algorithms

MDPI

Article

Trust in the Balance: Data Protection Laws as Tools for Privacy and Security in the Cloud

Darra Hofman, Luciana Duranti * and Elissa How

School of Library, Archival and Information Studies, University of British Columbia, Vancouver, BC V6T 1Z1, Canada; darra.hofman@gmail.com (D.H.); elissahow@gmail.com (E.H.)
* Correspondence: luciana.duranti@ubc.ca; Tel.: +1-604-822-2587

Academic Editors: Sye Loong Keoh and Khin Mi Mi Aung
Received: 24 January 2017; Accepted: 18 April 2017; Published: 27 April 2017

Abstract: A popular bumper sticker states: "There is no cloud. It's just someone else's computer." Despite the loss of control that comes with its use, critical records are increasingly being entrusted to the cloud, generating ever-growing concern about the privacy and security of those records. Ultimately, privacy and security constitute an attempt to balance competing needs: privacy balances the need to use information against the need to protect personal data, while security balances the need to provide access to records against the need to stop unauthorized access. The importance of these issues has led to a multitude of legal and regulatory efforts to find a balance and, ultimately, to ensure trust in both digital records and their storage in the cloud. Adding a particular challenge is the fact that distinct jurisdictions approach privacy differently and an in-depth understanding of what a jurisdiction's laws may be, or even under what jurisdiction particular data might be, requires a Herculean effort. And yet, in order to protect privacy and enhance security, this effort is required. This article examines two legal tools for ensuring the privacy and security of records in the cloud, data protection laws, and data localization laws, through the framework of "trust" as understood in archival science. This framework of trust provides new directions for algorithmic research, identifying those areas of digital record creation and preservation most in need of novel solutions.

Keywords: archival science; cloud services; trustworthy records; reliability; authenticity; accuracy; privacy; data protection; data localization

1. Introduction: Trust, Hope, or Desperation? Records in the Cloud

> Within the next three years, [...] more than four-fifths of all data center traffic, 83 percent, will be based in the cloud. What's more, most of this action will be going to public cloud services—there will be more workloads (56 percent) in the public cloud than in private clouds (44 percent).
>
> —J. McKendrick [1]

The data stored in the cloud—estimated to be in the range of 10 zettabytes (ZB) by 2019—include critical records that enable individuals, businesses, and even governments to continue functioning, such as identity and vital statistics records, bank and financial records, contracts, ownership and land records, and records related to the Internet of Things. Ensuring the continuing security, accessibility, and trustworthiness of these records is no small feat. Thus, cloud-based recordkeeping has become an entrenched part of many—if not most—people's and organizations' practices, often undertaken without a rigorous examination of the trustworthiness of the Cloud Services Providers (CSP) given charge over the records, and of their practices. Instead, "preserving information in the Cloud may be a black box process in which we know, at least ideally, what we put in for preservation, and we know what we want to access and retrieve—essentially the same things we put in—but often we do not know what technology is used by the CSPs to manage, store, or process our information" [2].

This lack of scrutiny of the trustworthiness of cloud services is deeply troubling; in 2016 alone, data breaches and hacks affected banks [3], the U.S. National Security Administration [4], and even, allegedly, the U.S. election [5]. As Duranti and Rogers state, "[e]ven as we have ever greater access to untold stores of information, our right to know comes at a rising cost to our privacy and anonymity, due to a complex web of data collection and surveillance, benign and not. These stores of information, furthermore, are accumulated and extracted from sources we often cannot know or evaluate" [6]. Even in the absence of malice, cloud-based recordkeeping poses a number of unanswered questions. The challenges identified by Duranti and Rogers [7]—including managing trans-jurisdictional data flows, attributing liability for and resolving data breaches, and establishing the chain of custody when a cloud service provider goes dark—remain unresolved. Given the potential risks, why individuals and organizations continue to store records in cloud environments? Is it because of a "blind trust" that the commercial providers will do right by us [7]? Is it hope that they will do so? Or is it a sense of desperation leading us to cross the metaphorical Mediterranean on a dinghy not because of trust in cloud providers but because of a perception that there is no better option in the face of an Internet-driven world?

There is, of course, constant technological innovation that tries hard to keep pace with both malicious and innocent challenges and ensure the trustworthiness of records in the cloud. What we know is that technical approaches alone cannot address them and solve the problems that arise from them; there is no technical solution to determined human misfeasance. Instead, technological tools must be supported by legal, social, and business structures that set the bar for minimum expectations for CSPs, ensuring that cloud recordkeeping will not endanger the preservation of our recorded memory and the continual functioning of society. While some organizations might indeed scrutinize the "reputation, performance, competence, and confidence" [7] of CSPs to verify their trustworthiness, research and experience show that many continue to be quite liberal with the choice of whom they entrust with their records. In such cases—where society is relying upon a service without assurances of its quality—the law often steps in to provide the certainty and trust that users cannot obtain on their own (see, e.g., [8]).

The abundance of records kept in cloud environments is forcing the law to both adapt old regulatory tools and create new ones. These tools are intended to balance ancient concerns—access, control, security, and trust—in a world where records have been loosened from the physical bonds that traditionally held them within a jurisdiction and in the care of a trusted custodian. By applying the archival framework of trust to some of the most frequently used legal tools for protecting records in the cloud—privacy, data protection, and data localization—this article highlights the ways in which those tools are meant to serve as surrogates for trust in CSPs, and examines ways in which they fail and they succeed. Ultimately, record security and privacy in the cloud will require innovation in the algorithmic sphere; simple necessities such as migration still pose a problem to proving the authenticity of records using current methods. Solutions to distributed recordkeeping problems—such as blockchain—will almost certainly require novel algorithms.

2. Materials and Methods: The Archival Paradigm for Trusting Records

The ultimate purpose of the archivist (the term "archivist" here is used to refer to all records professions, that is, to all occupations whose professional education is based on archival science) is to protect the trustworthiness of archival documents, or records. (In archival theory, the terms archival documents—referring to documents produced in the course of activity and kept for further use or reference—and records are synonyms.) As Eastwood states, "[the archival] discipline stands on two propositions [. . .]: that archival documents attest facts and acts, and that their trustworthiness is dependent upon the circumstances of their generation and preservation" [9]. These seemingly simple propositions, however, must be unpacked if one wishes to fully appreciate the archival paradigm of trust and its role in cloud-based recordkeeping. "Records" and "trustworthiness" are technical terms whose meaning brings sharp edges to the sometimes-fuzzy discussion of "trust" in both archival and legal contexts. Each proposition will be examined in turn.

While the general discourse often uses the term records as interchangeable with data, not all data rise to the level of "records." A record is properly understood as "a document created (i.e., made or received and set aside for further action or reference) by a physical or juridical person in the course of a practical activity as an instrument or by-product of such activity" ([7], citing [10]). This is no mere technical statement: the record is distinguished from all other data by its ability to serve as evidence of facts and acts:

> In order to conduct affairs, and in the course of conducting affairs, certain documents are created to capture the facts of the matter or action for future reference, to extend memory of deeds and actions of all kinds, to make it enduring. Inherent in this conception of the document's capacity to extend memory, to bear evidence of acts forward in time, is a supposition about the document's relation to fact and event or act. The matter at hand, the thing being done, produces the document, which then stands as a vehicle or device to access the fact and act. Documents of this type then came to be regarded as having what jurists called full faith or public faith—or, as we would say, as possessing trustworthiness as evidence of fact and act-if they were preserved in an appointed place according to fixed and well understood administrative procedures.
>
> —T. Eastwood [9]

Duranti and Rogers ([7], p. 524) examine in detail the necessary elements of a record: "(1) an identifiable context; (2) three identifiable persons concurring in its creation; (3) an action, in which the record participates or which the record supports either procedurally or as part of the decision-making process; (4) explicit linkages to other records within or outside the digital system, through a classification code or other unique identifier; (5) a fixed form; and (6) a stable content." Thus, there exist many records in electronic systems, although the conception of "fixed form" and "stable content" necessarily differs as compared to physical records: "A digital record has a fixed form if its binary content is stored so that the message it conveys can be rendered with the same documentary presentation it had on the screen when first saved, even if its digital presentation has been changed, for example, from .doc to .pdf. A digital record has a fixed form as well if the same content can be presented on the screen in several different ways but in a limited series of pre- determined possibilities; in such a case we would have different documentary presentations of the same stored record (e.g., statistical data viewed as a pie chart, a bar chart, or a table). Stable content means that the data or content of the record cannot be intentionally or accidentally altered, overwritten or deleted. The content is also considered stable when changes to what we visualize at any given time are limited and controlled by fixed rules, so that the same query or interaction always generates the same result, and we have different views of different subsets of content" ([7], p. 524).

In short, the archival paradigm centers upon preserving both records and their context in such a way as to enable future users—without privileging any particular future use—to evaluateparticular records' evidentiary capacity, and use records judged to be trustworthy as evidence of past facts and acts. Records, by their nature, testify in a way that mere data cannot; records that are disaggregated, de-identified, or otherwise stripped of their context in the cloud lose their capacity to serve as trustworthy evidence.

So what does it mean for a record to be trustworthy evidence? Eastwood's second proposition, "[records'] trustworthiness is dependent upon the circumstances of their generation and preservation" points to the elements underlying archival trustworthiness: reliability, accuracy, and authenticity. "A record is considered reliable when it can be treated as a fact in itself, that is, as the entity of which it is evidence. For example, a reliable certificate of citizenship can be treated as the fact that the person in question is a citizen" ([10], p. 6). Thus, a record's reliability is dependent upon the circumstances of creation; an unreliable record cannot be made reliable at some future point. Reliability is also dependent upon the competence of the author [7]; given the uncertain—and often undiscernible—creation and authorship of many records in the cloud, the hard truth is that individuals, businesses, and even governments find themselves depending on records that are per se unreliable. "[A]ccuracy is defined as the correctness and precision of a record's content, based on [its reliability] and on the controls on the recording of content and its transmission" ([7], p. 525). The third element of record trustworthiness is authenticity: "a record is authentic when it is the document that it claims to be. Proving a record's authenticity does not make it more reliable than it was when created. It only warrants that the record does not result from any manipulation, substitution, or falsification occurring after the completion of its procedure of creation, and that it is therefore what it purports to be" [10]. Even reliable records, when stored in the cloud through "black box" processes, become untrustworthy without procedures and safeguards in place to guarantee that they remain the same record that was originally stored in the cloud.

Although ensuring the continuing availability of trustworthy records is the aim of archival science, the archival concept of "trust" is much further reaching. InterPARES Trust, a "multi-national, interdisciplinary research project concerning digital records and data entrusted to the Internet" [11] defines "trust" as the "[c]onfidence of one party in another, based on alignment of value systems with respect to specific actions or benefits, and involving a relationship of voluntary vulnerability, dependence, and reliance, based on risk assessment" [12]. Ultimately, trust is not an absolute category; trust exists in degrees. "The level of trust required is proportional to the sensitivity of the material to be trusted and the adverse consequences of its lack or loss of trustworthiness" [7]. In the case of records in the cloud, it is increasingly clear that many parties are risking significant adverse consequences, with unsatisfactory assurances of the trustworthiness of CSPs. An InterPARES Trust study examining the issues of "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" ([13], p. 342) in cloud service contracts found significant gaps in how those issues—which all play a role in the broader issues of preservation—are addressed, if they are addressed at all.

If CSPs cannot be trusted (enough) with an individual's or organization's records, but they want or need to use the cloud for record creation and preservation, how do they go forward? As hinted earlier, where trust is needed for society to continue functioning efficiently, the law can serve as a surrogate source of trust between parties: "By giving legal assurances of remedies for breaches of trust, the law makes parties more likely to be both trusting (thanks to the hedging effect of the legal remedy) and trustworthy (to avoid sanctions). The broad category of institution-based trust 'is dependent on legal or other actions to enforce trusting behavior'" [8]. With the emerging legal and regulatory frameworks addressing records in the cloud, the law is in some ways acting as a regulator of trust, establishing minimum expectations for the treatment of records that users can (theoretically) rely upon even in the absence of knowledge about CSPs that could facilitate trust. Three legal regimes are emerging as dominant means of regulating trust in cloud-stored records: privacy, data protection, and data localization. This article examines each in turn, exploring how each serves to ensure that records in the cloud remain trustworthy. Although, "when it comes to digital records, trust isn't all" [7], it remains central to our ability to use records as sources of information or act upon them.

3. Discussion

3.1. Privacy and Data Protection: Old Regimes in a New Age

There exist, of course, many types of privacy, including bodily, spatial, communication, proprietary, intellectual, decisional, associational, and behavioral privacy [14]. While new technologies, such as remotely operated web cams and various forms of smart technology, have the potential to invade these forms of privacy as well, our focus here is on what is often termed "informational privacy" with regards to records and data in the cloud.

> Privacy is a concept in disarray. Nobody can articulate what it means. As one commentator has observed, privacy suffers from "an embarrassment of meanings." Privacy is far too vague a concept to guide adjudication and lawmaking, as abstract incantations of the importance of "privacy" do not fare well when pitted against more concretely stated countervailing interests.
>
> —D. J. Solove ([15], pp. 477–478).

Discussing "privacy" as a legal category is challenging at best; as Solove states, "Privacy seems to be about everything, and therefore it appears to be nothing" ([15], p. 479). The very conception of privacy is largely dependent upon context; according to Whitman, our conceptions of privacy result from our *"juridified intuitions*—intuitions that reflect our knowledge of, and commitment to, the basic legal values of our culture" ([16], p. 1160). The American *Black's Law Dictionary* defines privacy as nothing less than "the right to personal autonomy" [17]. More broadly speaking, when Canadians and Americans use the term, they are typically referring to "privacy as an aspect of liberty [...] the right to freedom from intrusions by the state" ([16], p. 1160). Thus, both Canadian and American privacy laws tend to focus on the freedom to control access to one's private life, and especially on the category of private information generally considered "personally identifiable information" (PII). It is an approach by which one can consent to the loss of privacy, and wherein the need for privacy is often explicitly counterbalanced by the need to use PII for any of a myriad of purposes. The European concept of privacy, by comparison, views it "as an aspect of dignity" ([16], p. 1160). Indeed, "there is no such word [as privacy] in French, even though the French are highly private and very much concerned about the protection of their personal information. If you look for a translation, you will find that 'privacy' is translated into French as '*intimité*', which is inaccurate, or very narrow. The French '*intimité*' is actually equivalent to 'intimacy' in English and has little to do with the American concept of 'privacy' or 'information privacy'" [18]. Unlike those of North America, European courts have voided property rights acquired through perfectly legal means to others' private images and information on the basis that their publication could harm the dignity of the person who originally consented to the use of his/her private information: "One's privacy, like other aspects of one's honor, [is] not a market commodity that could simply be definitively sold" ([16], p. 1176). The "juridified intuitions" underlying European understandings of privacy cannot abide human dignity as a commodity. The Canadian/American concept of "privacy" aligns much more closely with the European concept of "data protection."

Both the Canadian/American "privacy" and the European "data protection" seek to draw boundaries around information and records, fencing them off from public scrutiny. Such regulations—rightly—assume that not all people can be trusted with all information. For example, the Privacy Act of Canada seeks to "protect the privacy of individuals with respect to personal information about themselves held by a government institution," where "personal information means information about an identifiable individual that is recorded in any form" [19]. In an American example, the Health Insurance Portability and Accountability Act (HIPAA) focuses on "protected health information" (PHI), which is "individually identifiable health information" [20]. Traditionally, this meant controlling access to and, if and as needed, redacting paper records containing sensitive information. However, the growing impact of information and communications technologies on recordkeeping, along with a general confusion of "data" and "records," means that PII can now be treated as just a small subset of data, often divorced from its source record. This is a dangerous approach because it strips the data of its context, thereby removing the ability to determine whether the data is "private" for a particular purpose. For example, the Personal Information Protection Act of British Columbia permits covered organizations to use PII without consent if "the use is clearly in the interests of the individual and consent cannot be obtained in a timely way" ([21], Section 15(1)(a)). Determining whether the use is "clearly in the interests of the individual" is obviously a fact-driven analysis. In a simple case, the broader context of the situation might permit one to determine without the context of the record if the PII should be used—for example, pulling up the blood type of an unconscious individual who is bleeding out. However, not every case is simple. If law enforcement is trying to locate a person who might be in danger, knowing the context of the records from which addresses are being pulled is critical to determine the usefulness of those addresses. Establishing protection at record level, rather than data level, would lead to a better result.

Furthermore, data mining and other big data techniques are increasingly rendering data-level privacy protection worthless. There is a growing literature on algorithmic redlining [22], disparate impact from big data analytics ([23,24]), and the challenges posed to due process by algorithmic decision-making [25]. Presumably, no mortgage lender is out there programming decision-making algorithms to reject, for example, visible minority applicants. However, in a very real sense, they do not need to: based on the hundreds or thousands of other data points, the machine comes to same conclusions—with the same impact—as human evaluators reading coded references to an applicants' "urban lifestyle." Similarly, Target stores did not need to see a test to know if a customer was pregnant [26]. The big data decision-making algorithms' entire purpose is to discriminate, to make distinctions based on huge volumes and varieties of data.

Thus, the challenge with algorithmic discrimination—as in the case of a privacy approach that relies upon PII—is the loss of judgment. "The machine is incapable of determining whether a distinction is ethical or not. Unless we come up with a comprehensive theory of discrimination that can be represented algorithmically, we have no rigorous way of distinguishing between ethical and non-ethical machine-based discrimination [. . . however,] [s]ome of our ethical and moral criteria are so fragile, nuanced, and culturally dependent that it is not clear that the machine will ever be capable of appropriately weighing them" ([27], p. 5). Yet, the data-driven approach to PII assumes that, albeit redacting or pseudonymizing the most sensitive bits of data, we prevent the algorithm from "filling in the blanks" using the vast amounts of other data at its disposal. However, walling off bits of data as PII, while leaving all of the other data available open to whatever techniques clever data holders can devise, is a futile approach. The data-centric approach to privacy will be less and less useful in facilitating trust in cloud-based records.

The new European General Data Protection Regulation (GDPR) explicitly recognizes these challenges, and seeks to establish a higher standard of trust for its citizens. Recital 26 explicitly notes that, even though personal data may have undergone pseudonymization, "account should be taken of all of the means reasonably likely to be used [. . .] to identify the natural person directly or indirectly," distinguishing between pseudonynmized data and anonymous data [28]. Consider also that:

> Natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as Internet protocol addresses, cookie identifiers or other identifiers such as radio frequency identification tags. This may leave traces which, in particular when combined with unique identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them.
>
> —([28], Section 30)

While the GDPR does not explicitly solve the big data challenges to privacy (it is technology agnostic legislation, after all), it does provide a much firmer ground for European citizens to trust that their privacy will not be breached by clever data processors. Furthermore, the European Union provides a second line of legal protection. The new European General Data Protection Regulation (GDPR), citing Article 8(1) of the Charter of Fundamental Rights of the European Union, states that "The protection of natural persons in relation to the processing of personal data is a fundamental right" (Section 1). However, the right to the protection of personal data is an entirely separate fundamental right from the right to "respect for private and family life" [Article 7(1)]. As discussed supra, this European right to privacy is not merely data protection, nor does it support consent to data loss. Even if the sharing or use of one's private information were within the bounds of the GDPR, it could still be challenged as an assault on human dignity.

It must also be noted that, despite the problems identified above, privacy is and will remain critically important for trustworthy records in the cloud, and particularly for their security. Information security often focuses on the "CIA Triad" of confidentiality, integrity, and availability [29]. Confidentiality ensures that unauthorized subjects cannot access records; integrity requires that unauthorized modifications cannot be made to them; and availability guarantees that authorized subjects can access the records in a timely manner [29]. Ultimately, all three elements of information security—confidentiality, integrity, and availability—raise the question of who should or should not access a particular record. The answer must be determined, in large part, by privacy principles and considerations. Both legal and regulatory requirements and broader principles must undergird decisions as to which users are authorized to access specific records. Rustici states that "the GDPR makes it impossible to demarcate legal calls from data architecture choices and from business decisions" ([30], p. 2). However, the truth is that it has never been possible for organizations to have truly secure, private, trustworthy systems without considering the legal, infrastructure, and business dimensions of decisions about those systems in an integrated way.

The traditional archival function—with its focus on protecting records and their trustworthiness—has long served as an interface where the simultaneous and sometimes competing legal, infrastructure, and business needs for an organization's information could be addressed explicitly, in a principled manner. The shift from record-level privacy to data-level privacy has undermined security by fragmenting control and decision-making. Organizations are taking on legal and operational risk without even knowing that there is a risk:

> [P]ersonal data, or personally identifiable information, is scattered among the rest of a business corporate data and bundled off to unidentifiable server farms. The issue, labelled as "storage strategy", or "choice between on-premises, cloud and hybrid cloud", is handled by back office: as a result the scale of the personal data flows problem is often hidden from the only decision makers in the boardroom who have the authority to make that call and take that risk.
>
> —C. Rustici ([30], p. 9).

Securing information is impossible when no one knows where it is; protecting privacy is impossible when no one knows who should or should not have access.

3.2. Data Localization and Data Portability: A New Regime for Old Problems

In the face of concerns about the security and privacy of data in the cloud, lawmakers in jurisdictions around the world are turning to data localization measures. These legal or quasi-legal tools include laws, regulations, and policies designed to ensure that data and records are accessed, processed, and stored within a specific jurisdiction. Data localization measures attempt to assert greater control over the privacy rights of data owners whose records cross jurisdictional borders. This approach is a direct result, then, of the recognition that, when data crosses jurisdictional lines, that data might well be subject to the laws of the other jurisdiction or jurisdictions into which it crosses. In a world that is increasingly embracing the cloud, records that are there stored or accessed will most certainly cross jurisdictional lines in the absence of concerted efforts to keep them in a particular jurisdiction.

Data localization measures concern scholars and commentators for a number of reasons, including the arguments that they a) enable oppressive regimes to establish tighter control over the records of their citizens, b) stand in the face of the idealistic concept that the cloud should support a free and open internet, and c) harm the very security of data by pooling records in one place which might invite the attention of those with ill-intent. Others hold that data localization laws serve as a "pernicious form of non-tariff barrier which harms the growth of trade in a digitally powered world" [31]. Data localization advocates, on the other hand, focus on how data localization measures can help address the legal and security imbalance created by the multijuridictional reality of records in the cloud. For example, scholars at the University of Toronto have considered the risks to those who use e-communications such as cloud services. In their report titled *Seeing Through the Cloud*, the authors underscore that "... moving to the global Cloud requires Canadians or those living in Canada to forfeit their rights and protections as citizens and residents—particularly their constitutional protections..." [32]. Put another way, if a Canadian citizen entrusts his/her records to a CSP who physically stores them at a server farm located in the United States, the Canadian's records will enjoy less privacy protection than similar records stored in the same place by an American citizen. Until such a time when records stored in the cloud can remain secure and private, those that advocate the blanket cessation of data localization measures risk taking away one way to address the privacy imbalance that currently exists with records in the cloud.

The usefulness of a particular data localization law as a tool for either trust or security is of necessity fact-dependent. Data localization measures must not be viewed as being all passed for the same reasons by the same countries; much like in the case of "privacy", data localization laws embrace a diversity of legal tools which apply to different records in different ways. For example, data localization laws in British Columbia, Canada, are primarily contained in the Freedom of Information and Protection of Privacy Act (FIPPA), which is limited in its application to public bodies. Kazakhstan, by contrast, requires almost all information in the "kz" domain to be hosted domestically ([31], p. 139). The effectiveness of a particular data localization law must be evaluated in the context of that regulation and its application. That said, the general principles behind data localization laws can and should be scrutinized for their effectiveness in improving the trustworthiness and security of records in the cloud.

The first assumption of data localization laws is that, by limiting the jurisdictions in which records and data can be accessed, processed, and stored, those records will be protected from bad actors against whom another jurisdictions' laws would provide no recourse. This is a problematic assumption. Any records and data that are accessible over the Internet can ostensibly be accessed and harmed by bad actors in almost any jurisdiction. Whether or not the jurisdiction in which the records are located can provide remedy will rely on more than just localization laws. The second assumption behind data localization laws is that records hosted locally will be more secure. This, however, depends on adequate solutions and expertise being available within the jurisdiction to provide cloud services. While this is the case for jurisdictions such as British Columbia and Nova Scotia, for whom Microsoft is building local data centers [33], it is not automatically the case in every jurisdiction. Where local expertise and services are lacking, data localization laws could in fact decrease the security of records and data by forcing organizations to use less secure, less trustworthy CSPs. The third assumption behind data localization laws is that local custody is a preferable means of protecting records and data and assuring their trustworthiness. However, this assumption elides the evaluation of trustworthiness that any CSP—local or distant—should undergo. The fourth—and perhaps most important assumption—is that data localization laws provide stability should cloud services prove untrustworthy or insecure, because they provide some certainty as to which jurisdiction's laws will apply in resolving any disputes. (Though they're closely related, this differs from the first assumption: the first assumption focuses on limiting threats; the second on which law controls after a dispute has already arisen.) This is the greatest strength of data localization laws, and provides some confidence in using CSPs to host records and data. On the whole, however, data localization laws are a poor surrogate for trust and do little to enhance the security of records and data in the cloud.

4. Challenges for Algorithmic Research

Identifying and preserving digital records remain an open technical challenge. Most of the current tools—checksums, time stamping, and digital signatures, for example—leave open problems. For example, when records must be migrated, bit level changes can mean that the record's authenticity can no longer be verified. Furthermore, digital signatures, because they rely upon certificate authorities, are vulnerable if those authorities cease to exist. New algorithmic research could potentially develop solutions for verifying and authenticating records without reliance upon third parties such as certificate authorities; such algorithms could assure users that the records they entrusted to their CSPs are the records they get back, even if the records must be migrated.

One area of development for records' privacy and security in the cloud is the blockchain, the distributed ledger technology underlying cryptocurrencies such as Bitcoin and Ethereum. Because they are distributed ledgers, blockchains are essentially recordkeeping systems. Each blockchain relies on an underlying consensus mechanism—an algorithm for verifying transactions—to add transactions to the chain. If blockchain technology is going to fulfill its potential as a recordkeeping technology—suggested use cases include identity records, land records, personal property registration, and records of provenance for the Internet of Things—it will require algorithmic development that accounts for the archival principles that underlie strong recordkeeping systems. Both privacy and security will prove fundamental challenges in this area. Furthermore, many blockchain applications will still require strong support in the way of cloud technology, as the blockchain for many applications will only store pointers to external records and data; one area of development can be seen in Lemieux and Sporny's work on instantiating the archival bond in blockchain applications [34].

Ultimately, distributed computing recordkeeping—including cloud and blockchain—is going to require increasingly sophisticated means of identifying, securing, and preserving records, even as they move from system to system. It will also require tools for addressing the transjurisdictional nature of records in the cloud: while this paper looks at two models of privacy (North America and Europe, quite broadly painted), there exist many more legal and regulatory bodies with power over the records in question. The challenges—opportunities—are many.

5. Conclusions

It is highly unlikely that people and organizations are going to stop entrusting their records to the cloud, regardless of the current untrustworthiness of the CSPs. Some believe that surrogates—including law—will facilitate trust and trustworthiness. Privacy law is often cited as one such surrogate, an assertion of data subjects' right vis à vis the interests of data controllers. However, as discussed supra, the PII model will prove increasingly unworkable in the face of big data analytics, and is already flawed for the loss of context surrounding the data. Furthermore, considering the mutual dependence of privacy and security, inadequate privacy protection means that records and data in the cloud will be per se insecure. While trust always involves a degree of voluntary vulnerability, the risks involved in uncontrolled cloud based storage must be mitigated.

Given our juridified institutions, it is doubtful that Canada or the United States will ever (or would necessarily even want) to adopt a dignity-based model of privacy. One solution to the issues presented by different interpretations of privacy might be, in many ways, the oldest—a return to controlling privacy at the level of the record. Even within the most complex digital system, records can be identified, classified, and controlled (and must be, if the organization creating and maintaining them wishes to preserve and protect its information as both asset and evidence). However, that solution will require reframing the current privacy model from one focused on private "information"—a broad term that encompasses data and records both—to one focused on the co-existing challenges of private data (which undoubtedly exists outside of records) as well as private records. It will also require us to address the many concomitant problems that undermine the trustworthiness of records in the cloud and are related to ownership, custodianship, access rights beyond privacy, availability, and preservation, to name but a few. Entrusting records to the cloud should not be a fool's errand, but, to ensure that, each situation will require the expertise of archivists, technology experts, lawyers, and information professionals working together.

Acknowledgments: The article is based on the InterPARES Trust research, funded through a Partnership Grant by the Social Sciences and Humanities Research Council of Canada (SSHRC). The SSHRC grant has also covered the cost of publishing in open access.

Author Contributions: Darra Hofman has written the outline and contributed the content about privacy. Elissa How has contributed the content about data localization and portability. Luciana Duranti has integrated the content throughout and ensured consistency. The research was conducted by all authors in the context of the InterPARES Trust project. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. McKendrick, J. *Public Cloud Computing Growing Almost 50 Percent Annually, Cisco Says*; Forbes: New York, NY, USA, 2016.
2. Duranti, L.; Thibodeau, K.; Jansen, A.; Michetti, G.; Mumma, C.; Prescott, D.; Rogers, C. Preservation as a Service for Trust (PaaST). In *Security in the Private Cloud*; Vacca, J.R., Ed.; CRC Press: Boca Raton, FL, USA, 2016; pp. 47–72.
3. Quadir, S. *Bangladesh Bank Exposed to Hackers by Cheap Switches, No Firewall: Police*; Reuters: New York, NY, USA, 2016.
4. Fox-Brewster, T. *Shadow Brokers Give NSA Halloween Surprise with Leak of Hacked Servers*; Forbes: New York, NY, USA, 2016.
5. Director of National Intelligence. *Joint Statement from the Department of Homeland Security and Office of the Director of National Intelligence on Election Security*; Office of the Director of National Intelligence: Washington, DC, USA, 2016.
6. Duranti, L.; Rogers, C. Trust in Online Records and Data. In *Integrity in Government through Records Management: Essays in Honour of Anne Thurston*; Lowry, J., Wamukoya, J., Eds.; Ashgate: Farnham, UK, 2014; pp. 203–216.
7. Duranti, L.; Rogers, C. Trust in Digital Records: An Increasingly Cloudy Legal Area. *Comput. Law Secur. Rev.* **2012**, *28*, 522–531. [CrossRef]
8. Cross, F.B. Law and Trust. *Georget. Law* **2004**, *93*, 1458–1545.
9. Eastwood, T. What is Archival Theory and Why is it Important? *Archivaria* **1994**, *37*, 122–130.
10. Duranti, L. Reliability and Authenticity: The Concepts and Their Implications. *Archivaria* **1995**, *39*, 5–10.
11. InterPARES Trust. Available online: https://interparestrust.org/ (accessed on 13 January 2017).
12. Pearce-Moses, R.; Duranti, L.; Michetti, G.; Andaur, S.B.H.; Banard, A.; Barlaoura, G.; Chabin, M.-A.; Driskill, M.; Owen, K.; Pan, W.; et al. InterPARES Trust Terminology Database. Available online: http://arstweb.clayton.edu/interlex/term.php?term=trust (accessed 13 January 2017).
13. Duranti, L. What Will Trustworhty Systems Look Like In The Future? In *Building Trustworhty Digital Repositories: Tehory and Implementation*; Bantin, P.C., Ed.; Rowman & Littlefield: Lanham, MA, USA, 2016; pp. 336–350.
14. Koops, B.-J.; Newell, B.C.; Timan, T.; Skorvanek, I.; Chokrevski, T.; Galic, M. A Typology of Privacy. *Univ. Pa. J. Int. Law* **2017**, *38*, 483–575.
15. Solove, D.J. A Taxonomy Of Privacy. *Univ. Pa. Law Rev.* **2006**, *154*, 477–564. [CrossRef]
16. Whitman, J.Q. The Two Western Cultures of Privacy: Dignity versus Liberty. *Yale Law Soc.* **2004**, *113*, 1151–1221. [CrossRef]
17. Garner, B.A.; Black, H.C. *Black's Law Dictionary*; Wet: Saint Paul, MI, USA, 2009.
18. Gilbert, F.; Privacy, V. Data Protection. What is the Difference? Francoise Gilbert On Privacy, Security, and Cloud Computing, 1 October 2014. Available online: http://www.francoisegilbert.com/2014/10/privacy-v-data-protection-what-is-the-difference/ (accessed on 13 January 2017).
19. Government of Canada. Privacy Act, RSC 1985, c P-21. Available online: http://laws-lois.justice.gc.ca/PDF/P-21.pdf (accessed on 27 April 2017).
20. 45 CFR Part 160.102—General Administrative Requirements. Available online: https://www.law.cornell.edu/cfr/text/45/160.102 (accessed on 27 April 2017).
21. Government of Canada. Personal Information Protect Act, SA 2003, c P-6.5. Available online: https://www.canlii.org/en/ab/laws/stat/sa-2003-c-p-6.5/latest/sa-2003-c-p-6.5.html (accessed on 27 April 2017).
22. Yuvasri, P.; Boopathy, S. A Method for Preventing Discrimination in Data Mining. *Int. J. Adv. Res. Comput. Eng. Technol.* **2014**, *3*, 1541–1546.
23. Barocas, S.; Selbst, A.D. Big Data's Disparate Impact. *Calif. Law Rev.* **2016**, *104*, 671–743.

24. Pasquale, F.; Citron, D.K. Promoting innovation while prevenitng discrimination: Policy goals for the scored society. *Wash. Law Rev.* **2014**, *89*, 1413.
25. Citron, D.K. Cyber civil rights. *Boston Univ. Law Rev.* **2009**, *89*, 61–62.
26. Corrigan, H.B.; Craciun, G.; Powell, A.M. How does Target know so much about its customers? Utilizing customer analytics to make marketing decisions. *Mark. Educ. Rev.* **2014**, *24*, 159–166.
27. Tene, O.; Polonetsky, J. Judged by the Tin Man: Individual Rights in the Age of Big Data. *J. Telecommun. High Technol. Law* **2013**, *11*, 351.
28. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Date Protection Regulation). Available online: http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf (accessed on 27 April 2017).
29. Stewart, J.M.; Chapple, M.; Gibson, D. *CISSP: Certified Information Systems Security Professional Study Guide*; John Wiley & Sons: Indianapolis, IN, USA, 2012.
30. Rustici, C. *GDPR: The Functional Specifications of EU-Grade Privacy*; O'Reilly Media: Sebastopol, CA, USA, 2016.
31. Mishra, N. Data Localization Laws in a Digital World: Data Protection or Data Protectionism? *Public Sphere: J. Public Policy* **2016**, *2016*, 135.
32. Bohaker, H.; Austin, L.; Clement, A.; Perrin, S. *Seeing through the Cloud: National Jurisdiction and Location of Data, Serves, and Networks Still Matter in a Digitally Interconnected World*; University of Toronto: Toronto, ON, Canada, 2015.
33. Dingman, S. *Microsoft Opens Cloud Services to Select Canadian Clients with New Data Centres*; The Globe and Mail: Toronto, ON, Canada, 2016.
34. Duranti, L.; Rogers, C. Educating for trust. *Arch. Sci.* **2011**, *11*, 373–390. [CrossRef]