Preservation in the Cloud.
Towards an International Framework for a Balance of Trust and Trustworthiness

Luciana Duranti

APA / C-DAC Conference

New Delhi 4-6 February 2014

The Cloud

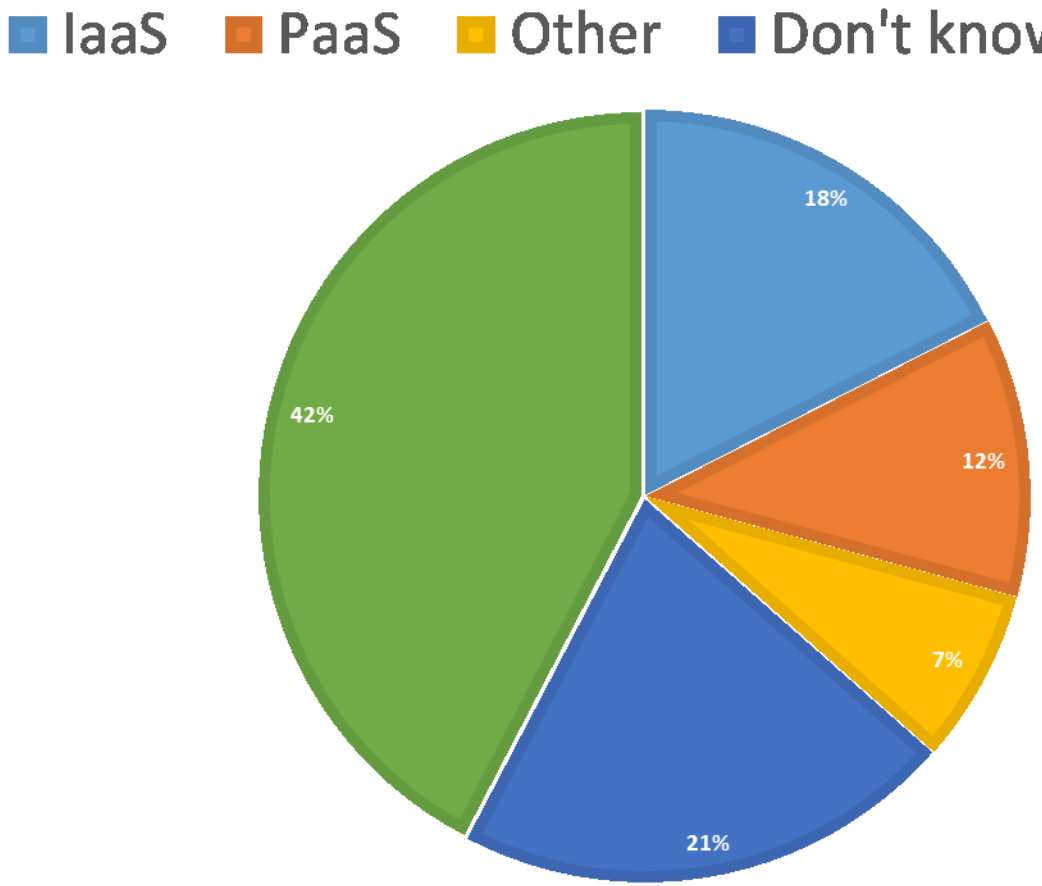
Budapest Convention on Cybercrime, 2001

Often the Internet is referred to as the Cloud. Technically this is a misuse of terms. Internet providers are "entities providing users the **ability to communicate** through a computer system **that processes or stores computer data** on behalf of such communication or users." Therefore, there are three "actions" related to the definition of provider: **communication, data processing** and **data storage**

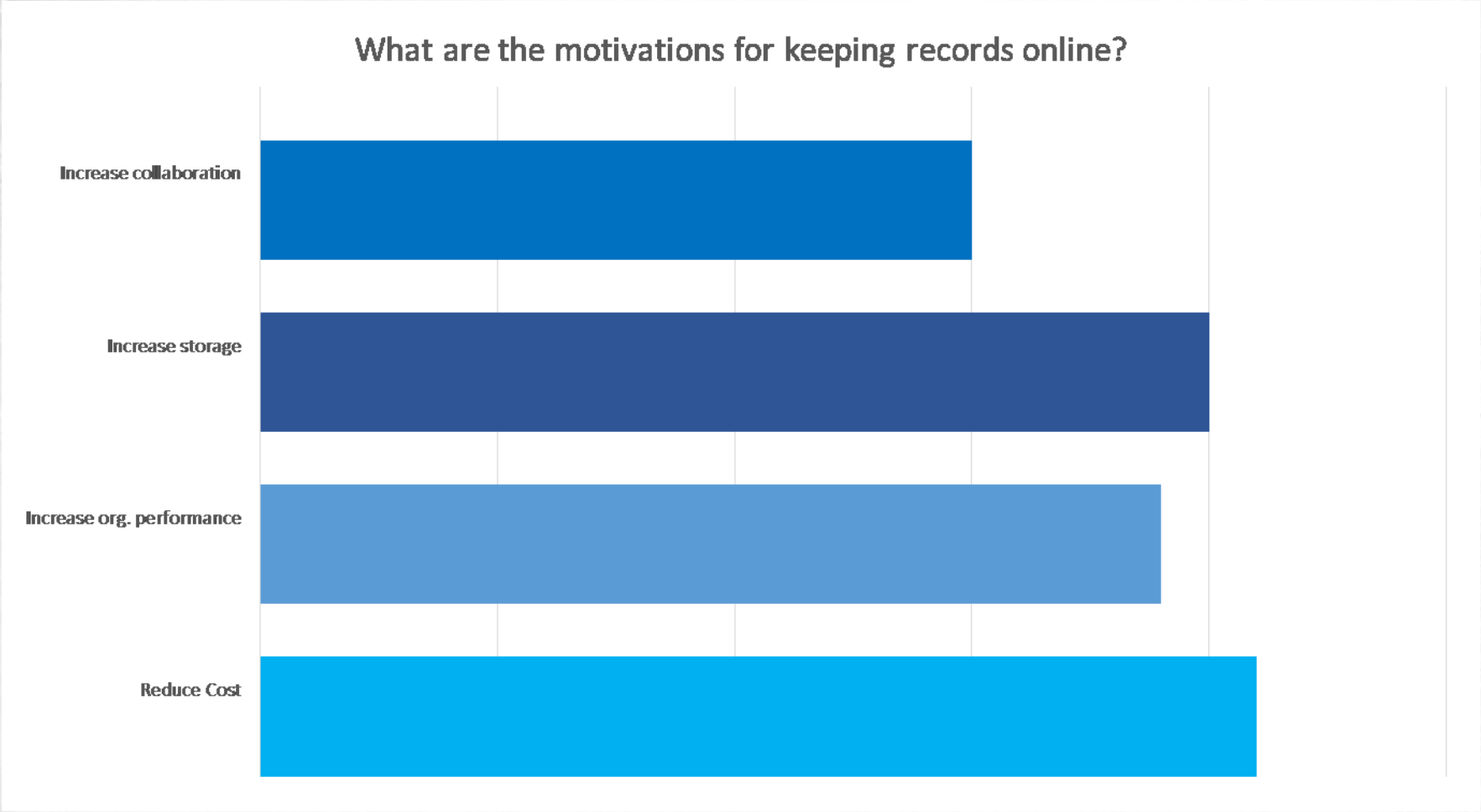
National Institute of Standards and Technology

"Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

Type of Service in Use



Motivations



Benefit

Reduced Costs

- ✓ No owning of hardware/software, so no huge upfront costs.
- ✓ Lower energy costs.
- ✓ Reduced IT personnel costs, as they don't have to implement or maintain a Record Keeping or Preservation System.
- ✓ Even in a private cloud, shared-tenant system allows pooling of resources to get more for less-better hardware/software and network.
- ✓ You can get whatever you need, and only pay for what you use.
- ✓ You can track and measure use.

Risk

Cost Issues

- ✓ If you calculate transfer, implementation and subscription, costs are not insignificant. One can get unexpected license fees.
- ✓ Variability of costs-no set monthly fee.
- ✓ There is a significant per-request charge, to motivate access in large chunks.
- ✓ In Amazon, for example, although you are allowed to access 5% of your data each month with no per-byte charge, the details are complex and hard to model, and the cost of going above your allowance is high.

For **long-term storage**: a) it can be rented, as for example with Amazon's S3 which charges an amount per GB per month; b) It can be monetized, as with Google's Gmail, which sells ads against your accesses to your e-mail; c) it can be endowed, as with Princeton's DataSpace, which requires data to be deposited together with a capital sum thought to be enough to fund its storage "for ever".

Benefit

Reliability

- ✓ Always there on demand, big or small.
- ✓ Available from anywhere, using a browser.

Risk

Provider Reliability Issues

- ✓ Public providers can go bankrupt, disappear or be sold. Your records might be gone.
- ✓ Public and private providers can lose records, and sometimes can’t get them back or backups fail.

Benefit

Security

- ✓ Security can be more robust than any one organization or unit could afford otherwise-both physical and virtual.
- ✓ Data sharding and data obfuscation requires a critical mass of data and complex technologies
- ✓ Centralized control on data easier to secure.

Risk

Security Issues

- ✓ Unauthorized access, sub contractors, hackers. It is not a matter of *if* but *when* a breach will occur. Are you told when it does?
- ✓ Documents can be stored anywhere and can be moved at any time-without you knowing.
- ✓ Encryption might not be done-in transit or in cloud. A security firm found last month that nearly 16% of the Amazon directories in which business customers store data could be perused by anyone online, revealing thousands of files containing sales records, passwords and personal data. It is a relatively new technology accessible to non-technical users.
- ✓ Shared servers could intermingle information.
- ✓ Law enforcement may seize servers for 1 person's actions. If 50 persons used it, it may take them days to get access to their records.

Benefit

Collaboration

- ✓ Allows for easy collaboration as all files are in consistent format, viewed in web browser.
- ✓ Can access and distribute information across distant geographic areas.
- ✓ Think Google Docs, Dropbox.

Risk

Control

- ✓ You have no real control over records online.
- ✓ No control over who shares your servers with you or to whom services are delegated.
- ✓ Terms of service or privacy policy may change.
- ✓ Backup may be done without you knowing and may not be disposed of as needed
- ✓ Records might be deleted without you knowing or may not be deleted according to the retention schedule.

A Question of Trust

- In fact we know very little about what happens on the Internet. The **standard of trustworthiness** for it is that of the ordinary marketplace, *caveat emptor*, or **buyer beware**

- Trust is defined in legal theory as a relationship of **voluntary vulnerability, dependence and reliance,** based on **risk assessment**

- In business, trust involves confidence of one party in another, based on **alignment of value systems** with respect to **specific benefits**

Trust in the Internet

- In everyday life, trust involves acting without the knowledge needed to act. It consists of **substituting the information that one does not have with other information**

- Trust is also a matter of **perception** and it is often **rooted in old mechanisms** which may lead us to trust untrustworthy entities

- The nature of trust relationships on the Internet is fraught with risks, weaknesses, and fault-lines inherent in the management of records and their storage in rapidly changing technologies where **authorship**, **ownership**, and **jurisdiction** may be questioned.

Questions We Should Be Asking

- How can **confidentiality** of records and data privacy be protected in the Internet?
- How can **forensic readiness** of an organization be maintained, compliance ensured, and e-discovery requests fully met?
- How can an organization's records **accuracy, reliability, and authenticity** be guaranteed and verifiable?
- How can an organization's records and information **security** be enforced?
- How can an organization maintain **governance** upon the records entrusted to the Internet?
- How can the **preservation** of records of permanent value be ensured?

The Classic Response

- Choosing the Internet is a **Risk Assessment** decision where Risk = probability x impact. It is a question of comparison. If one cannot have everything, what does one give up?

- The first choice offered us is **between Transparency and Security**: the Internet offers "trust through technology." Security involves location independence: a core aspect of Internet services delivery models.

- The second choice offered us is **between Control and Economy**: the Internet offers "trust through control on expenditures."

- But there is a necessary tension between laws that protect records in a traditional way and the abdication of custody and process without responsibility. Archives should be aware of this tension.

Archives in the Cloud

Archives are regarded as the **trusted custodians** of our documentary memory. Yet, institutions are storing their holdings in the Cloud because:

- Many of the records they are mandated to preserve already exist in the Cloud
- Access would be possible from any location to anyone who can use a browser
- A trusted digital repository satisfying ISO standards as well as basic archival preservation requirements is not affordable
- The knowledge to deal with records produced by complex technologies is not commonly available among archival professionals
- Strong protection measures are often confused with preservation measures

But "Archives in the Cloud, as well as "preservation in the Cloud," is still an oxymoron.

Archives as a Place

Justinian Code (534 A.D.)

"an archives is *locus publicus in quo instrumenta deponuntur* (the public place where records are deposited), *quatenus incorrupta maneant* (so that they remain uncorrupted), *fidem faciant* (provide trustworthy evidence), and *perpetua rei memoria sit* (and are perpetual memory of facts)"

Ahasver Fritsch (1664 A.D.)

Archives receive trustworthiness from the fact that 1) the place of storage belongs to a public sovereign authority, 2) the officer forwarding them to such a place is a public officer, 3) the records are placed both physically (i.e., by location) and intellectually (i.e., by description) among authentic records, and 4) this association is not meant to be broken.

Key Issue #1: Location Independence

A fundamental issue with keeping records/archives in the Cloud remains the distinction between the **entity responsible for their preservation and accessibility** and the **entity storing them**, and the possibility that the **jurisdiction** under which either exists is different from that in which the records exist.

India is planning to impose a ban on the use of foreign cloud-based email services to send official communications, before the end of the year. It would prevent civil servants from using Gmail, Yahoo! or Outlook.com. Instead they would be required to use a service provided by the country's own National Informatics Centre (NIC).

Brazil's president has confirmed her country plans to set up its own secure, encrypted email service to 'prevent possible espionage'.

Europe does not allow the data of European citizens to be stored outside Europe.

Models to Consider to Respect the Archival Right/Duty in the Cloud

Maritime rules of shipping, which centered on the recognition of the authority of the port state, the flag state and the coastal state

Early international maritime agreements established that the nationality of the transport vessel (the **flag state**) would establish jurisdiction, and by extension, the laws that would be in effect

Following the abuse of such rule, the **port state** was given greater control to inspect vessels coming within its territorial waters by the Law of the Sea Convention in 1982

Similarly, **coastal states** through whose waters the flagged vessels transit, have authority over the safety and competency of the ship and its crews and are also allowed inspection and enforcement while the vessel is in the coastal state's waters regardless of the flag of either the vessel (flag state) or its destination (port state)

Making an Analogy

A Canadian university could place its archives into the care of an American CSP which in turn maintains its data centers in Brazil. Following the maritime example then, the American company would be the 'flag state' that would be 'moving the goods' through 'coastal states' to their ultimate destination in the 'port state' of Brazil.

This analogy becomes problematic not only because the Canadian University owning the archives would have no jurisdiction, but also with regards to the rights of the coastal state, in that the 'pipe' used to move the records can transit through several countries (coastal states) as they are routed along the way.

Traditionally, 'coastal states' have not been granted access to inspecting packets of records as they move along the internet**. The rules of conduct then become very difficult, if not impossible, to enforce by any of the parties involved.**

Alternatives

The **territoriality principle** is not applicable because it is not possible to know the location of the records at any given time

The **nationality principle** is not applicable because nationality is an attribute of persons, not records, and the principle cannot be used to connect persons to records

The **power of disposal** principle, which "connects any data to the person or persons that obtain sole or collaborative access and that hold the right to alter, delete, suppress or to render unusable as well as the right to exclude others from access and any usage whatsoever" can be considered

By analogy, it could be possible to consider a **power of preservation principle** that assigns jurisdiction to the institutions controlling the archives as the trusted custodian and the place guaranteeing authenticity, **but jurisdiction with responsibility without custody defeats its entire purpose, even in a community cloud**

Key Issue #2: Records Trustworthiness

Reliability

The trustworthiness of a record as a statement of fact,

based on:

- the competence of its author
- the controls on its creation

Accuracy

The correctness and precision of a record's content

based on:

- the competence of its author
- the controls on content recording and transmission

Authenticity

The trustworthiness of a record that is what it purports to be, untampered with and uncorrupted

based on:

- identity
- integrity
- reliability of the system containing it

Authenticity: Identity

The whole of the attributes of a record that characterize it as unique, and that distinguish it from other records.

Identity metadata:

- names of the persons concurring in its creation
- date(s) and time(s) of issuing, creation and transmission
- the matter or action in which it participates
- the expression of its documentary relationships
- documentary form
- digital presentation
- the indication of any attachment(s)
- digital signature
- name of the person handling the business matter

Authenticity: Integrity

A record has integrity if the message it is meant to communicate in order to achieve its purpose is unaltered.

Integrity metadata:

- name(s) of persons handling the matter over time
- name of person(s) responsible for keeping the record over time
- indication of annotations made to the record
- indication of technical changes
- indication of presence or removal of digital signature
- time of planned removal from the system
- time of transfer to a the designated preserver or destruction
- time of access to the public
- existence and location of duplicates outside the system

Metadata in the Cloud

- ✓ how does metadata follow or **trace** records in the cloud from the creator to the preserver?
- ✓ how is this metadata **migrated** as a preservation activity over time?
- ✓ who **owns** the metadata created by the service providers related to their management of the records (integrity metadata)?
- ✓ Is metadata **intellectual property**? Whose?
- ✓ How can this metadata be **accessed** by the public and what are the responsibilities of the provider towards archival users?

Transparency, Stability, Permanence

Transparency:

- An unbroken chain of legitimate custody is not possible or demonstrable
- Records reliability cannot be inferred from known creation and management processes
- Records authenticity cannot be inferred from their documentary context and from a known preservation process

Stability:

- Archives requires that each record's context be defined and immutable, with all its relationships intact and this is difficult to demonstrate in the dynamically provisioned environment of the Cloud.

Permanence:

- What happens when hardware/software become obsolete? Is there a known migration plan?
- Termination of contract: how is records portability and continuity ensured?
- Termination of provider: how is records sustainability ensured?

Balance of Trust

If we decide to entrust our documentary memory to the Cloud, we must establish a **balance between trust and trustworthiness** that is valid across jurisdictions, because of the location independence which characterizes the Cloud.

The **trustworthiness** we should focus on is then not of the trustees but **of the historical records** that would be entrusted to them, keeping in mind that historical records, a society documentary memory, always start their life as current records and **their trustworthiness must be protected from creation.**

Protecting the trustworthiness of the documentary heritage of society goes **well beyond the security** guaranteed by the Cloud.

Conclusion

To establish a "balance of trust" requires enabling the development of trustworthy procedures and contractual conditions. We can do so by

- identifying the changes required in our paradigms of trust in records/archives and preservation systems, and
- developing an **internationally shared framework** that both providers and users can live by.

Only then we can require and expect **stability**, **transparency**, **accountability**, and **permanence** in addition to **security** and **economy**, develop a Trust in the Cloud founded on the Trustworthiness of the material it stores, and ensure that the expression **"Preservation in the Cloud" be not an oxymoron**.

www.interparestrust.org

THANK YOU