



Data Residency Challenges

A Joint Paper with the Object Management Group®

Table of Contents

- Introduction and Background..... 4
- Data Residency Defined.....4
- Data Residency Issues and Risks.....6
 - A Taxonomy of Sensitive Data.....6
 - Generic Data Residency Risks.....8
 - Specific Examples of Risks 13
 - Overall Risks to the IT Industry..... 15
 - The Impact of the Internet of Things..... 16
 - How Organizations Perceive Data Residency Risks 17
 - Governance of Data Residency..... 17
- Laws and Regulations18
- Applicable or Related Standards22
- Cloud Customer Needs.....24
 - Documentation and Education..... 24
 - Standards..... 25
- Challenges.....26
 - Collaboration Challenges..... 26
 - Implementation Challenges 27
 - Potential Pitfalls..... 27
- Conclusion27
- Appendix A – Laws and Regulations.....29
 - International Landscape 29
 - Specific Laws and Regulations 29
- Appendix B – References.....31

Preface

This discussion paper is issued in two closely related versions in collaboration between the Cloud Standards Customer Council and the Object Management Group (OMG, www.omg.org). This collaboration recognizes that cloud computing has become one of the key sources of data residency challenges, as is explained in the body of this paper. Cloud customers have a particular vested interest in understanding this issue and expressing their requirements to cloud providers.

The OMG version of this paper was approved for release in March 2017 [28]. The main difference between the two versions of the paper is that the version under OMG cover discusses that organization's potential roadmap for standardization. This is because the OMG, contrary to the CSCC, is a standards development organization.

Acknowledgements

The contributors to this paper are:

- Mike Abramson (ASMG)
- Brian Arbuckle
- Claude Baudoin (cébé IT & Knowledge Management), editor
- Jayant Dani (Tata Consulting Services)
- Michael DiPaula-Coyle (IBM)
- Preetam Gawade (nCryptedCloud)
- Mangesh Gharote (Tata Consulting Services)
- David Harris (Boeing)
- Sridhar Iyengar (IBM)
- Christian Jung (Fraunhofer IESE)
- Kiran Kumar Nutheti (Tata Consulting Services)
- Dennis O'Neill (Dennis M. O'Neill & Associates)
- Justin Scaduto (General Dynamics)
- Karolyn Schalk (IBM)
- Dr. Reinhard Schwarz (Fraunhofer IESE)
- Nick Stavros (Jackrabbit Consulting)
- Denise Tessier (IBM)
- Alex Tumashov (Schlumberger)
- Char Wales (MITRE)

The contributors' opinions are their own and not intended to represent the opinions or positions of their companies on the topic.

The paper addresses topics that may have legal implications for organizations. It does not constitute legal advice. The reader is advised to seek appropriate legal guidance in making decisions related to data residency.

Introduction and Background

The Object Management Group’s work on data residency started as a result of an inquiry made in the spring of 2015 by an OMG member whose business consists of deploying very large-scale, distributed databases that may span countries or regions.

It had been known for some years that the growing utilization of distributed computing resources across the globe, with data being regularly moved from country to country for a variety of reasons, raised a concern for data owners that their data and the mechanisms of its movement may violate various international, national or local laws and regulations, or at minimum expose their data to unintended access. [6] [12] [35]

Together with the member inquiry, this observation led to the formation of an OMG Data Residency Working Group, which has met on a quarterly basis starting in June 2015. Simultaneously and coincidentally, the importance of the issue was highlighted by the elaboration of a new European Union directive on data protection reform [5]. Just a few months later, the “safe harbor” provision between the EU and the United States was invalidated by the European Union Court of Justice.

The early meetings of the working group served to understand these issues, as well as to investigate issues of data residency beyond the obvious ones related to the storage of personally identifiable information (PII). In fact, we ended up writing that *“this topic [...] also concerns the right to move ‘sovereign’ data, such as oil reserves data; the international licensing of genomics data; the distribution of biometrics data for security purposes; etc.”*

In January 2016, the OMG issued a 20-question Request for Information (RFI) in order to gather input from any willing organization [27]. By then, it had been proposed that the information collected from the RFI, together with the discussions we had in our various meetings, would be transformed into an OMG discussion paper as a preliminary step before considering a roadmap for developing any appropriate standards. The results of the RFI were described in a public webinar held in July 2016.

The CSCC, in particular members of its Security Working Group, became involved in this effort from November 2016 to March 2017. The principle of issuing two almost identical papers in parallel was adopted, and this paper is the result of the CSCC part of this process.

Data Residency Defined

The RFI authors wanted to tread a fine line. On the one hand, we needed to explain in the RFI what we meant by “data residency,” lest we receive responses of lesser value due to a misunderstanding; on the other hand, we wanted to let the respondents tell us what *they* understood it to mean, thus potentially surfacing some aspects we had not thought about.

Based on the RFI itself and the responses received, we offer the following definition:

Data residency is the set of issues and practices related to the location of data and metadata, the movement of (meta)data across geographies and jurisdictions, and the protection of that (meta)data against unintended access and other location-related risks.

In this document, the terms “data” and “information” will generally be used interchangeably, in spite of the well-known distinction between “data” in the sense of raw bits and bytes, and “information” as the result of processing and interpreting data in a specific context. However, the accepted term for the subject of this paper is “data residency” while the “I” in PII or PHI stands for “information.”

The rest of this section highlights some of the scoping considerations that should support understanding the definition.

“Data residency” logically starts with (a) understanding where the data is, (b) determining what risks exist, and in particular what laws and regulations apply, based on that knowledge; but it must also include (c) controlling where the data is or goes in order to minimize such impact. Thus, data residency is not just a “spectator sport” – organizations affected by it need to establish policies and execute certain actions to address the issue.

As noted in the Introduction, “data” is not limited to PII or patient health information (PHI), or generally to data covered by data protection regulations that typically focus on the privacy of individuals. This is why **data residency is not the same as privacy, even though they are related**. The confusion between the two concepts is common and understandable, but the distinction is important:

- An organization may keep certain data within a single location, even on its own premises, thus avoiding data residency issues, and still violate privacy. For example, this would be the case of a hospital whose administrative personnel could access confidential health information about patients.
- An organization may place data that has no privacy implications in a location that causes a data residency issue. For example, this would be the case of a bank that stores corporate financial records about a client in a country whose authorities may demand access to the information as a result of a tax dispute.

Data residency issues do not only arise in the context of cloud computing solutions, even though such solutions have certainly heightened them. Even organizations that do not use cloud solutions are often exposed to data residency issues. This is true, for example, if an organization:

- consolidates multiple data centers from different countries into a single location in a different country from some of the initial locations,
- remotely backs up or mirrors data across borders to provide disaster recovery and business continuity,
- receives warranty or technical support from personnel located in another country,
- or outsources some business processes to another country, for example to lower costs or to access expertise not present at home.

Data residency issues are most often raised in the context of legal and regulatory constraints, as in the above examples. However, the above definition also covers the potential loss or degradation of reliability, service delivery, security, speed, or business continuity.

The use of the phrase “geographies and jurisdictions” should be noted. In federal states, a jurisdiction may be smaller than a country (e.g., California). In supranational organizations, a jurisdiction may be multiple countries (e.g., the European Union).

“Residency” is not synonymous with “location” – the use of the term “residency” implies the presence of issues that require addressing, and the need for best practices to address those issues, while physical “location” of data is just one element of the analysis. Data residency is also different from “data transfer” laws, which govern cross-border movement of data to protect privacy, and from “data retention” policies, which set minimum time periods to preserve information.

Data Residency Issues and Risks

A Taxonomy of Sensitive Data

It is fair to say that *all data could be affected by potential data residency laws or regulations*. This statement, however, hardly helps users or providers focus on actionable measures. Some specific types of sensitive data have been mentioned in the preceding sections. The focus of this section is to provide a more complete classification of data that is particularly likely to be subject to residency requirements.

Sensitive data (in the sense used in this paper – other laws and regulations may define “sensitive” differently for other purposes) can be identified by examining the types of regulations that may restrict the location or movement of data. While there is a vast range of specific regulation and compliance requirements across countries and organizations that need to be considered, they can be broadly classified under three main categories:

- Personal data regulation
- Trade and Customs data regulation
- Industry and Government data regulation

The same data can fall under several categories at the same time, which means all relevant regulations have to be considered. Various taxonomies of sensitive data have been defined by organizations, and include such things as shown in Table 1. In Europe, the GDPR defines additional “special categories of personal data” that are the object of specific protection (information about race, ethnicity, political opinions, religion, etc.).

This taxonomy clearly indicates that while Personally Identifiable Information (PII) and Patient Health Information (PHI) are obvious cases of regulated information, there are other types that are often overlooked. As explained in the previous section, this is why there is an important distinction between privacy and data residency.

Table 1: Taxonomy of Sensitive Data

Personal Data

Information that is used (or that could be used) to identify or trace an individual's identity. Each country may identify and regulate this information in different ways.

- Personal Identifiable Information (PII) (1)
 - Social security or national ID number
 - Name, address, phone number, date of birth
 - Driver license number
 - Biometric data
 - Financial account numbers and related banking or financial market information
 - Employment history and background check information
 - Tax-related information
 - Data associated with imagery and user-supplied Web content, including image geocoding (due to risk of re-identification)
 - Tracking/location data (2)
 - Behavior data that may be used for authentication purpose
 - Criminal records
- Personal Health information (PHI) (3)
 - Healthcare insurance ID information
 - Patient health records
 - Condition-centric registries (e.g., for communicable diseases)
 - Transactional information (visits to doctors and hospitals)
- PCI (Payment card industry) information (4)
 - Cardholder data (Primary Account Number or PAN, cardholder name, expiration date, service code, card verification value, etc.)

Information subject to trade controls

- Technology information subject to restrictions related to sanctioned and embargoed countries
- Information subject to intellectual property regulations (e.g., application source code)

Information subject to industry and government data regulations

- Military information
- Government information
- Information on natural resources, especially oil and gas reserves (geological, geophysical, petrophysical data) and information on radioactive minerals
- Marine geological data
- Other highly regulated data (varies by country) or so-called "sovereign data"

Table Notes:

- (1) “Sensitive personal data” or SPI in the U.S., or “special categories of personal data” under GDPR are a subset of PII, not to be confused with the way sensitive data is defined for this paper.
- (2) The kind of information that allows an individual person to be identified may also include dynamic or ephemeral data such as tracking or location information. As mobile and personal devices become more sophisticated and ubiquitous, the kinds of dynamic and ephemeral data that may lead to individual identification or location determination will also evolve.
- (3) Healthcare content can range from extremely sensitive (sexual health, mental health) to relatively benign public health statistics.
- (4) PCI is a proprietary information security standards for organizations that handle branded credit cards from major cards schemes. Cardholder data (CD) is any personally identifiable information associated with a person who has a credit or debit card.

Generic Data Residency Risks

What Are the Risks?

Across all business domains, and across many parts of the world, there are a number of common risks that arise when the location and movement of data are not well understood, monitored or managed:

- Violation of a government law or regulation, leading to penalties.
- Unintentional access to the information by a foreign organization.
- Demand by a foreign government to access information due to the investigation of a suspected crime, verification of tax payments, etc.
- Industrial spying by a foreign company or a foreign government.
- Exposing the sensitive data of a vendor – not just one’s own – to unintended access.
- Demand by a foreign government to be provided with secret keys to inspect encrypted data.
- Violation of “domestic content” policies – requiring the hiring of local staff or purchase of local goods and services, including hardware and software, from in-country sources.
- Increased cost of doing business in a given country, due to the obligation to set up in-country data storage – including backup or replication capabilities.
- Inability of a multinational organization to provide shared employee services, such as payroll and benefits, from a single centralized system.
- Losing business to a local competitor because of that cost, or because of customer concerns about the location of their data.
- Inability to qualify for government or private contracts that contain a data residency mandate.
- Increased risks of cyberattacks due to the multiplication of locally managed data centers with smaller and less experienced security teams.
- Diminished disaster recovery capabilities, especially in a small country or jurisdiction that may be affected in its entirety by the same event (earthquake, unrest, etc.)
- Delays in business transformation and technology modernization due to the multiplicity of data locations and disproportionate fears of non-compliance.
- Consumer and citizen mistrust of technology, organizations and governments.

Data residency may also be invoked by some IT managers as a reason to resist the adoption of cloud solutions. Their motivation may genuinely be to protect their company from exposing sensitive data, or it may be to protect the status quo – including the jobs of the people managing on-premises resources.

Finally, there is also a risk of paralysis resulting from an attempt to avoid any risk. An organization may incur higher costs or delays in procuring solutions if it overestimates the issues. The “fear, uncertainty and doubt” factor is thus a risk in itself. Managing risks does not mean eliminating them completely, but it means applying sound risk management methodologies to assess and mitigate risks until the residual risk is deemed acceptable.

Who is at Risk?

In information governance and policies, one usually distinguishes between the **data subject**, the **data controller**, and the **data custodian** or **data processor**. Data residency issues may put the data subject’s information at risk, but from a legal or regulatory standpoint it is the data controller (responsible for the purpose and means of processing personal data) and the data custodian (who operates the infrastructure where the data is stored) who are typically considered to be in violation.

The potential confusion between subject, controller and custodian is a frequent cause of finger-pointing and legal maneuvers when a data residency issue arises. The situation gets more complicated when the control chain includes additional parties such as “subprocessors.”

In some cases, one may have *as many as four actors* in different countries: the subject, the controller, the custodian, as well as the user of the system who accesses the data.

Note that this terminology can vary. According to the data protection laws of some countries, *the data owner may be the data subject* – for example, the bank is only considered the custodian of your financial record, not its owner.

What Factors Cause or Heighten the Risk?

Data residency issues tend to arise in the following situations:

- Large multinational companies wish to consolidate data centers from multiple countries into a smaller set of locations (data center consolidation).
- Organizations migrate some of their services to the cloud or to a hosted solution managed by an outsourcing company located in another country. “Services” is a very broad term here, and a risk arises simply if a remote backup solution stores the backup data in another country.
- A business process outsourcing (BPO) solution, or a managed helpdesk solution, causes agents in a different country to have access to protected information in order to perform the contracted service.
- Employees travel across borders, carrying sensitive data with them on their laptops and smartphones.

The level of risk mostly depends on the following factors:

- the nature of the data,
- the severity of the laws and regulations to which the data may be subjected,
- the nature of the services being contracted,
- the level of awareness (or ignorance) of the issues by users.

In a cloud solution, the question becomes: what knowledge of and access to the customer’s data does the custodian (cloud service provider) have, and who is at risk for storing data that may contravene local laws? What access does the provider/processor have to the data? Are they stepping in the shoes of the controller, can they access the data, or are they simply storing the data? The cloud provider may explicitly decline, by means of language in the Customer Service Agreement (CSA), any responsibility for the information stored by its customer. This exclusion is generally legitimate in the case of an infrastructure or platform service (IaaS, PaaS) – obviously even more so if the data is encrypted. In the case of software-as-a-service (SaaS), however, it is harder for a cloud provider to claim that they do not know that the client is storing certain types of sensitive information in their database. For example, the provider of a Customer Relationship Management cloud service can hardly claim that they were unaware that data such as names, addresses and phone numbers were stored in their system, since this is the very purpose of the service.

Certain countries have adopted “open data” policies that allow access to government information. For example, the directory of employees of agencies of the federal government of Mexico must be available to the public. If this requirement is combined with a decision to use a cloud storage solution based outside of the country, it may result in a conflict with laws against locating such data outside the country.

The risk level may also depend on the *currency* of the information. Storing past data in another location may be less sensitive than storing current data if the value of the information decreases with time – as is the case with stock prices, for example.

So far, we have only considered **data at rest** – that is, data that is stored in a file system or database. The risk can increase if data passes through other jurisdictions on its way between the points where it is generated, stored, and used. The mesh architecture of the Internet, and its very principles of path redundancy, mean that it may not be easy to restrict the locations through which traffic passes between international locations. Thus, **data in transit** may also pose residency issues, even though it is a largely untested area, and one that some consider only a theoretical risk. Encrypting data in transit may help, although it may be prohibited by some countries, and it is still possible to reveal sensitive information (such as who is communicating with whom) through traffic analysis.

Data Residency Use Case Matrix

Table 2 summarizes the situations that may arise based on the combination of locations we need to consider.

The table is color-coded as follows:





-  No data residency-related risk
-  Low risk – assess and monitor risk
-  Medium risk – specific measures are strongly desirable
-  High risk – strong specific measures are required

Table 2: Data Residency Use Case Matrix

Use Case Description	Data Source Location	Data Storage Location	Application Execution Location	Network Path	End User Location
Classical in-house hosted process	In-house	In-house	In-house	In-house	In-house
Hybrid Cloud execution services (data mining, seismic processing) with in-country cloud provider	In-house	In-house	In-country	In-country	In-house
In-country public cloud-based process (e.g., a CRM solution)	In-house	In-country	In-country	In-country	In-house
Outsourced (3rd party location), In-country cloud process	In-house	In-country	In-country	In-country	In-country
As above, with the data also supplied from outside of the organization's premises (e.g., data entered on an ATM)	In-country	In-country	In-country	In-country	In-country
Emerging world location with in-house hosted process, external network paths (e.g., satellite ground station, Internet routing)	In-country	In-house	In-house	External	In-house
Emerging world location with In-country cloud and external network paths (e.g., satellite ground station, Internet routing)	In-country	In-country	In-country	External	In-country
Hybrid Cloud execution services (data mining, seismic processing) with an out-of-country cloud provider	In-house	In-country	External	External	In-house
As above, with end users also located out of the country	In-country	In-country	External	External	External
Citrix/MTS access to host country. Data loading, QC (i.e., user processes/manipulates but does not typically view data)	In-country	In-country	In-country	External	External
Citrix/MTS access to host country. Metadata access (job logs, backups, DBA)	In-country	In-country	In-country	External	External
Offshore-hosted and outsourced business process	In-country	External	External	External	External
Cloud/hosting services outside of host country	In-country	External	External	External	In-house
Restricted data on mobile devices when end user is out of host country	In-House	External	External	External	External

The following definitions apply to the terms used in Table 1:

- **In-country:** the data is physically present within the boundaries of the jurisdiction in question. When this location is on the physical premises of the data custodian, it is equivalent to in-house.
- **In-house:** the data is present within the physical premises of the data custodian. Whether this is a single (computer) room, building or campus is not germane. What is germane is whether the storage, servers and network infrastructure are all privately controlled by the data custodian. As a specific case, if two locations are physically separate and connected by an Internet connection, this criterion would be violated. Examples of data sources that are in-country but not in-house are a wellsite sensor on a private owner's oil lease and an automatic teller machine (ATM) in an International airport, on a ferry or on a cruise ship.
- **External:** one or more infrastructure components (storage, servers, network) are outside the jurisdiction in question. An example is a seismic vessel acquiring data within the territorial waters of a country. The acquisition process is being monitored by personnel physically within that country. The data is transmitted via satellite to a ground station located in another country (e.g., Russian Arctic via a Norwegian ground station, offshore Indonesia via a Singapore ground station, etc.) and then via the Internet into the company's home country.

Table 2 omits certain situations that can make the use cases even more complex. For example, sensitive data may be downloaded to a laptop or a smartphone while a user is in-country or even in-house, at which point there is no data residency issue. But if the user takes his/her device to another country, the data has now left the country of origin – even if the computer is not turned on or connected to a network.

Hybrid clouds or “multi-clouds” can also render the situation more complex if some resources are in-country and others are not.

Strategic and Operational Issues in Data Center Placement

As the above matrix indicates, the relative placement of the data, the subjects of the data, the applications and the users is key in determining whether data residency issues arise and, ultimately, whether the envisioned configuration is even feasible without creating excessive business risk.

Here, we delve deeper into this issue as it affects the geographical placement of an organization's data centers – a strategic issue for most enterprises, whether acting as service providers for clients or managing their own data.

Both data protection laws and data residency regulations make it increasingly important and urgent to understand the impact of the placement of data centers on the cost of doing business. Organizations need decision models to aid in locating data centers. The data center marketplace recognizes the need for static or dynamic matching of the provider location to the customers, using their preferences and subject to the capacity of each location. The relationship between the stakeholders can be of different types:

- Scenario 1: The customers and data centers are part of the same controlling entity – for example in the case of traditional data center consolidation or of a private cloud.
- Scenario 2: The customers are individual entities while the data storage facilities are controlled by a single vendor – for example, a Google Drive or Amazon S3 type of provider.
- Scenario 3: The customers are individual entities and the storage facilities are competitors in a free market (or semi-free market if there are restrictions on location as in the EU).

In the latter scenario, the one-to-one mapping between customer and storage location would be a matching optimization problem with the aim to minimize the total vulnerability. Several solutions would exist, from customer-optimal to provider-optimal and everything in between. Characterizing these solutions from a data protection or data residency standpoint is challenging, in particular because of all the decisions that may be needed: whether to buy or rent a facility, where to expand capacity, where to backup data, and which clients to assign to each backup facility.

Specific Examples of Risks

This section highlights some industry-specific examples of risks that may not be immediately obvious or are not generally known.

Natural Resources Data

The petroleum exploration and production industry (also called “Upstream Oil & Gas”) is a complex value chain that includes “operators” (including major multinational companies, national oil companies, and small independent operators in North America), oilfield service providers, and equipment suppliers. Certain processes that lead to the estimation of reserves or the optimal operation of production are outsourced along that supply chain. A company that is contracted to collect and process geophysical measurements may wish to perform that process centrally in a large facility staffed with international experts and equipped with large computer clusters or supercomputers. Remote monitoring of operations has also become a source of higher efficiency and safety, allowing fewer personnel to be posted in potentially risky areas.

Attempts to achieve such efficiencies and improvements through remote processing and control run into the issue that the data that needs to be processed is considered a national asset – often called “sovereign data” by many countries. Examples are Russia and Venezuela, where subsurface data about oil reserves is considered a state secret. Therefore, the transmission of such data outside of the country may be prohibited or severely restricted [2]. Some of the potential impacts are as follows:

- An operating company or service company may need to build a smaller remote operating center in each such country instead of a small number of centers covering entire continents.
- An earth station may need to be installed in a country in order to receive satellite transmissions from distant in-country sites instead of being able to receive the data in another country and routing it back through terrestrial links.
- Some countries have raised the possibility of imposing a value-added tax (VAT) when raw data is sent out of the country for processing and the results (which have higher value) are sent back into the country.

Life Sciences Data

International collaboration between research organizations is common, and the sharing of data across countries is often key to permitting experiments on large enough data sets. An early example of this process is the way in which data from the Large Hadron Collider at CERN in Geneva is made available to research institutes around the world.

The sharing of such data across countries may be subject to complex licensing terms, and the fact that the data can be shipped in an intangible manner across networks makes it difficult to monitor whether such licensing agreements are being followed. All it takes is an unprotected directory on a server in location A for a researcher in location B to be able to copy the data. After all, CERN is precisely where the World Wide Web was invented in order to permit this very kind of remote access.

When the data concerns the life sciences (genetics, genomics, etc.), there is an added data residency concern, which is that the data is often originally linked to living humans. Even if the data has been “anonymized” – stripped of individually identifying information – the techniques of “reidentification” demonstrated by various researchers might be applicable in some cases. Metadata such as gender, age, location, and other markers might be sufficient to establish again the identity of the subject who donated a DNA sample described by the data. Therefore, this data may be subject to data protection laws and data residency regulations.

Law Enforcement Access to Cloud-Based Information

As mentioned, the massive expansion in cloud services over the last decade is one of the causes of data residency concerns. While users all over the world are using common cloud services such as Google Mail (Gmail), Office 365, or Facebook, the services themselves are primarily located in a few technologically advanced countries, with a strong concentration in North America and Western Europe. More generally, a user’s email and documents may now be located in a different country than the user, and even of the company that acts as the data custodian.

This was illustrated in what is known as the “Microsoft Ireland case” from 2013, in which U.S. authorities attempted to compel Microsoft to provide access to private emails stored on one of its servers located in Ireland. While the case is still moving through the judicial system, an appeals court stated in July 2016 that “§ 2703 of the Stored Communications Act does not authorize courts to issue and enforce against U.S.-based service providers warrants for the seizure of customer e-mail content that is stored exclusively on foreign server” [9]. An extra twist was added when a different ruling was reached on 5 February 2017 in a similar case involving Google [33]. Because the situation is so new and the judicial branches of government have little precedent or expertise to rely on, these various cases are still winding their way through the court hierarchy.

While law enforcement authorities always have some right of “search and seizure” when properly authorized by a court’s order, the questions raised here are (a) how far the authority of a country extends when it comes to seizing data, and (b) the potential overreach by the authorities of a government that may attempt to obtain information related to regime opponents.

Data Residency as a Pawn in International Relations

The situation gets further complicated due to the changing scope of certain regulations such as the European Union's General Data Protection Regulation. The decision by the United Kingdom to leave the European Union, following a June 2016 referendum, may lead to a situation, by the end of the formal exit process in 2019, in which UK companies are not able to house information about their European customers on UK soil, and vice versa. [3]

In some cases, the risks of data residency violations are leading cloud service providers to change their hosting policies. For example, Microsoft announced in November 2016, and opened on 24 January 2017, a "differentiated option" of its Office 365 platform hosted at a data center in Germany. [17] The data center belongs to T-Systems, a German company, which acts as trustee and manages the service in compliance with German and EU laws. Requests for data access or data deletion must comply with those laws or are denied.

It is sometimes difficult to untangle the motivations of a government in restricting the movement or storage of information pertaining to its citizens outside of its country. Such motivations may include:

- A genuine concern for the privacy of its citizens
- A wish to generate domestic jobs by forcing foreign cloud providers to set up facilities inside the country
- A desire to make it easier for law enforcement to spy on its own citizens

All three motivations were mentioned by analysts after a 17 November 2016 decision by the Russian communications regulator, Roskomnadzor, to block access to the LinkedIn professional network. [29]

In a specific case from a European organization, "the usage of an external storage solution has been rejected due to information leaving the European jurisdiction. In this case, the data resides within Europe, but information such as log data about data accesses would be handled outside Europe. The security officer rejected the use of the solution."¹

Overall Risks to the IT Industry

The above examples hint at the fact that there is a larger issue looming. The entire IT industry has been moving toward a virtualization of its services across continents, such that the location of its personnel and servers could be made independent of the location of the users and customers, with decisions made based on labor costs, infrastructure robustness and other technical and economic criteria.

Data residency issues threaten this trend. Mandated data locality could make cloud solutions or traditional outsourcing services difficult, limited, costly or too risky. For example, a mandate to store data locally in a certain country may be impractical if there are few trained data center administrators and computer security specialists in that job market. This would impact both the industry and its clients: providers would not be able to expand into certain markets, and users in those markets would lose access to cost-effective services.

¹ OMG Data Residency RFI response from the Fraunhofer Institute.

Data residency restrictions could discourage innovation and investment in modernization, either because compliance costs drag future investment, or because companies hesitate to try new technologies that could be beneficial to them or to their clients, but may exacerbate data residency issues.

In general, the IT services industry is of course in favor of the free movement of data across borders, and within certain areas, such as the European Union, it is seen as imperative for today's global economy. Overly restrictive data localization requirements can stifle competition, disadvantage companies in all sectors and seem contradictory with the Internet's distributed infrastructure that enables optimal resource distribution. From that perspective, any exceptions to the free flow of data, even if intended to protect personal data privacy, should be limited to legitimate public policy objectives and be in full compliance with the provisions of the General Agreement on Trade in Services (GATS). This being said, and as we will point out in the "Potential Pitfalls" section, consortia such as the OMG or the CSCC cannot take a position in debates related to government policies or politics.

The Impact of the Internet of Things

When asked about the future of data residency issues, there is a remarkable consensus among professionals that the Internet of Things (IoT) will raise additional issues. Since the IoT implies (a) connection to the Internet and (b) extensive data analytics, it follows that the data acquired by smart devices may be transferred to a different country or jurisdiction and may result in the discovery of certain information about the users or organizations that are the data subjects.

In the case of consumer IoT applications, once again the potential issues tend to be related to privacy. For example, data from medical monitoring devices worn by patients in a European country might be sent in the future to a support center in Asia, or to a server at the World Health Organization for detection of emerging epidemics.

People tend to be less concerned about the Industrial Internet of Things (IIoT), and yet this is an area fraught with risk because of the potential advantage that organized crime, terrorist organizations, and foreign governments may reap from accessing certain data. While blood pressure readings may not interest that many people and a patient may not care if they are stored in a foreign country, the output level of a factory or the failing pressure readings of an oil reservoir may be very useful to industrial spies. Companies that acquire and process this data may be outsourcing their data center or using cloud storage abroad. Such industrial data may actually relate to the client of a service company, not to the company itself, creating an unexpected liability when clients contract an IT service that might place the data outside the country.

In a May 2016 case, the German Automotive Association ADAC revealed intense data collection in modern cars [1]. The collected data, as well as the fact that it was collected, was mostly unknown to the car owners. ADAC analyzed data traffic from four different car brands and requested more transparency and end user control. The article caused concern in the German public, in part because in the case of foreign-made cars the data may be transmitted to a manufacturer's facility located abroad.

How Organizations Perceive Data Residency Risks

The responses received by the OMG to its RFI on Data Residency indicate a certain immaturity in how organizations view the risks. When asked whether the risk is low, medium or high, most organizations say that they view it as moderate, even when they operate in multiple countries, and that they are confident that their assessment is correct. However, when we examine in the same response documents the incomplete understanding that people have of the scope and ramifications of the issue, it is very possible that in fact most organizations *underestimate* the risk.

It is true that if one separates data residency *per se* from privacy, some organizations that have serious risks related to privacy are not incurring any substantial additional risk related to residency. This would be the case, for example, of a healthcare facility in a high-technology country that outsources its IT operations to an in-country provider and does not use an offshore helpdesk service. However, in an increasingly global economy where cloud services and managed IT services are easily supplied across borders, there are organizations whose data owners are not yet aware of the risks posed by IT's procurement decisions.

Governance of Data Residency

The general finding on governance from the OMG RFI responses is that the responsibility to identify and manage data residency issues is generally diluted across many functions in the organization. Some of the responses we obtained were:

- “A cross-functional Data Residency Working Group under the office of the General Counsel”
- “Various roles including policy, security, technical, legal, information modeling”
- “Collaboration of engineering, compliance, and legal”
- “A combination of privacy and security experts in engineering, legal, and public policy”

When multiple people across the organization are all responsible for something (not just data residency), then it is typically a sign that nothing is going to be done, since everyone is going to expect that someone else is worrying about it – a situation we can summarize as “if everyone is in charge, then no one is in charge.”

In some cases, in an attempt to assign responsibility more clearly, it is given to a senior executive (e.g., the CEO) who probably does not have the time or the expertise to make adequate strategic decisions about it.

Whether it is more appropriate for an organization to assign responsibility to an existing manager (e.g., Risk and Compliance Manager), a new role (Chief Data Officer or Data Protection Officer) or to a cross-functional team is a decision that depends on company culture as well as on the nature of the business. No generic rule can be asserted. Regardless, a mechanism and structure for data residency governance must be put in place, just like for any other enterprise information management capability such as security or privacy.

The data residency governance person or team must:

- Understand what data is potentially concerned, what its lifecycle is, and what is the impact of data protection
- Define its expectations and requirements in the form of policies or rules
- Ensure that all relevant data is covered by those policies and rules
- Understand and monitor the applicable regulations and laws (which are still rapidly evolving and vary from country to country)
- Communicate those policies so that all data stakeholders and processors understand them (a data residency implementation program will fail if those people do not understand the reasons or do not see the value of implementing the rules)
- Monitor compliance with the rules, identify any deviations and take concrete action in reaction to them
- Put in place a mechanism to handle exceptions and address concerns
- Report regularly and accurately to management
- Report to data protection and regulatory authorities as required.

To meet these requirements, we suggest the following steps:

1. Establish the governance structure – typically a team with representation from business lines, IT security, legal or compliance, etc., including the strategic, tactical and operational levels (i.e., people charged with implementation must also be part of the structure, directly or through representation) and representatives from the organization’s geographic areas.
2. Ensure proper metadata management by having a complete enterprise “data landscape” or information model in one place. A data dictionary should map the conceptual information model with actual logical/physical implementations (databases, file systems, cloud storage, etc.). As necessary, more complex models or ontologies can be developed to link models of geographically dispersed data.
3. Define all the policies and rules on sensitive data elements – what can be located where, what needs to be anonymized or encrypted, etc.
4. Establish reports to monitor the application of policies; measure how much data resides in which country or jurisdiction; and identify deviations.
5. If possible, implement tools to track the provenance and pedigree of information – and how it moves across boundaries as it gets processed and transformed.

Laws and Regulations

Data residency would be easier to understand and manage if there were only a few consistent texts to control what organizations can or must do to store and transmit their data. Unfortunately, this is far from being the case. In fact, *laws and regulations that relate to data residency are fragmented and inconsistent across countries and regions, sometimes contradictory, almost always ambiguous, quite complex, and often untested in court.*

As mentioned earlier, this complex web of laws and regulations is a response to multiple motivations, such as:

- protecting a country's citizens' personal rights to privacy by ensuring that data only flows to countries with similar guarantees of personal data (privacy) protection;
- ensuring that a government has access to the records and information related to its domiciled businesses and citizens, usually in the name of law enforcement, tax enforcement, or national security;
- protecting, favoring, or stimulating domestic employment, manufacturing industries, service providers, and/or intellectual property (IP) providers at the expense of foreign competitors, particularly those operating in innovative industries;
- monetizing the flow of data (charging transaction fees or value-added taxes).

Regardless of the mix of motivations, the result of the regulations is usually to compel organizations to keep the data they collect within a certain territory, to require companies to establish data centers and other infrastructure within the country, or potentially to prevent them from offshoring remote management and helpdesk services.

Several organizations are in the process of compiling catalogs of applicable regulations [7] [8] [10] [14] [34]. This is a Sisyphean task; no sooner is a list compiled that something changes and the list must be updated. Typically, only large companies with a significant stake in protecting their right to operate in multiple countries have the resources and motivation to undertake such an effort. Some consulting organizations, such as KPMG, publish databases and alerts for their clients to keep them current on the evolution of such laws and regulations.

The European Centre for International Political Economy (ECIPE), a Brussels-based independent and non-profit policy research think tank dedicated to trade policy and other international economic policy issues of importance to Europe, has released the *Digital Trade Estimates* database and search tool [10]. This free resource went online in late 2016 and is being adopted by the World Trade Organization (WTO). It covers 13 trade policy topics, including data transfer restriction laws and regulations, across 65 initial countries. This tool will feature a rating of how stringent each national regulation is, as well as country-to-country comparisons of laws. IBM is collaborating with this project.

As an example of how regulations change over time, the European Commission Directive 95/46/EC on data protection was a "directive," not a regulation, meaning that its implementation was basically optional and did not avoid a patchwork of local laws, some more stringent than others. The replacement of the Directive with the General Data Protection Regulation (GDPR) in 2016 gave much less leeway for individual governments to change what they could put in their national regulations. Still, in the case of the United Kingdom, the implementation of the GDPR could remain partial or even be abandoned in that country in 2019 as a result of its exit from the EU.

In addition to the rapid evolution of the laws and regulations, their *interpretation* by the courts is also evolving. The Microsoft and Google examples cited in the section, "Specific Examples of Risk," show that the guidance that can be derived from a single law may vary across jurisdictions within a single country as well as over time.

“Appendix B – Laws and Regulations” provides a snapshot of the regulations that existed as of the publication date of this document.

One particularly challenging impact of the existing web of regulations is the possibility of a conflict that cannot be resolved. For example, a bank in country A may not be able to let a citizen of country B open an account because (a) the data cannot be held in country A due to the data residency laws of country B, and (b) the data cannot be held solely in country B because the bank would not be able to comply with financial regulations of country A.

If such a situation arises, no technical solution or standard can satisfy the conflicting regulations. The only practical solution may be for the organization in country A to refrain from doing business in country B. But this may be difficult for an online business, which may not be able to determine with certainty where its clients are located, let alone what country or countries they are citizens of.

The Role of Encryption

While encrypting data is often seen as the solution to security risks in the cloud, its role in data residency is less clear, and it can be either a help or a hindrance.

From a trade control perspective, where regulations aim to prevent leakage of sensitive information to foreign companies and governments, some jurisdiction have clarified their regulations over time to declare that encrypted data can be located and transmitted anywhere. This is, for example, the case of the Canadian military. However, there are caveats to this exemption:

- the encryption technique must meet a certain standard of robustness,
- the exemption ceases as soon as the data is decrypted, which is usually required to use it.

Therefore, when employees of an organization are trained on trade compliance, they need to be taught that while the transfer of encrypted technical data from country A to country B may not be considered an “export,” it is as soon as the customer or end user in country B decrypts it and uses it unencrypted.

Moreover, the storage or transmission of encrypted data to a country other than its origin may expose the data owner or data custodian to another risk, called *mandatory key disclosure*. Certain countries give its law enforcement authorities the right to obtain a decrypted copy of any data stored in encrypted form within its borders. This can take three forms:

- Mandatory decryption – the data custodian, or sometimes a telecommunications provider or Internet service provider without the knowledge of the data custodian or data owner, may be compelled to provide decrypted content without actually revealing the key.
- Key disclosure – the government requires that the secret keys escrowed with a government agency – in effect given themselves wide latitude to examine the content of data whenever they want.

- Strength limitation – the government preserves its ability to access data through brute force decryption by making it illegal to use the stronger algorithms or key lengths.²

An additional issue with encryption is that the custodian of the information (e.g., a cloud storage solution provider) is not able to inspect the data, and therefore cannot be sure whether the data is or is not in violation of a data residency regulation. Certain cloud providers may decide to avoid the risk by forbidding the storage of encrypted data – which weakens the protection that the customer was trying to obtain.

In summary, encryption can sometimes lower the risk of violating one regulation, but it can increase the risk of violating another.

Data Residency and Trade Compliance

The transfer of data across jurisdiction may violate trade regulations rather than those specifically designated as data residency ones.

For example, in the U.S., there are requirements related to the following:

- ITAR (International Traffic in Arms Regulations) from the Department of State. This concerns military technology *and military-specific software*. There are penalties when information is accessed or released and it cannot be established where it was exported to.
- EAR (Export Administration Regulations) from the Department of Commerce. This is the equivalent of ITAR for commercial non-military technology.

Figure 1 below represents some of the scenarios in which various regulations may apply to transborder data transfers.

In this context, it may be tricky to establish who a person is, where they are, where is the data they need to access is, where the application using that data is, and therefore which regulatory requirements and artifacts govern the access request. Some of these requests might require remotely accessing a directory that, for data residency reasons, cannot leave the country of the requesting user.

The existence of “embargoed countries” also imposes certain prohibitions on the transfer of data (including software) to those countries – even though it may be difficult to track data that is lawfully sent by person A to person B, who might then re-export it to person C in an embargo country because this is not prohibited by B’s country.

² See www.cryptolaw.org/cls2.htm#busigov

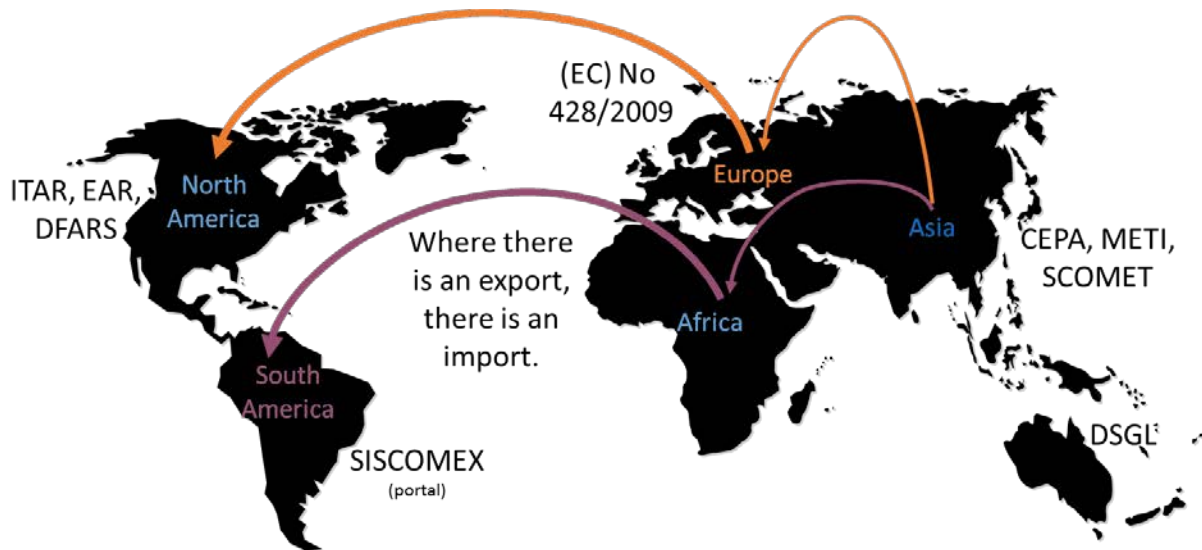


Figure 1: Trade Compliance Regulations

CEPA	Counterfeit Electronic Part Avoidance (a Boeing program)
DFARS	Defense Federal Acquisition Regulation Supplement (USA)
DSGL	Defence and Strategic Goods List (Australia)
EAR	Export Administration Regulations (USA)
EC 428/2009	European Commission Regulation on Dual Use Export Control Regime (EU)
ITAR	International Traffic in Arms Regulations (USA)
METI	Ministry of Economy, Trade and Industry (Japan)
SCOMET	Special Chemicals, Organisms, Materials, Equipment and Technologies (India)
SISCOMEX	Sistema Integrado de Comércio Exterior – Integrated Foreign Trade System (Brazil)

Applicable or Related Standards

A number of guidelines and standards have been defined and are in various stages of adoption concerning *data protection and privacy*, but there are no standards (as of this publication’s date) that specifically address *data residency* or offer technical means to detect residency issues or to enforce constraints about the location of data.

Some of the existing guidelines and standards that include aspects indirectly touching on data residency are:

- NIST Special Publication 800-144: “Guidelines on Security and Privacy in Public Cloud Computing” (Dec. 2011) [18].
- NIST Special Publication 500-299: “NIST Cloud Computing Security Reference Architecture” (May 2013) [19].
- NIST Special Publication 1500: “NIST Big Data Interoperability Framework (NBDIF)” Version 1 (September 2015) [20].
- ISO/IEC 27001 – Information Security Management.

- ISO/IEC 27017 – Code of practice for information security controls based on ISO/IEC 27002 for cloud services.
- ISO/IEC 27018 – Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors.
- Other ISO/IEC standards in the 27xxx series.

The Cloud Security Alliance (CSA) established in 2012 an International Standardization Council (ISC) whose mission is to coordinate all aspects of standardization efforts within the CSA. The ISC has established liaisons with:

- ISO/IEC JTC 1/SC 27 (IT Security Techniques)
- ISO/IEC JTC 1/SC 38 (Cloud Computing and Distributed Platforms)
- ITU-T, the Telecommunication standardization sector of the International Telecommunications Union (ITU)

As of early 2017, ISO/IEC JTC 1/SC 38 is considering starting some work on data residency and localization. The Subcommittee has inquired about the Object Management Group’s work including the present discussion paper.

Also in early 2017, the group Cloud Infrastructure Services Providers in Europe (CISPE) established a voluntary data protection code of conduct aimed at meeting the requirements of the GDPR. [30]

Consumer Reports announced on 6 March 2017 that it is leading an effort to develop a “new open-source privacy standard for the Internet of Things” [4]. The draft document, still incomplete as of this writing, is more a guideline than a standard, but it is a useful compilation of areas of concern, including security, privacy, ownership, governance and compliance.

Certain standards related to accessing information, while not written with data residency issues in mind, may provide suitable approaches to control access to and movement of information in a data residency context. Those standards include:

- OMG’s Information Exchange Framework (IEF), an initiative to establish a family of specifications for the enforcement of information sharing and safeguarding (ISS) policies that govern email exchange, file sharing, instant messaging, structured messaging, web services, etc. [24]
- XACML (eXtensible Access Control Markup Language), a standard that addresses access control, including a policy language, architecture and the processing behavior. IEF leverages XACML. As an example of how that standard might address data residency, IEF could employ XACML to express access control policies that enforce residency restrictions embedded in the metadata.
- Open Digital Rights Language (ODRL), a proposed language for digital rights management including standardization activities to express rights information over content [36].

The OMG’s Command, Control, Communication, Computers and Intelligence (C4I) Task Force started an effort in 2007 toward a specification for Data Tagging and Labelling for Security and Privacy. A Request for Information was issued in 2007 and a Request for Proposals in 2001 [22] [23]. The effort was suspended but is being revived in 2017 due to strong interest from several military organizations.

The Distributed Management Task Force’s Redfish® interface specification, designed to facilitate the management of large, scalable distributed sets of computer servers, has shown some promise to help resolve certain residency constraints. It can be used, for example, to force the remote execution of a query (for example, a user directory validation) that can only run in a certain country because of residency rules, returning a result (that is not itself residency-sensitive) to a remote user. This would address the trade compliance determination problem presented at the end of the “Laws and Regulations” section.

Software-Defined Networks (SDN) is a technology that may allow better control and monitoring of residency-sensitive traffic in a communication network. [15]

Cloud Customer Needs

Considering the input from IT providers as well as from end users, and the current lack of specific data residency standards, we attempt in this section to describe areas in which cloud customers need help in order to dispel the uncertainties created by data residency.

Documentation and Education

Generating materials to educate businesses and technical personnel about data residency, which was consistently requested by the respondents to the OMG RFI, is well aligned with the mission of the CSCC, at least to the extent that the issues are related to the use of cloud services. Generally, other situations (e.g., outsourcing or remote operations) will also benefit from such materials.

While the present paper should partially fulfill this interest, additional documents may need to be produced before enough stakeholders are sufficiently informed about the issues.

More specifically, what has been requested by RFI respondents includes **papers or presentations** that educate management, customers, regulators, legislators, suppliers, etc., about data residency, by:

- providing good definitions of data residency issues;
- dispelling myths about data residency – and by extension about cloud computing – and replacing them with objective analysis;
- explaining the issue of data in transit vs. data in storage;
- explaining the issues with storing encrypted data;
- collecting more information on how various IT service providers are currently handling the challenge.

A **catalog of international laws and regulations** on data residency is highly desired. Appendix A of this paper is a first effort, but it is necessarily a partial snapshot, at a given time, of the complex web of such regulations. A thorough effort would imply that the catalog is conscientiously maintained as an “evergreen” online publication. Note that some catalog efforts already exist, as mentioned in the “Laws and Regulations” section – but these catalogs are textual documents in a natural language, with the potential ambiguities this implies. A formal machine-readable language or ontology to encode existing catalogs could be useful to cloud customers.

A comprehensive **bibliography** is also of interest. Appendix B provides a starting point. Again, this is a static snapshot and would require a serious commitment in order to make it into a living reference.

Some respondents also asked for a **regular communication and education** tool, such as an electronic newsletter on data residency. This is outside the current capabilities of the CSCC, although the data residency mailing list maintained by OMG since 2015 can informally serve this purpose – as long as enough subscribers contribute their own findings and observations. *To subscribe to that mailing list, send a message to request@omg.org asking, in plain text, to be added to the “dataresidency” mailing list.*

Standards

Various specifications or standards may be developed by organizations such as NIST, the Cloud Security Alliance (CSA), OMG, the Organization for the Advancement of Structured Information Standards (OASIS), the World Wide Web Consortium (W3C), ISO and others. Cloud customers should participate in discussion to prioritize these efforts. When they have the technical skills necessary, they should also participate in the development of the standards so that these are not purely defined by the cloud providers.

Such standards may include, without limitation:

- A comprehensive **data classification and taxonomy** for describing sensitivity to data location and transfers.
- A **data residency maturity model**, structured in a manner similar to the Capability Maturity Model (CMM) of the Software Engineering Institute, to help organizations assess their progress toward the proper understanding and governance of data residency issues. Such a model would indicate processes and capabilities that must be deployed in order to move from one level of maturity to the next – for example, a compliance program that includes employee training and certification, similar to what many organizations have put in place to ensure trade regulation compliance.
- A formal specification of **residency metadata** that would describe the provenance of data and restrictions on its location, movement and use.
- A standard for the **specification of data residency laws and regulations**. Such a standard would allow reasoning on a set of rules, removing some of the ambiguity and difficulty of interpreting the legalese language in current documents (not to mention the fact that these documents are written in a variety of natural languages, which does not help their analysis by an affected party).

Used in combination, the last two standards – a formal description of the data sensitivity and a formal description of international regulations – might allow an inference engine to determine, for example, that “data D cannot be stored in country C.” To protect data in transit, this mechanism might be implemented in network routers to determine that certain packets must be sent down a certain path rather than another.

TOSCA (Topology and Orchestration Specification for Cloud Applications) [21] is an OASIS language to describe a topology of cloud based web services. This standard is relevant to the issue of where a particular Web service can reside, and therefore where data may be transferred as services are deployed and orchestrated. The impact of data residency on TOSCA should be examined, potentially leading to work within either OMG or OASIS – or both.

Such standards may be related to, or may exploit, existing specifications used to break down information and documents into pieces that are tracked separately – such as the S1000 and S2000 technical publication specifications for integrated engineering, manufacturing and logistics support.³

Similarly, the Hyperledger open source project to implement blockchain technologies could be useful in maintaining data access provenance in an immutable and trusted manner.⁴

Certain existing ontologies related to privacy could be adapted to include coverage of data residency issues. They include:

- the Privacy Ontology in OWL for Services (OWL-S), developed by the DARPA Agent Markup Language (DAML) program [32];
- the Privacy Preference Ontology for Linked Data, from the Digital Enterprise Research Institute in Galway, Ireland [31].

Challenges

All these efforts might eventually converge and culminate in a standard way to **manage data residency**, for example by monitoring or preventing the movement of sensitive data across a network. This is an ambitious goal, which will face two basic kinds of challenges.

Collaboration Challenges

Data residency issues may not be completely defined and understood, but they have already had a chilling effect on the willingness of some actors to collaborate openly.

This is understandable since acknowledging the existence of a data residency issue might cause customers and investors to discover a risk they were unaware of, and to question the ability of a supplier to perform as expected. In fact, several organizations with a clear stake in data residency did not respond to the OMG RFI, and others responded with the caveat that they did not want their comments associated with them. We had anticipated these concerns and offered to anonymize the comments we received. One organization even went further, asking OMG not to mention that they had responded at all.

If organizations are reluctant to disclose that they consider data residency to be an issue, they are probably going to be even less disposed to participate in a publicly observable effort to develop standards.

³ https://en.wikipedia.org/wiki/S-Series_of_ILS_specifications

⁴ <https://www.hyperledger.org/>

As we provide more education about the issue, and various incidents (such as those we mentioned about Microsoft, Google, or LinkedIn) get mentioned in the press, secrecy may become less credible as a means for organizations to claim that they are not being affected. In fact, we can hope that at some point it will be seen as a positive fact that an organization – whether it is a producer or consumer of distributed IT services – is actively contributing to the management of data residency issues.

CSCC Working Groups may provide a suitable mechanism for organizations to influence the course of action on data residency *as a group* rather than as individual contributors. This approach would “anonymize” the participants’ inputs to the data residency efforts.

Implementation Challenges

Some of the specifications that may be developed, as stated in the “Standards” section, may imply changes in software or middleware. These changes may reach deep down into the control of storage and networking infrastructure, especially if we want to handle data in transit. The disruption to the network fabric, with millions of instances of legacy equipment that may not be upgradable to handle the new specifications, may be similar to what occurred with the move from IPv4 to IPv6. References [11], [13] and [16] point to work that has been done by researchers on the subject of policies and policy enforcement at the network level.

Conversely, such new technical facilities could present a competitive advantage to the first company that would implement the new standards.

Even for data at rest, the ability to monitor, let alone prevent, the movement of data subject to residency controls to a location where the laws and regulations are in conflict is a significant challenge.

If the technical barrier to the adoption of a data residency standards is too high, IT suppliers may have trouble committing to the implementation – at least until they have no other choice because the status quo is no longer tenable.

Potential Pitfalls

Data residency is unavoidably related to the economic policies and politics of various countries and states. While CSCC members may individually express opinions or lobby legislators, the CSCC as a whole should not comment on the pros and cons of laws and regulations, let alone argue with governments or legislators about their positive and negative effects. We can educate legislators and private organizations about the issues, but the political, legal and economic inferences are theirs. The CSCC should focus on advising cloud customers about managing data residency in the existing legislative and regulatory context.

Conclusion

Data residency is an emerging and complex challenge to the movement of data and the provision of services in an increasingly interconnected world. Left ignored or unmanaged, it can complicate or impede the operations of companies whose business and supply chain extends across jurisdictions. The

IT services industry in particular is faced with a systematic risk of being prevented from offering services across borders, including but not limited to cloud solutions.

Moreover, the related laws and regulations are rapidly changing. Staying informed about the latest changes across dozens of countries is difficult. An organization may achieve compliance with all applicable laws and trade rules at a certain point in time, but run into a compliance problem at a later point due to a change in legislation (or, as in the case of Brexit, due to a country leaving a certain bloc and therefore being no longer covered by the same regulations as before).

Compliance should be explicitly governed, and no development on the technology front will remove the need for an organization to seek proper legal advice before establishing a policy and rules about data location and movement. This means that organizations need to know their data and assess how sensitive it is to its physical location.

In the case of cloud services, a Cloud Services Agreement should explicitly specify where data is going to be stored. The customer has the primary responsibility to understand its data, and to choose services which will meet *both* its business needs and the applicable regulatory requirements. Providers or data processors should work with clients to offer the necessary controls on where the data may (or may not) reside. The impact of data residency on cloud offerings means that cloud service providers have a heightened obligation to know and understand the applicable laws and regulations of the countries in which they are based, in which they have facilities, and from which their customers come. The client has the ultimate say, as data controller, whether the services offered will adequately protect their data.

While certain simple IT measures can often be implemented with today's technology (e.g., blocking certain types of traffic to/from embargo countries or countries that demand the right to inspect proprietary data), new standards or improvements to existing standards are likely to be required in the future to allow more complete management of data residency compliance. These technologies will need to be implemented in conjunction with governance.

The reader is encouraged to consider the following actions:

- Get legal assistance, understanding that data residency laws can be about more than just the protection of personal information, but applies to companies, industries, and different types of data, resulting from different governmental motivations.
- Stay informed, since laws and regulations are evolving.
- Know what kind of data is being collected, used, processed, stored, transferred, backed up or otherwise managed on your IT systems
- Be prepared to discuss and document data residency and transfer issues with any vendor providing IT services.
- Support emerging standards to address or streamline data residency – in particular, participate in related efforts by organizations such as the Cloud Standards Customer Council and the Object Management Group.

Appendix A – Laws and Regulations

Caveat: *this appendix is a snapshot of a constantly evolving landscape of international laws and regulations. As a result, the reader should check the currency of any information listed here before using it.*

International Landscape

This section describes the general principles that prevail in a number of countries that (a) represent a significant population and economic power, (b) have specific laws and regulations in place.

Those laws and regulations mostly obligate companies to store certain types of data in-country:

- Electronic health records (Australia)
- Information produced by national, state/province or local governments (Canada, Denmark, France, Norway)
- Financial transactions data (Canada, China, Germany, Indonesia, Korea, Nigeria, Russia, Turkey, Venezuela, proposed in Ukraine)
- All data related to citizens of the country (Russia, proposed in India)
- Data considered a “state secret,” such as subsurface oil and mineral reserves (Russia, Venezuela)
- All data generated within the country (China, Malaysia, Vietnam)

A particular case is that of the European Union’s General Data Protection Regulation (GDPR), which applies across all EU nations. The “Safe Harbor” provision that allowed US companies to self-certify compliance with the previous EU privacy directive was invalidated in 2015 and ultimately replaced by the EU-US Privacy Shield. The Shield is potentially subject to further legal challenges, still unresolved as of this writing.

Even within the E.U., certain countries have adopted more stringent constraints. Those may be deemed compatible with the GDPR and other European laws, as is the case of Germany; or at variance with them, as in the case of Greece, which prevents physical media created in Greece from leaving the country, a potential violation of the free movement of goods within the EU single market.

For current information, consult the European Centre for International Political Economy (ECIPE) *Digital Trade Database* [10].

Specific Laws and Regulations

Australia: Australian Privacy Principles. www.oaic.gov.au/individuals/privacy-fact-sheets/general/privacy-fact-sheet-17-australian-privacy-principles

Canada: Personal Information Protection and Electronic Documents Act (PIPEDA). www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/

China: Cybersecurity Law. www.chinalawtranslate.com/cybersecuritylaw/?lang=en

European Union: “Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.” Generally known as the General Data Protection Regulation (GDPR). <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN>

European Union: Directive 95/46/EC. Replaced by the General Data Protection Regulation since 2016.

European Union: EU Court of Justice Case C-131/12, so-called “Right to be Forgotten” ruling. <http://curia.europa.eu/juris/documents.jsf?cid=1661329>

Germany: Federal Data Protection Law (Bundesdatenschutzgesetz). www.gesetze-im-internet.de/englisch_bdsch/index.html. See in particular Section 11.

Germany: Telemediengesetz. www.gesetze-im-internet.de/tmg/.

Germany: Federal Cloud Policy (Resolution 2015/5 of the federal government’s IT Council). www.cio.bund.de/SharedDocs/Publikationen/DE/Bundesbeauftragter-fuer-Informationstechnik/IT_Rat_Beschluesse/beschluss_2015_05.pdf?__blob=publicationFile

Netherlands: Dutch Personal Data Protection Act (2016). www.akd.nl/t/Documents/17-03-2016_ENG_Wet-bescherming-persoonsgegevens.pdf

Russia: Russian Federal Law on Personal Data No. 152-FZ. <https://pd.rkn.gov.ru/authority/p146/p164/>

Russia: Ministerial Decree No. 540 (August 1992), “On Measures to Regulate Export of Geological Information on Subsoil Resources.” www.ecolex.org/details/legislation/ministerial-decree-no540-of-1992-regarding-the-arrangements-for-the-regulation-of-export-of-geological-information-on-subsoil-lex-faoc038833/

United States: Department of Commerce Export Administration Regulation (EAR). <https://www.bis.doc.gov/index.php/regulations/export-administration-regulations-ear>

United States: Federal Bureau of Investigations (FBI) – Criminal Justice Information Services (CJIS) Security Policy. <https://www.fbi.gov/services/cjis/cjis-security-policy-resource-center/view>

United States: Health Insurance Portability and Accountability Act of 1996 (HIPAA). <https://www.hhs.gov/hipaa/>

United States: International Traffic in Arms Regulations (ITAR). https://www.pmdtc.state.gov/regulations_laws/itar.html. Category XIII of the regulations equates certain “information security or information assurance systems and equipment” with munitions.

United States: Internal Revenue Service Publication 1075, “Tax Information Security Guidelines for Federal, State and Local Agencies.” <https://www.irs.gov/pub/irs-pdf/p1075.pdf>

United States: US Department of the Treasury Sanction Programs. <https://www.treasury.gov/resource-center/sanctions/Programs/Pages/Programs.aspx>

Also see references [7], [10], [14] and [34] in Appendix B below.

Appendix B – References

- [1] ADAC: “Was Geschieht mit den Fahrzeugdaten” (“What Happens to the Vehicle Data”). May 2016. <https://presse.adac.de/meldungen/technik/was-geschieht-mit-den-fahrzeugdaten.html>
- [2] Baudoin, Claude R.: “Data Residency Challenges in the Oil & Gas Industry.” OMG document datares/15-06-02, June 2015. www.omg.org/cgi-bin/doc?datares/15-06-02.pdf
- [3] Bedell-Pearce, Jack: “Data custody: the pawn in Brexit’s messy divorce.” DatacenterDynamics, February 2017. www.datacenterdynamics.com/content-tracks/design-build/data-custody-the-pawn-in-brexit-s-messy-divorce/97747.article
- [4] Consumer Reports: *The Digital Standard*. Evolving as of March 2017. www.thedigitalstandard.org/the-standard
- [5] Council of the European Union: *EU data protection reform: Council confirms agreement with the European Parliament*. Press release, December 2015. www.consilium.europa.eu/en/press/press-releases/2015/12/18-data-protection/
- [6] De Souza, Evelyn: “Data Residency and Public Cloud: Why We Care and Techniques to Think About.” Wired Innovation Insights, August 2014. <http://insights.wired.com/profiles/blogs/data-residency-and-public-cloud-why-we-care-and-techniques-to#axzz4KFEadsGr>
- [7] Determann, Lothar: “Local Data Residency Requirements for Global Companies.” Baker McKenzie, August 2015. www.bakermckenzie.com/en/insight/publications/2015/08/local-data-residency-requirements-for-global-com
- [8] DLA Piper: “Data Protection Laws of the World.” https://www.dlapiperdataprotection.com/index.html#handbook/law-section/c1_RU
- [9] Elligsen, Nora: “The Microsoft Ireland Case: A Brief Summary.” LawFare blog, July 2016. www.lawfareblog.com/microsoft-ireland-case-brief-summary
- [10] European Centre for International Political Economy (ECIPE): *Digital Trade Database*. <http://ecipe.org/dte/database>
- [11] Fraunhofer IESE: “IND²UCE – Integrated Distributed Data Usage Control Enforcement.” www.iese.fraunhofer.de/content/dam/iese/en/dokumente/Fraunhofer-IESE_IND2UCE_e.pdf
- [12] Hewlett Packard Enterprise: “Meeting Data Residency and Compliance Challenges in Global Enterprises.” www.hpe.com/h20195/V2/getpdf.aspx/4AA6-0217ENN.pdf
- [13] Hilty, M., A. Pretschner, D. Basin, C. Schaefer, T. Walter: “A Policy Language for Distributed Usage Control.” In *Computer Security—ESORICS 2007* (pp. 531-546). Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/978-3-540-74835-9_35
- [14] Information Technology Industry Council (ITI): “Data Localization Snapshot.” Updated July 2016. <https://www.itic.org/public-policy/SnapshotofDataLocalizationMeasures7-29-2016.pdf>
- [15] Internet Society: *Let SDN Be Your Eyes: “Secure Forensics in Data Center Networks.”* www.internetsociety.org/sites/default/files/01_1-paper_0.pdf

- [16] Jung, Christian, Andreas Eitel, Reinhard Schwarz: *“Enhancing Cloud Security with Context-aware Usage Control Policies.”* INFORMATIK2014: Big Data – Mastering Complexity. September 2014. www.researchgate.net/publication/269276170_Enhancing_Cloud_Security_with_Context-aware_Usage_Control_Policies
- [17] Microsoft Corp.: *“Office 365 Germany now available to help address the needs of the most regulated organizations in Europe.”* January 2017. <https://news.microsoft.com/europe/2017/01/24/office-365-germany-provides-unparalleled-data-security-regulated-organisations-europe/>
- [18] National Institute for Standards and Technology: *“Special Publication 800-144: Guidelines on Security and Privacy in Public Cloud Computing.”* December 2011. <http://csrc.nist.gov/publications/nistpubs/800-144/SP800-144.pdf>
- [19] National Institute for Standards and Technology: *“NIST Special Publication 500-299: “NIST Cloud Computing Security Reference Architecture.”* May 2013. http://collaborate.nist.gov/twiki-cloud-computing/pub/CloudComputing/CloudSecurity/NIST_Security_Reference_Architecture_2013.05.15_v1.0.pdf
- [20] National Institute for Standards and Technology: *“NIST Big Data Interoperability Framework.”* September 2015. https://bigdatawg.nist.gov/v1_output_docs.php
- [21] OASIS: *Topology and Orchestration Specification for Cloud Applications Version 1.0.* November 2013. <http://docs.oasis-open.org/tosca/TOSCA/v1.0/os/TOSCA-v1.0-os.html>
- [22] Object Management Group: *“Data Tagging and Labeling for Security and Privacy RFI.”* OMG document omg/07-09-04, September 2007. www.omg.org/cgi-bin/doc?omg/07-09-04.pdf
- [23] Object Management Group: *“Security Tagging and Labeling RFP.”* OMG document omg/11-09-04, September 2011. www.omg.org/cgi-bin/doc?omg/11-09-04.pdf
- [24] Object Management Group: *“Information Exchange Framework (IEF™).”* www.omg.org/intro/IEF.pdf
- [25] Object Management Group: *Semantics of Business Vocabulary and Rules™ (SBVR™).* May 2015. <http://www.omg.org/spec/SBVR/1.3>
- [26] Object Management Group: *“OMG Data Residency Survey.”* OMG document datares/15-09-01, September 2015. www.omg.org/cgi-bin/doc?datares/15-09-01.pdf
- [27] Object Management Group: *“Data Residency Request for Information.”* OMG document mars/2015-12-07, December 2015. www.omg.org/cgi-bin/doc?mars/15-12-07.pdf
- [28] Object Management Group: *“Data Residency Challenges and Opportunities for Standardization.”* OMG document mars/2017-03-22, March 2017. www.omg.org/cgi-bin/doc?mars/17-03-22.pdf
- [29] Petroff, Alanna: *“Russia has banned LinkedIn.”* CNN Money, November 2017. <http://money.cnn.com/2016/11/17/technology/russia-linkedin-banned/>
- [30] The Register: *“Cloud industry body sets up new data protection code.”* February 2017. www.theregister.co.uk/2017/02/16/cloud_industry_body_sets_up_new_data_protection_code

- [31] Sacco, Owen and Alexandre Passant: “A Privacy Preference Ontology (PPO) for Linked Data.” Digital Enterprise Research Institute, University of Galway, March 2011.
<http://events.linkeddata.org/ldow2011/papers/ldow2011-paper01-sacco.pdf>
- [32] SRI: “OWL for Services (OWL-S) – Security and Privacy.” Last updated February 2010.
<http://www.ai.sri.com/daml/services/owl-s/security.html>
- [33] Stempel, Jonathan: “Google, unlike Microsoft, must turn over foreign emails: U.S. judge.” Reuters, February 2017. <http://mobile.reuters.com/article/idUSKBN15J0ON>
- [34] Thomson Reuters Practical Law: “Data Protection in the United States – Overview.”
<http://us.practicallaw.com/6-502-0467>
- [35] Winstead, B.K.: “Data Residency and Legal Questions About the Cloud.” ITPro Windows, July 2011.
<http://windowsitpro.com/blog/data-residency-and-legal-questions-about-cloud>
- [36] World Wide Web Consortium: *Open Digital Rights Language (ODRL) Version 1.1*. September 2002.
www.w3.org/TR/odrl/

© 2017 Cloud Standards Customer Council.

All rights reserved. You may download, store, display on your computer, view, print, and link to *Data Residency Challenges, A Joint Paper with the Object Management Group* on the Cloud Standards Customer Council Web site subject to the following: (a) the Guidance may be used solely for your personal, informational, non-commercial use; (b) the Guidance may not be modified or altered in any way; (c) the Guidance may not be redistributed; and (d) the trademark, copyright or other notices may not be removed. You may quote portions of the Guidance as permitted by the Fair Use provisions of the United States Copyright Act, provided that you attribute the portions to the Cloud Standards Customer Council *Data Residency Challenges, A Joint Paper with the Object Management Group* (2017).