Libraries [COPPUL]) met à l'essai un service de conservation basé sur un nuage informatique qui utilise le système de conservation numérique Archivematica. Le service est offert aux institutions membres du COPPUL qui souhaitent effectuer la conservation de leur fonds documentaire numérique, mais qui n'ont pas la possibilité ni l'intention d'installer et de gérer une implémentation Archivematica locale. Ce service est un effort conjoint de COPPUL, de Artefactual Systems Inc. (les développeurs principaux d'Archivematica) et de la bibliothèque de l'Université de Colombie britannique (*UBC Library*), qui est le fournisseur du stockage en nuage. COPPUL a la responsabilité de la promotion du service, de l'enrôlement de nouveaux établissements et de la dissémination des coûts uniques de mise en place; Artefactual Systems fournit

l'administration des comptes, l'installation, l'administration des serveurs et le soutien technique aux utilisateurs; et la Bibliothèque de l'Université de Colombie britannique offre le service d'hébergement payant sur ses serveurs et un service de stockage d'objets numériques. Cet article examine le modèle de service Archivematica en général et en tant que service de COPPUL, et couvre certains aspects de la mise en œuvre de manière plus détaillée, pour conclure par une discussion sur les orientations futures.

Mots-clés : conservation numérique, Archivematica, COPPUL, consortiums universitaires, informatique en nuage

Introduction

This article discusses a program developed by the Council of Prairie and Pacific University Libraries (COPPUL) to offer cloud-based digital preservation services to its members. The service is offered to COPPUL member institutions that wish to preserve digital holdings but are unable or unwilling to install and manage local Archivematica instances. Archivematica is a free and open-source digital preservation system that is designed to maintain standards-based, long-term access to collections of digital objects. This COPPUL service (which is known as Archivematica as a service) demonstrates many of the benefits of a community-based digital preservation model as well as some of the preconditions necessary for its success, including institutional cloud-based computing, experience with an open source vendor, a history of working cooperatively, and trust within the membership.

After introducing COPPUL and highlighting previous shared services of its Digital Preservation Working Group (DPWG), the article discusses Archivematica as a service, including COPPUL's rationale for the offering, details of the service, support and communication mechanisms, and a description of Archivematica and how it was chosen to fill digital preservation needs. The article considers the benefits for both the provider and the service users and concludes with a discussion of future directions and challenges for the service, including issues of sustainability, scale, and shared governance.

About COPPUL and the DPWG

COPPUL "provides leadership in the development of collaborative solutions addressing the academic information resource needs, the staffing development needs, and the preservation needs of its member institutions" (COPPUL n.d.a). The consortium comprises twenty-three university libraries located in Manitoba, Saskatchewan, Alberta, and British Columbia, and fifteen affiliate members that participate in licenses for electronic resources, discounted pricing, and favourable terms on licensed resources. Beyond consortial licenses to resources, benefits for COPPUL members include networking and information sharing for directors and staff, shared expertise to advance collaborative projects, workshops and continuing education for staff at member libraries, and the opportunity to participate in working groups, including the DPWG. In its 2010–15 strategic

directions framework, in addition to identifying a role for COPPUL as an research and development incubator, digital and electronic collections are identified as one of the three main areas of focus, and digitization and digital preservation are further identified as areas on which COPPUL members will work on collectively (COPPUL n.d.b).

The DPWG is one of several working groups that are active within COPPUL. Other working groups focus on scholarly communications, research data, collections, and return on investment. COPPUL also offers several programs, of which Archivematica as a service has perhaps most in common with the Shared Print Archive Network (SPAN), a distributed retrospective print repository program (COPPUL n.d.c). Twenty-one COPPUL libraries participate in this program to preserve an optimal number of printed journals and provide access to shared print archives. In many ways, including a level of comfort with cost sharing, a commitment to working together, and a shared leadership model, this history of collaborating to preserve print archives has paved the way for COPPUL to also enable its members to work together to preserve their digital materials.

As noted in its statement of purpose, the work of the DPWG is "informed by significant developments in digital preservation currently underway in the memory institution community." Various COPPUL and DPWG members are engaged in related digital preservation efforts including participation in several Private LOCKSS Networks (PLNs), participation in the Global LOCKSS Network and/or Portico, local implementations and use of Archivematica, use of Archive-It to archive websites, and digital preservation policy development. Within this context, the DPWG was tasked with, among other things, developing options for a common approach to digital preservation for COPPUL libraries, with a particular focus on solutions that require consortial-level, or inter-institutional, cooperation for their effectiveness.

Prior to the DPWG (which was formed mid-2012), many of the members had experience working together to subscribe to, or develop, other shared digital preservation services. The most relevant example of this is the COPPUL PLN, which was established in December 2007, when the directors of COPPUL libraries agreed to support a two-year pilot project. The mission of the COPPUL PLN is to "preserve digital collections of local interest to COPPUL members that are not being preserved elsewhere, other than local backup" (COPPUL Digital Preservation Working Group n.d.). All locally created collections that are at risk of being lost are candidates for the network, and material that has been contributed includes open access journals, most created using the Open Journal Systems platform, as well as born digital government publications, theses and dissertations, and locally digitized materials. Any COPPUL full member institutions that are able to meet the COPPUL PLN membership requirements are eligible to participate as members in the COPPUL PLN. Although there must be a minimum of seven members for the COPPUL PLN to function, there is no maximum number of members, and since its inception there have been at least nine participating nodes in the COPPUL PLN. The

operations of the PLN were originally overseen by a Steering Committee and supported by a technical committee, but since the creation of the DPWG in 2012, this oversight has become a function of the DPWG.

Archivematica as a service

The examples of the SPAN and the PLN illustrate some of the experience COPPUL members share in working together to develop preservation-related programs of mutual benefit. However, although the need for digital preservation and the benefits of shared services are well understood by DPWG members, the services to support these services are not consistently in place at the local level, particularly in some of the smaller institutions. The COPPUL PLN provides redundant storage for certain types of collections, but other features and functions of preservation were missing, including format migration, preservation metadata, and more. To address the gap between preservation storage and full preservation services, some members were using the Archivematica digital preservation software, and the DPWG had formed a working group to discuss testing results and the experiences and workflows of a few local production implementations. Several other COPPUL institutions were interested in implementing Archivematica but did not have the local infrastructure in place to support it.

Although members shared a broad level of knowledge and interest in the Archivematica software generally,[1] interest in Archivematica as a service was first generated at a digital preservation workshop organized by the DPWG and held in Vancouver in March 2013. Library directors were asked to participate in the workshop along with one staff member from each institution. Many of the presentations and much of the discussion concerned members' use of Archivematica digital preservation software. Directors saw the potential for a shared service, and the DPWG was asked to follow up with a proposal. This proposal was presented to the directors at their September 2013 meeting, and after a funding model was agreed upon, several members of the DPWG began working with Artefactual Systems, the lead developers of Archivematica, to operationalize the service.

In developing the service to provide hosted instances of Archivematica, it was envisioned that hosting and digital object storage would be provided by one or more COPPUL institutions. Artefactual had identified potential commercial cloud hosts, but few Canadian options existed (a necessity for potential users in terms of privacy regulations, especially around the storage of potentially sensitive archival data). Coincidentally, the University of British Columbia's (UBC) information technology (IT) group had just launched a new cloud hosting service, branded as EduCloud, which seemed promising. The EduCloud service is a cloud-computing service based on the UBC Vancouver campus that allows self-management from a web portal and self-deployment from templates. Importantly, for BC clients, this service meets BC provincial privacy requirements under the Freedom of Information and Protection of Privacy Act.[2] In addition, it offers the benefits of a virtual server service such as server consolidation, resource pooling, high service availability, and regular backups. Multiple

(V9 15/7/15 14:42) UTP (6"×9") AGaramond 1323 CJILS 39:2 pp. 235–244 1659 CJILS 39.2_09_Sprout (p. 238)

consumption models are available, ranging from capacity-as-you-go to reserved pools (UBCIT n.d.). EduCloud appeared to meet the needs of the Archivematica service offering, and UBC Library offered to act as a liaison between UBCIT and COPPUL/Artefactual.

Artefactual chose EduCloud partly because high-volume discounts on EduCloud's virtual machine (VM) ware licenses meant that its prices were highly competitive with commercial providers. A more important factor was that all of the parties felt that partnering with UBC added a level of accountability to the service that would be missing if Artefactual, a private company, had developed its own private branded cloud service using a third-party commercial cloud provider such as Amazon Web Services. This is particularly true in the case of EduCloud, whose goal is to provide low-cost computing and storage infrastructure to scholarly institutions. This goal meshes well with COPPUL's support for university libraries and Artefactual's mission to provide open-source software to the heritage community.

With UBC identified as the cloud service provider, the roles of the three parties involved were defined: the service would be a joint effort of COPPUL, Artefactual Systems (the Archivematica lead developers and support providers), and UBC, the cloud storage provider. Responsibilities for the service have been divided along functional lines: COPPUL is responsible for promoting the service, signing up new institutions, and subsidizing Archivematica technical support; Artefactual Systems provides account administration, installation, server administration and user technical support, and end-user training with each significant upgrade; and UBC provides fee-based server hosting and digital object storage service. Artefactual deploys and manages the VMs on EduCloud with UBC Library acting as a liaison. Individual VMs for member institutions are installed on EduCloud and could be moved in-house if they decide to withdraw from the COPPUL service.

The proposal was structured to include options for members who wanted full preservation and access services and those who wanted to gain experience with Archivematica while making less of a financial commitment. The service and the fee structure were designed based on a tiered model, with a range of storage, functionality, and support available. Three different service levels were offered to member institutions (see Table 1).

Participating institutions derive substantial benefits from the service, including the ability to use an existing digital preservation platform; training and technical support services from experienced Archivematica developers and digital preservation specialists; centralized system administration at a much lower cost than paying for a local system administrator; and annual maintenance and software upgrades subsidized by COPPUL. Participating institutions also benefit by being part of an existing community of Archivematica users. The DPWG as a whole benefits from the added dimension that the service offers its preservation discussions, and, indeed, the model of the Archivematica service suggests a possible future for a shared preservation network. Since Archivematica is open source, all users also benefit from a larger community of clients and non-clients

Table 1: Archivematica as a service levels.

Bronze Service Level	Silver Service Level	Gold Service Level
Basic ingest and storage management:	Full ingest micro-services:	Full ingest micro-services, plus DIP upload to AtoM and full AtoM support:
• assign universal unique identifier to each object	• all services provided in tier 1	• all services provided in tiers 1 and 2
• transfer digital holdings into Archivematica	• prepare Submission Information Packages	• generate DIP (access copies)
• assign UUID (universal unique identifier to each object	• assign descriptive metadata	• upload DIP to AtoM
• calculate checksums	• normalize (generate preservation copies)	• display digital objects in AtoM
• extract packaged files	• generate PREMIS metadata	• enhance metadata and manage accessions in AtoM
• generate METS file	• generate AIP METS file	• eight CPUs, 48 GB RAM, 2 TB disk space
• scan for viruses	• index METS file	• fifteen support tickets
• clean up filenames (remove prohibited characters)	• package contents in Library of Congress BagIt format	• online Archivematica and AtoM training
• extract technical metadata	• compress AIP	
• identify format	• place AIP in secure storage	
• validate format	• periodically verify checksums of stored AIPs	
• index transfer	• four CPUs, 32 GB RAM, 1 TB disk space	
• assign rights metadata	• ten support tickets	
• place transfer in secure storage	• online Archivematica training	
• periodically verify checksums of stored transfers		
• two CPUs, 16 GB RAM, 400 GB disk space		
• five support tickets		
• online Archivematica training		

alike, who share knowledge including technical support in a public forum—the Archivematica discussion list.

For Artefactual, the benefits of a centrally hosted model mean that the lower-level issues are managed without the individual clients needing to know about them. Artefactual has direct contact with UBC IT and has been able to resolve some issues before the end users are aware of them. Further, a level of standardization has been achieved that is more difficult to attain when each institution hosts its own server. From a technical perspective, EduCloud provisions a pool of compute resources (central processing unit, memory, and storage) and

Artefactual provisions VMs for each institution, with resources allocated as defined by the subscriber's service level. Artefactual then deploys Archivematica (and the AtoM archival description software) on these VMs and configures the applications for use by the subscribers.

In terms of application support, when hosted service users have questions about how to use the software, or experience technical problems when using the software, the support process is the same as for other Artefactual clients. Whereas the cost of a support contract from Artefactual may be too high for smaller institutions to take on by themselves, the COPPUL hosting service levels the playing field for these users.

Technological infrastructure

As stated earlier, the service is hosted on the UBC's EduCloud server platform. OVH, another cloud infrastructure provider, was selected as a backup host. An important goal in developing the service was to allow for the hosting of Archivematica on a variety of cloud platform providers.

After the initial research and selection of EduCloud, work started on building the infrastructure required to deploy and manage what amounts to a private cloud. After successful initial test deployments in EduCloud (with OVH as a backup and test environment), Artefactual developed a suite of deployment tools based on an open-source automated configuration management system called Ansible. Prior to the development of these tools, installing Archivematica required a high level of technical expertise and three to four hours of time. The Ansible tools brought the deployment time down to twenty to thirty minutes. Even more importantly, the tools allow all of the configuration information to be documented and easily reproducible. This has numerous benefits, including improved backup and disaster recovery processes and the ability to reproduce perfectly a production site in a test environment to replicate bugs reported by users. It is worth noting that once the COPPUL infrastructure was completed, Artefactual received funding from other institutions to improve and extend the original Ansible tools. Since Archivematica is an open-source project, these tools are being released under the AGPL3 open-source software license for others to use and enhance. In this way, COPPUL supported not only its own member institutions but also the digital preservation community at large.

The resulting infrastructure highlights the advantages of a hosted service over siloed local installations in diverse hardware and software environments. Artefactual systems has easy, standardized access to all of the client installations, which makes error diagnosis and software upgrades much simpler than they are when working with the same number of installations in diverse locations with varying hardware, network, storage, and security infrastructures. Moreover, when multiple institutions pool resources in a virtualized cloud-hosting environment, it is a simple matter to allocate resources based on processing needs. This means that institutions with large numbers of video files, for example, can purchase additional processing power at a relatively low price. A final and very important advantage of the pooled hosted service is that all of the clients are able to rely on

UBC's IT department to deal with storage, security, backup, and other issues. This can make a tremendous difference to small institutions with limited IT resources.

Experience so far

To date (about half-way through the first year of the service at the time of writing), progress in implementing Archivematica at the subscribing institutions has varied for several reasons. Artefactual has been working with all of the subscribers on training, preservation planning, and operationalization of the service, and, overall, progress has been substantial. During this initial implementation period, several interesting issues have surfaced. First, one institution has been required to satisfy their campus legal staff that preserving organizational records using Archivematica is consistent with the university's privacy policies. Security considerations are an important part of the Open Archival Information System (OAIS) Reference Model and are outlined in Annex F of the "Magenta Book" (Consultative Committee for Space Data Systems 2012). These considerations are "informative" and not "normative," which means that they simply point out security issues and define what a compliant system must do to address them and do not stipulate a specific technical security model. Archivematica, as a digital preservation system that aims to be compliant with the OAIS functional model, implements specific access controls on content under its purview, but these controls must also be consistent with local policies dealing with security, privacy, and records retention.

Another issue is that implementing Archivematica requires considerable resources that have little to do with a library's ability to provide technological infrastructure. Many of the subscribing institutions do not have comprehensive digital preservation policies or frameworks, and the lack of a digital preservation framework that defines preservation priorities and policies has forced subscribing institutions to spend staff resources addressing these questions early in their implementations of the service. The absence of a comprehensive digital preservation framework before implementing a system such as Archivematica is not necessarily negative. For many sites, implementing a system offers them a concrete opportunity to focus on their priorities and to develop policies around the operational strategies that Archivematica offers.

A third issue, related to the previous one, is that integrating Archivematica with content repositories such as DSpace requires working with their campus' central IT department. In cases where these repository platforms are hosted on behalf of the library by central IT departments, requirements arising from integrating the platforms with external applications such as Archivematica may not have been anticipated when the repositories were implemented or may not be possible given the security policies applied to the local infrastructure. For example, Archivematica can accept exports from DSpace (Artefactual Systems), but taking advantage of this feature requires access to the exported content in ways that many central IT departments may find problematic to configure, especially if they were not anticipated when DSpace was originally provisioned. Archivematica's requirements are not unreasonable or insecure, but they may

pose barriers to implementation in some situations. Nonetheless, at least one subscribing institution has started ingesting DSpace exports in their hosted instance—their access to Archivematica as a hosted service provided them the opportunity to work with their central IT staff to investigate and implement the integration.

From Artefactual's perspective, the experience so far has been positive in several ways. First, COPPUL's service has enabled them to work with a group of clients who may not otherwise have implemented Archivematica on their own because they lacked the local technical infrastructure to do so. Second, in preparation for implementing the COPPUL service, Artefactual developed a suite of deployment tools based on an open-source automated configuration management system called Ansible. Prior to the development of these tools, installing Archivematica required a high level of technical experience and three to four hours of time. The Ansible tools brought the deployment time down to twenty to thirty minutes. Even more importantly, the tools allow all of the configuration information to be documented and easily reproducible. This has numerous benefits, including improved backup and disaster recovery processes and the ability to reproduce perfectly a production site in a test environment to replicate bugs reported by users. This work has allowed Artefactual to develop hosting services with new partners. For example, in August 2014, Artefactual Systems and DuraSpace announced a collaborative service to host Archivematica on the DuraCloud platform (DuraSpace n.d.), which has been branded "ArchivesDirect" (ArchivesDirect).

Future directions for the service

In the immediate term, encouraging more COPPUL members to subscribe to Archivematica as a service is a priority. The funding and sustainability models used by COPPUL's Archivematica service, combined with a flexible model for provisioning the necessary server and storage infrastructure to meet demand, will allow the number of subscribers to the service to expand incrementally within the next few years. The DPWG is also exploring the development of additional shared digital preservation services for COPPUL members, modelled after the Archivematica service. These new services may also incorporate aspects of SPAN, where applicable. The most obvious such service would be to transform the current COPPUL PLN so that COPPUL members that do not host nodes in the network can have access to shared storage capacity. Another may be development of a shared service to use the Internet archives' Archive-It service to ensure that institutions that do not subscribe have input into collaborative web archiving initiatives. The cost-sharing and service models developed for Archivematica as a service can serve as a template for these and other shared digital preservation services within COPPUL.

Acknowledgements

The authors would like to thank Evelyn McLellan, Justin Simpson, and Courtney Mumma of Artefactual Systems for their invaluable assistance in the preparation of this article.

Notes

1. Archivematica is a free and open-source digital preservation system that is designed to maintain standards-based, long-term access to collections of digital objects. Archivematica uses a micro-services design pattern to provide an integrated suite of software tools that allows users to process digital objects from ingest to access in compliance with the International Organization for Standardization OAIS functional model. Users monitor and control the micro-services via a web-based dashboard (Artefactual Systems).
2. Freedom of Information and Protection of Privacy Act, RSBC 1996, c 165.

References

Consultative Committee for Space Data Systems (CCSDS). 2012. *Reference Model for an Open Archival Information System (OAIS): Recommended Practice CCSDS 650.0-M-2: Magenta Book*. Washington, DC: CCSDS Secretariat. http://public.ccsds.org/publications/archive/650x0m2.pdf.

COPPUL. N.d.a. *About Us*. http://www.coppul.ca/about-us.

COPPUL. N.d.b. *2012–2015 Strategic Directions Framework*. http://www.coppul.ca/sites/default/files/uploads/StratFramework.pdf.

COPPUL. N.d.c. *Shared Print Archive Network (SPAN)*. http://coppul.ca/programs/shared-print.

COPPUL Digital Preservation Working Group. N.d. *PLN Subgroup*. http://coppuldpwg.wordpress.com/committees/pln-subgroup/.

DuraSpace. N.d. *DuraSpace and Artefactual Partner to Offer New Hosted Service*. http://duraspace.org/articles/2211.

UBCIT. N.d. *EduCloud Server Service*. http://it.ubc.ca/services/web-servers-storage/educloud-server-service.