Data Controllability for Risk Management in Smart and Intelligent Systems

Manuel Munier, Vincent Lalanne
LIUPPA, EA 3000
Universite de Pau et des Pays de l'Adour E2S UPPA
Mont-de-Marsan, France

Abstract

Information has become a major asset in companies that have based their business on the production and exploitation of this information, but also in traditional companies that exploit their information with a view to continuously improving their processes. This is the case in collaborative systems where companies are interconnected but also in intelligent systems that have many information exchanges with there environment. It is important that companies keep control of the information they import, process and distribute. In this article, in the context of a risk management approach, we present a new security criterion: controllability.

1. Introduction

Nowadays, whatever their sector of activity, information has become the center of concern for companies. This concerns not only their informational capital as such, but also all information flows in and out of the company. We are now in a (digital) information society where some companies produce information while others are consumers.

In this context, the information system (IS) is the nerve center of companies. If its constituent elements (personal, hardware, software,...) make it possible to acquire, process, store and communicate information. But the main purpose of an IS is no longer limited to being a "shared storage". Depending on the level of maturity of companies with respect to their information capital, the IS function can go beyond the role of support function (operational level, data warehouse, collaborative platform) and position itself as a business partner (decision-making function, economic intelligence). Companies must consider all factors related to the effective use of information. This is all the more true in the context of collaborative systems or smart and intelligent systems (like smart buildings, smart cities). For this, the current forms of IS governance must evolve to explicitly take into account the use of information, especially from the point of view of information security.

And in our opinion, information security can no longer be based solely on computer security mechanisms (hardware, software, networks,...). We have to take into account qualitative and organizational criteria to have a global approach to information control in the company. In this article we propose a new security criterion, the controllability, to evaluate the ability of the company to control its information following a risk management approach related to the information security.

This article is structured as follows. Section 2 explains how information became a corporate asset. Section 3 introduces various tools for information security, as well as the limitations of traditional security criteria with respect to our need for information (value) control. Section 4 provides an overview of some of the work related to data quality management. In Section 5 we propose a new criterion, the controllability, to quantify the level of control of an organization in the information it handles. The definition of this criterion is based on a risk management approach by presenting the vulnerabilities, threats and risk scenarios associated with this criterion. We then give in Section 6 our vision to implement these mechanisms in an architecture where the terms of use of information are specified in contracts whose semantics are formally defined. Finally, Section 7 concludes this article by mentioning some possible perspectives for this work.

2. Information security in digital economy

In many cases, it is common not to differentiate between the words "information" and "data". This abuse of language leads us to recall the following definition: « information is a set of data aggregated for human use ». Data is the elementary description, represented in coded (digital) form, of a reality (thing, event, measure, transaction,...) intended to be:

- collected, recorded
- processed, manipulated, transformed

- stored, archived
- exchanged, disseminated, communicated

Taken individually, this basic data is rarely useful. The context must be considered for the data to become information. In order to clearly distinguish these two notions, let us take the example of information « the weather is nice ». It is built from several data such as temperature, sunshine, humidity, wind strength,... However, depending on the context (e.g., geolocation of data in London, Tahiti, Mount Everest), from the same data set we will not necessarily deduce the same information.

We can thus consider that the data (sometimes called raw data or primary data) only represent factual elements or numerical values. These are just pieces of information, not the information itself. We only talk about information once the data has been processed, interpreted, organized or structured. The way in which the data is presented also makes it meaningful or useful: the criteria for constructing the data set are clearly explained, the different column headings are explained, the codes used are translated to be understandable by a human,... And this work will be all the more usable if it is translated into metadata (literally data on data).

In the context of building management system (BMS) for example, the temperature of each room in a building is a primary datum, one piece of data. The average temperature of a floor or of the whole building (possibly taking into account the layout of the premises) is information that can be derived from the data collected.

It is obvious that information has become the center of concern of companies. This resource is now vital and strategic whatever the field of activity of the organization. This concerns not only their informational capital as such (data, business information, published content and knowledge), but also all information flows entering or leaving the company. However, if everyone indeed considers information to be one of the most important business assets of his organization, few of them manage it with the same discipline as their physical or financial assets, or even their human capital. One of the explanations for such a situation can be found at the level of the accounting profession which, although we are in the midst of the Information Age, still refuses to recognize information as a balance sheet asset, and therefore to value it as such.

To answer this problem, a new theory emerged in the late 1990s: infonomics is the concept and practice of treating information as an actual corporate asset. The analyst Doug Laney who coined this term nevertheless had an economic vision of information [1]: "for most practical purposes it makes perfect sense (and dollars) to monetize your information in a variety of ways". In his definition, infonomics aims to consider information as a new class of assets whose economic value can be measured. "Infonomics provides the framework businesses and governments need to value information, manage it and wield it as a real asset" [2]. It allows organizations to transform from just using information to optimizing it, from an economic point of view.

If there are important strategic, operational and financial advantages to having good management of information assets, the starting point is obviously the "quality" of the information. We use quotes because the term quality can represent very different concepts depending on the area of application: precision of measurements (at primary data level), consistency of data correlations to enrich them, accuracy of data interpretation to deduce information, etc. Our research axis more specifically concerns the ability of an organization to control its information, both in terms of its internal processes and in terms of its incoming and outgoing flows. Indeed, new information and communication technologies (ICT) have brought an evolution of computer systems from a standalone architecture to an architecture where the systems, belonging to various companies, are interconnected to exchange more and more information. In doing so, companies become dependent on each other from an informational perspective. Hence the importance for a company to keep control of the information it processes. However, the security criteria on which the main tools and methods for information security are based (confidentiality, integrity and availability) mainly concern the data handled, and not information in the sense of distinction made at the start of the section. In this article we therefore propose the introduction of a new security criterion: "controllability".

3. Information security tools

In this section we will briefly remind the three basic security criteria, namely confidentiality, integrity and availability. Other concepts such as traceability and accountability have subsequently been introduced in risk management methods to formalize other aspects of information security.

3.1. Key concepts

The information system is an essential asset of the organization, which should be protected. IT security is to ensure that an organization’s hardware or software resources are only used within the intended framework. To define the security of information, it is necessary to study its two components:

- **The information**, which can be presented whatever its form of storage, processing or transmission. Here we can talk about a piece of paper, an oral exchange, a binder, a digital structure coupled with a method of transmission by telecommunications…
- **Security**, evaluated by various defined criteria that qualify the security of information.

Security can indeed be qualified by different elements. We are talking here about CIA triad (for Confidentiality, Integrity and Availability), which are three terms used to set the guidelines for the security of information:

- **Confidentiality**: When we talk about confidentiality of information, it's about protecting the information from disclosure to unauthorized parties. Confidentiality is roughly equivalent to privacy.
- **Integrity**: It ensures that the information remains unchanged from its substance or form during its transmission, processing or storage. More generally, it is the guarantee that the information is "true" (in the "undistorted" sense by a third party). We therefore seek to be certain that the information is legible in its accuracy and also in its entirety.
- **Availability**: It is ensuring that the data is accessible (readable, searchable). From the point of view of the informational assets, available information has value and represents an added value and strength for the company. Information that is not or no longer searchable when we need it represents nothing and returns to the same point as non-possession of information.

Originally, these criteria were developed to assess the safety of the various constituent elements of an IS: hardware, software, network,... Computer security has then evolved into information security, but without focusing on the "value" of information.

3.2. Methods for the information security

Based on the criteria mentioned above, the information system security officer (ISSO) must be a proactive part of the company that will use risk analysis and qualification tools by classifying, on the one hand, their severity (in terms of impact and effect) and on the other hand the likelihood that they will occur. Different methods of risk analysis for an information system exist. In France, the Marion method was followed by two other methods: Mehari[1] and EBIOS[2]. In England, the preferred risk analysis method is CRAMM[3]. The United States uses OCTAVE[4]. Internationally, we use ISO/IEC 27005:2011 [5], which is an international standard that responds point by point to the requirements of ISO/IEC 27001. This is the most recent standard, moreover it is easily applicable because pragmatic.

These methods thus make it possible to qualify the risk and the impact of an attack or an event on the information system and make it possible to structure, think and implement a security that is adapted to it. They are based on the three basic criteria of the CIA triad. These criteria are sometimes not enough and alternative models have been proposed where other properties, such as authenticity, traceability, accountability, non-repudiation and reliability can also be involved (ISO/IEC 27000:2009 [6]). For Internet-scale collaborative systems these are essential concepts.

- **Traceability** (or "proof", or auditability): Guarantee that the accesses and access attempts to the considered elements are traced and that these traces are preserved and exploitable. The traceability of an information represents knowing where it comes from, where it went and where it ended.
- **Authentication**: Identifying and authenticating users (the proven fact that someone proves that they're, they assert they are) is key to managing access to relevant workspaces and maintaining trust in exchange relationships.
- **Non-repudiation** and **accountability**: No user shall be able to challenge the operations he has performed as a part of his licensed actions, and no third party shall be able to claim the actions of another user. Accountability is also a legal notion that expresses the ability to enforce something to someone else or something else. In concrete terms, accountability expresses the possibility of attributing responsibility for an act against a legal person.

3.3. Traceability and accountability

From NIST.IR.7298r2 "Glossary of Key Information Security Terms" [7], accountability is that the security goal that generates the necessities for actions of associate degree entity to be copied unambiguously thereto entity. This supports non-repudiation, deterrence, fault isolation, intrusion detection and hindrance, and after-action recovery and proceedings.

The traceability criterion (also often used under different names: accountability, auditability or evidence) has therefore been added to the three basic CIA criteria in order to identify responsibilities in case of failure with respect to one of the three CIA criteria: internal to the company, external actor,... It is mainly a legal tool to protect oneself in terms of responsibilities. Traceability rules and good practices are often determined by national or international standards and/or control bodies.

[1] Mehari (harmonized method of risks analysis) was developed by CLUSIF in 2010.

[2] EBIOS (expression of needs and identification of security objectives) [3] was developed since 1995 by the National Agency for Security Information Systems (ANSSI).

[3] CRAMM (CCTA Risk Analysis and Management Method) was developed by the British government organization CCTA (Central Computer and Telecommunications Agency).

[4] OCTAVE (Operationally Critical Threat, Assets, and Vulnerability Evaluation) [4] was developed by Carnegie Mellon University in 1999.

From our point of view, this approach to traceability remains very "static": the system collects and stores information as evidence to identify liabilities in the event of a dispute. In business intelligence (BI) this information can however also be used to calculate some ad-hoc indicators that will be reported in the dashboards. BI is the latest management catchphrase for next-generation data warehousing. Where data warehousing focuses on data integration, BI is concerned with data governance, i.e. the practice of using integrated data to make strategic business decisions about expenditures, workflow and product quality. In such a context, the control of information can be seen as a new security criterion.

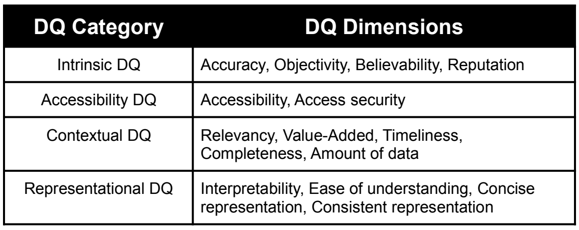
4. Literature review

In this section we will review the concepts and mechanisms necessary for the implementation and understand of the controllability criterion. Some come from the field of Computer-Supported Cooperative Work (CSCW), others from fields such as Big Data or Deep Learning. But all have in common the concern for the "quality" (in a very broad sense) of the data shared between the various components of a collaborative system.

4.1. Data Quality (DQ)

Research on data quality began in the early 1990s. Different definitions of data quality and methods of distinguishing data quality have been developed, in particular in [8], [9] (Fig. 1). Author has also defined categories related to the context of use of these data, in particular characterized by the fact that this means that utility and usability are important aspects of quality.

In our case, it is in fact the context that is important, it makes it possible in particular to determine the source of the data and the confidence that can be placed in it.



DQ Category	DQ Dimensions
Intrinsic DQ	Accuracy, Objectivity, Believability, Reputation
Accessibility DQ	Accessibility, Access security
Contextual DQ	Relevancy, Value-Added, Timeliness, Completeness, Amount of data
Representational DQ	Interpretability, Ease of understanding, Concise representation, Consistent representation

Figure1. DQ categories and dimensions

4.2. Provenance, lineage, traceability

Because the Web allows information to be shared, discovered, aggregated, filtered and circulated in an unprecedented way, it is also becoming very difficult to reliably identify the original source that produced a piece of information on the Web. Provenance is information that describes the origin or experience of a data. It appears as a "first class" criterion for Cloud content providers [10]. Lineage gathers facts, a flow of how data is or will move and transform between systems, tables, data domains. These data line diagrams often produce end-to-end flows that non-technicians would consider unusable. Traceability contains most of the components that compose the data management stack. Systems, profiling rules, tables and information columns will be extracted from their relevant systems or from a technical metadata layer. The true power of traceability (and data governance in general) lies in the information that business users can add to it.

How to materialize this traceability if not by adding metadata. A first approach was made in [11] by the constitution of an XML schema allowing to distinguish between "why" provenance (refers to the source data that had some influence on the existence of the data) and "where" provenance (refers to the location(s) in the source databases from which the data was extracted). Since then, many other provenance models have been defined, a survey of which can be found in [12]. In [13], authors describe some areas in which data provenance is finding applications and is opening up new lines of research. And [14] aims to establish a starting point towards implementing data provenance in IoT.

4.3. Trust and reputation

Trust and reputation are concepts very closed. Trust is the extent to which a party is prepared to depend on something or someone in a given situation with a sense of relative security, even if negative consequences are possible [15]. Reputation is what’s usually same or believed a few person’s or thing’s character or standing [16]. The difference between trust and reputation can be illustrated by the following perfectly normal and plausible statements [17]:

1. "I trust you because of your good reputation."
2. "I trust you despite your bad reputation."

Reputation can be considered as a collective measure of trustworthiness based on the referrals or ratings from members in a community [18]. In addition to security mechanisms to protect data (Integrity) and communications (Confidentiality), we must protect ourselves from information providers who may provide false or misleading information, yet traditional security mechanisms are not able to protect against this type of threat. Trust and reputation systems can offer protection against this type of abuse. Based on these concepts, many systems have been set up to quantify (probabilistic [19], Bayesian [20]), rank or rate [21] the reputation of information sources.

4.4. Contracts, SLA

A Service Level Agreement (SLA) represents an agreed document in which requirements about the quality of a service are established. In the context of web services, SLA often refers to the performance and security properties of the service. In general terms, the life-cycle of a machine-readable SLA is composed of a six-step approach, namely: template definition, publishing and discovering, negotiation, implementation, execution and monitoring. Each of these steps represents specific challenges either from the business or technological point of view. For example, the definition of the Quality of Service (QoS) parameters is not an easy task due to the difficulty to formalize the criteria under which the service will be evaluated. Similarly, for composed services, the SLA template definition should consider the separation of duties for each partner, a more complex monitoring of the QoS parameters, the construction of the business workflow, and negotiation strategies, specially in case of decentralized service provision [22].

Several machine-readable SLA (mainly based on XML) have been proposed in the literature, such as SLANG [23], WSLA [24], WSOL [25] and WS-Agreement [26]. Most of the existing approaches are modifications of two of the most accepted specifications, i.e. WS-Agreement and WSLA [24] including new aspects such as the number of signatory parties or new monitoring techniques. So, due to their measurability characteristic, most of the current SLAs only define performance guarantees at the infrastructure level (hardware availability, power availability, network availability, outage notification guarantee,...) as well as a minimum set of business levels including penalties and payments. Despite its limited expressiveness, SLA is an important reference for this work, allowing to identify the modeling and machine-readable representation of non-functional service requirements.

4.5. Smart and intelligent systems

The definition of Smart systems (Intelligent systems) includes a set of functions composed of sensors, actuators, all integrated in looped systems that allow to adapt and predict the system's response. Smart systems can be simple (temperature control, speed control) but are also represented through the control command of an industrial production unit.

Cyber-Physical Systems (CPS) are initially defined as smart systems but they also have important digital connections with the outside world around them. CPS can produce and use data available on the Internet: these permanent exchanges and interactions are the main key to the definition of a CPS. As an example we have the autonomous car, the industrial production units directly connected to the commercial and supply chains (Industry 4.0) up to the smart city.

Moreover, in this connected environment, employee training using innovative technologies such as virtualization, augmented reality or tele-presence robots (Education 4.0 concept) will allow training for constantly changing learning.

To illustrate these different interactions (Information and communications technology - ICT), we have taken the example of the Smart City. This concept developed in the 1990s with the emergence of the digital transition. First seen from a technical point of view (management of energy, water, transport), it was driven by digital companies who wanted to propose technological solutions to the problem of the increase in cities. Thus, in order to automate the management of these activities, the masses of data began to be collected and exchanged.

Data is usually collected via the Internet of Things (weather, traffic, Mac address of BYODs). They are typified by various quality problems and by different types of sensitive information. In [27] a framework to rationalize intelligent data management, including data collection, cleansing, anonymization and publication. The data are classified into three categories: sensitive, quasi-sensitive and open to the public.

5. A new security criterion

In this article we propose the definition of a new information security criterion aimed at capturing this idea of "control of information" contained and processed in our information system. As we said before, whatever the organization, its informational capital is an essential asset. And before implementing mechanisms to ensure its confidentiality, integrity and availability, it must first have confidence in its information and be certain of their quality and relevance. Hence the term "controllability" for "information control".

5.1. The premises

With the emergence of new needs, the interconnection of information systems is the next step. This evolution is a fact in particular through the development of service-oriented architectures (SOA) that have gained great popularity because they allow the creation of new services by composition (orchestration, choreography,...) of existing services over the Internet. Web Services (WS) is one of the most widely used technologies for such SOA. But these paradigms also raise new problems for the security of information, hence the emergence of new concepts like traceability, trustworthiness and controllability.

In [28] it has been proposed to add to Annex D of ISO/IEC 27005:2011 a new type "service" to capture the vulnerabilities and threats inherent in WS-oriented technologies and services provided. However, the threats presented mainly concerned standard CIA criteria. The concept of controllability "to ensure complete control over services used" was only a perspective. The idea of this article is to take up this notion of service by emphasizing the trust/quality/... that we can have in the information that enters (or leaves) our IS through these services.

The concept of controllability had already been introduced in [29] in the context of cloud computing via the question "Can I control my data ?". In this context, this criterion joined availability and confidentiality. The underlying questions were: "Can I act freely on my data ?", "Do I know where my data is ?", "Do I know who has access to my data ?", "Have I the property of my data ?", " Can I easily change provider ? ".

To define this concept of controllability, the rest of this section is organized in the same way as risk management approaches: definition of the (information) security criterion and proposal of a scale of needs; identification of the vulnerabilities and threats associated with this criterion in the form of a non-exhaustive list of scenarios; overview of some existing security measures.

5.2. How to define controllability ?

In the context of our work, if we always approach this notion of information control through the question "Can I control my data ?", it is nevertheless with a broader vision: "Can I trust my data ?", "What is its quality/relevance/… ?", "Am I able to protect the information it contains ?",... We therefore propose the following definition:

Definition 1: Controllability is the ability for an organization to ensure the value of its information.

Behind this definition one can find many concepts such as, but not limited to, information flow control (both internally and externally), traceability, quality management, trust management, exchanges management (e.g., workflows),…

In order to be able to assess the risks associated with a given criterion, it is first necessary to define a scale of needs. Such a scale is usually ordinal (the objects are ordered in order of magnitude, the numbers indicate ranks and not quantities) and composed of several levels to classify all the studied assets. Each level reflects a possible business need with respect to the security criterion considered. It must be easy to determine the level required for each asset. The different levels should be very explicit, unambiguous and with clear limits. The main issue regarding a scale of needs is that it is understood and usable by the people who will express the security needs of the assets. This scale must therefore be adapted to the context of the study, i.e. its development will ideally be carried out in collaboration with the people who will determine the needs. Thus, each value will have a real meaning for them and the values will be coherent. For the criterion of controllability of information, a scale of needs could be the following (Table 1):

1. **limited**: This is not to say that control of the data is useless; the information system would no longer be necessary ! Nevertheless, certain domains using fuzzy logic techniques, artificial intelligence,... can accommodate poorly controlled data. Examples: fuzzy reasoning, artificial intelligence.
2. **weak**: A lower quality and/or reliability of certain information does not jeopardize the proper functioning of the organization’s processes. Data processing used tolerate a certain amount of uncertainty in the data. Of course, the minimum level required will depend on the resilience capabilities of the system in managing incomplete or uncertain data. Examples: Big Data, statistical processes.
3. **strong**: The quality of the data is essential for the proper functioning of the processes. Any uncertainty can have a significant impact on the business of the organization. Examples: customer orders, software development, smart grids, smart and intelligent systems and applications.
4. **absolute**: Uncontrolled data engages the responsibility of the company, which can thus jeopardize its durability. Even minimal uncertainty can have major consequences for both the organization and its ecosystem. Examples: medical data, nuclear power plant.

Table 1. Required level of controllability

Level	Detailed description
1 – limited	The quality or reliability of the information is not necessary.
2 – weak	It is possible to accommodate some erroneous and/or unreliable information without impacting the processes.
3 – strong	The quality of the data is essential for the proper functioning of the processes.
4 – absolute	Uncontrolled data can cause damage to the sustainability of the business.

5.3. Threats, vulnerabilities and consequences

In the previous section we gave a definition for the controllability criterion as well as an example of a scale of security requirements related to this criterion. The next step is to illustrate the integration of the controllability criterion into a risk management approach alongside traditional CIA criteria. A risk is usually defined using what is called the "risk equation":

$$Risk = T hreat \times V ulnerability \times Impact$$

To fully understand the concept of risk, it is important to look at each of its components. First of all the threat (the source of the risk) is the possible attack of a dangerous element for the assets. It is the agent responsible for the risk. Vulnerability is then the characteristic of an asset constituting a weakness or vulnerability to security. Finally the impact represents the consequence of the risk on the organization and its objectives. Threat and vulnerability, representing the cause of risk, can be qualified in terms of likelihood. The impact can be qualified in terms of level of severity.

After the asset identification step to inventory all assets within the scope of the audit, the threats must be identified. A threat is likely to damage assets such as information, processes and systems and, therefore, organizations. The sources of both accidental and deliberate threats should be identified. Some threats may affect more than one asset and may therefore have different consequences depending on the asset assigned. The identification of these threats can be based on a standard threat repository such as Appendix C of the ISO/IEC 27005 standard.

We will not dwell here on the step of identifying the existing security measures performed during the risk assessment, which will be discussed in the next section.

The next step is to identify vulnerabilities that can be exploited by threats to harm assets or the organization. It should be noted that a security measure improperly implemented, or malfunctioning, or incorrectly used may be a vulnerability. A security measure may or may not be effective depending on the environment in which it is implemented. Conversely, a threat that does not match any vulnerability may not result in risk. Examples of vulnerabilities and vulnerability assessment methods are available in Appendix D of the ISO/IEC 27005 standard.

To identify the consequences we introduce the notion of "incident scenario". An incident scenario is the description of a threat exploiting a vulnerability. The impact of this incident must be identified. This leads to the creation of a list of incidents, with their respective impacts. An incident scenario can affect one or more assets. The consequence assessment then consists in quantifying as much as possible the consequences of the previously identified scenarios. Naturally, the valuation of the consequences is directly linked to the value of the assets concerned. The consequences will be much more serious for assets considered to be very important than for assets whose value is lower. Finally, for each scenario, it will be necessary to estimate the likelihood of its occurrence.

Then, we can estimate the risk level, which consists in relating the consequences of the scenarios to their likelihood. This will result in a list of all the risks to which a value will be assigned. This is called "risk levels". Finally, the risk assessment stage will draw up a list of risks, valued and, above all, ranked objectively in order of importance.

In the context of such a risk assessment process, our objective in this section is to propose different generic scenarios presenting threats, vulnerabilities and consequences related to the controllability criterion. Through these scenarios, the idea is to highlight incidents that are not captured by the CIA criteria. This work on scenarios could also result in an extension of the threat and vulnerability repository (example: ISO/IEC 27005 – Annex D) as shown in Table 2.

1. **Lack of traceability**: Traceability of information from external services is not provided. As a result, the reliability and/or trust of the information can not be guaranteed. The company therefore incurs the risk of generating and disseminating information calculated on the basis of unreliable data (and thus to engage its own responsibility).

2. **Accuracy of data not guaranteed**: Like personal data in the context of the GDPR[5], the principle of data accuracy is important. There are obvious risks to IS controllability if inaccurate data are processed. Therefore controllers are responsible for ensuring that the processing performed is done with accurate and up-to-date data if necessary. In this case, every effort must be made to update the data and erase or rectify the data that is out of date.

3. **Lack of resilience**: The resilience of a computer system can be defined as the "ability of the system to perform well in the presence of a limited number of faults". Although in information system security we immediately think of software or hardware defects, we can also talk about resiliency with respect to the quality, reliability, confidence of the manipulated data.

4. **No contract**: From the legal point of view, only the existence of a contract makes it possible to invoke contractual liability in the event of a dispute between the customer and the provider, in particular for the identification of responsibilities.

5. **Absence of implementing rules**: The client-provider relationship has actually been contracted, but the clauses setting the terms and conditions for the delivery of the service are not sufficiently precise. If the provider "does not respect deadlines" (very subjective notion, suddenly), the customer may experience delays in his business process without the possibility of appeal.

6. **Obligations of parties to the contract are imprecise**: Suppose that the provider broadcasts, in good faith (e.g., subcontracting), information provided by the customer for the proper performance of the service, but that this information is confidential. If the obligations of the parties to the contract are not specified (e.g., confidentiality obligation), the customer's liability may be incurred.

[5] The General Data Protection Regulation (GDPR) is the new European regulation that aims to strengthen the protection of personal data.

7. **No acknowledgment of receipt**: For its part, the provider may also ask the customer to validate the service rendered: acknowledgment of receipt of information, receipt (by the customer) attesting to the "quality" of information received,... If this is not the case, it will be difficult for the provider to prove its good performance of the contract.

8. **Poor data quality**: The results are not usable because the quality of the data processed is not sufficient to meet the required level.

Table 2. Threats and vulnerabilities related to data controllability

Type	Examples of vulnerabilities	Examples of threats
Software	(3) Lack of resilience	Computation of erroneous information
Service	(1) Lack of traceability of the service provided	Breach of trustworthiness of information
	(4) No contract	Breach of trustworthiness of information
	(5) Absence of implementing rules	Breach of trustworthiness of information
	(7) No acknowledgment of receipt	No proof of contract performance
	(8) Poor data quality	Computation of erroneous information
Personnel	Insufficient security training	Error in use (erroneous information entry)
	Lack of security awareness	Error in use (erroneous information entry)
Organization	(2) Accuracy of data not guaranteed	Computation of erroneous information
	(6) Obligations of parties to the contract are imprecise	Dysfunctioning relationships with other actors

Once the risks have been analyzed and evaluated in order to prioritize and rank against their evaluation criteria, the final step will be the treatment of the risk. This is the process of selecting and implementing controls. This will firstly be done through the identification of security objectives: determination of the modes of treatment and taking into account the elements of the context. The objectives thus identified will constitute the specifications of the risk treatment process. Then, security requirements will be determined to meet the security objectives and describe how to deal with the risks. To define the treatment options, the risk and the cost of treatment must be matched. There are four options for dealing with risk:

- **Refusal** or **avoidance**: the risk considered is too high, the activity leading to the risk must be removed.
- **Sharing**: the risk will be transferred to another entity (an insurer, a subcontractor) capable of managing it.
- **Reduction**: the risk must be reduced. This is to reduce the impact of the risk so that the (residual) risk is acceptable.
- **Risk retention**: the risk is maintained as is.

The risk treatment plan thus established must also assess the residual risks. This plan will be finally submitted to the acceptance decision of the organization's managers.

5.4. Existing controls

Obviously, a risk assessment methodology like ISO/IEC 27005 assumes that the company has already listed all its existing controls (and their effectiveness) before applying new ones. During a risk analysis, this stage takes place between the identification of the threats (on the assets) and the identification of the vulnerabilities (with regard to the controls already implemented). But in this article, for the sake of understanding, we preferred to present the controllability criterion first. Controls that are planned to be implemented according to the risk treatment implementation plans should be considered in the same way like those already implemented.

Although the concept of controllability as presented in this article is a new criterion, much work has already been done on the notion of "quality" of data (in the broad sense). A synthesis was presented in Section 4. The first area that immediately comes to mind is obviously the data quality itself seen from the perspective of indicators such as accuracy, completeness, reliability, relevance and how up to date it is. Another area of research is traceability, which is a general term indicating something can be linked to another artifact. Traceability has been broken down into several subdomains, including the data provenance and the data lineage (more details in Section 4.2).

Finally, another interesting area deals with the contractualization of interactions between the different actors of a collaborative activity. Many studies have studied Service Level Agreements in the context of traditional client-provider relationships. Another track [30], [31] proposes to provide a formal model for contracts not only indicating the obligations of the parties, but also the requirements, purposes and expected evidence. With well-defined semantics for service contracts in order to express controllability policies, this approach aims to provide both tools for monitoring the proper execution of contracts and "traceability" mechanisms to, a posteriori, identify the responsibilities in case of litigation for example.

5.5. Methodology

Whichever method is used (EBIOS, ISO/IEC 27005,...), the complete development of a risk analysis and its various stages as presented in Section 5.3 can be a tedious job to complete. The complexity of such an approach is partly due to the comprehensiveness of the elements to be taken into account (vulnerabilities, threats, consequences) and the simultaneity of the analyzed security concepts (confidentiality, integrity, availability criteria, different gravity scales). Addressing all of these points at the same time is usually confusing for a non-expert person. In order to make the analysis more accessible, it should be possible to target it to a specific need; in our case, the control of information.

With the recent entry into force of the GDPR (General Data Protection Regulation), we have seen similar concerns. The application of the new European Regulation GDPR requires, in part, the completion of a Privacy Impact Assessment (PIA). This approach is based on a security risk analysis (in the cybersecurity sense) focused solely on the risks affecting personal data and their impact on the rights and freedoms of the persons concerned by these data. It is based on the EBIOS methodology and can be summarized in 4 steps:

1. **Context** (scope of the PIA): Describe the processing(s) of personal data under consideration, its (their) purposes and stakes, the data retention periods, the responsibilities related to the processings. Describe the processes and personal data supporting assets (hardware, software, networks, people, papers) for the entire personal data life cycle (from collection to erasure). The purpose of this step is also to evaluate "the necessity and the proportionality of the processing operations with regard to the purposes".
2. **Controls** (the compliance components): Identify existing or planned controls. These controls are the legal controls imposed by the regulation (rights of persons and information to be provided to them) and the risk-treatment controls to protect them (organizational and technical).
3. **Risks** (potential privacy breaches): The proposed definition of risk breaks it down into two parts to facilitate identification and analysis. On the one hand, we evaluate what is feared about the processes (loss, disclosure, data corruption with the impacts on privacy) and the level of severity of these dreaded events. On the other hand, the threats and their source targeting the assets of the processes and which can lead to the dreaded events. The likelihood of realization of these scenarios is evaluated by considering identified or potential vulnerabilities and security controls. The combination of dreaded events and threat scenarios provides risks, and risk levels are assessed by considering the likelihood of the scenario and the severity of the impacts identified.
4. **Decision** (validation of the PIA): Decision-making consists of validating the choice of existing and planned controls to address risks. Thus, if the controls are not sufficient, an action plan is defined to propose new ones. The analysis is then revised to take into account these new parameters, until the risk levels make it possible to make the decision to accept them.

A PIA is clearly a lightened risk analysis that focuses solely on the security of personal data (whether digital or paper). However, since this approach respects the terminology and principles of a traditional risk analysis, a PIA can then be easily integrated into an EBIOS or ISO/IEC 27005 analysis to take into account the "personal data" aspect.

As part of our work, our goal will be to develop a similar approach: a security risk analysis focused solely on risks aimed at controlling information within the organization. This analysis will take up the specific points (vulnerabilities, threats, consequences) that we presented in Section 5.3.

6. Our approach for an implementation

As we have already mentioned, the problem that we address in this article concerns the control of information exchanges between the systems managed by different organizations. In certain research areas we also speak of collaborative systems or systems of systems. The underlying idea is that each system is completely independent of the others, except for exchanging information. We are now in a (digital) information society where some companies produce information while others are consumers.

The evolution of ICT has spawned new economic relationships where information is now seen as an economic good. The production and development of these information goods are the source of significant economies of scale and their network distribution is the source of powerful club externalities (network effect: the usefulness of a good for an agent depends on the number of other users). In this perspective, the relevant paradigm does not consist in thinking of ICT as the natural technological support of free trade, but rather as the instrument of "distributed coordination" between agents, a coordination without explicit institutional representation.

And it is precisely this absence of overall control that is at the heart of our approach. Whether producing or consuming information (or both), each company has its own objectives, its own economic challenges, its own constraints, etc. It is a completely decentralized peer-to-peer architecture where nodes can negotiate the terms of information exchange with each of their partners on an equal basis. The conditions of use are generally contractualized in an SLA. However, due to their measurability characteristic, most of the current SLAs only define performance guarantees at the infrastructure level (hardware availability, power availability, network availability, outage notification guarantee,...) as well as a minimum set of business levels including penalties and payments.

In a very simplified way, we can consider that SLAs mainly specify the conditions of access to information. Our approach consists in being inspired by this concept of contract to define clauses which this time relate to the controllability of information in order to take into account the "qualitative" aspect and the use of information: usage control policy, commitments / proofs / penalties / etc. on compliance with the rules set. In such a decentralized architecture, each actor is free to negotiate the terms of each of its relationships with the other actors. There is no global policy or even global control. Each actor must assume its own responsibilities by verifying that its information exchanges are compliant with the contracts that it has negotiated with its partners.

Our approach is on the same guideline as the GDPR, which aims to strengthen the security of privacy by empowering stakeholders not only with regard to their own processing, but also by obliging them to contract their relations with their subcontractors. In our case, our concern is the control of the information imported, processed and disseminated.

In [30], [31] we presented the technical side of our approach, namely a model formalizing the semantics of service contracts using ontologies. In this work, contracts explicitly represent policies governing the relation between clients and providers regarding the use of assets. Ontologies are not used here with the purpose of creating a taxonomy, but a vocabulary for adding a clear semantics to the contractual terms that will be used for describing the policy. The model of the contract is formalized by using a subset of the first order logic, specifically, the DL formalism. The OWL 2 language is used as the concrete syntax of the model, which allows its machine-readable representation. The second stage of this work is the implementation of the proposed semantic contracts (as well as the logs that contain the evidences) within a platform to audit the interaction between clients and providers. Such a platform is able to calculate and trace indicators (e.g., controllability level) about the correct execution of contracts, relevance of contract's rules (the frequency of a rule violation gives useful insights about the relevance of the commitment). Those indicators can also be used to propose a model of trust and reputation, which, in terms of risk management is an additional security parameter to consider when choosing a service provider.

At this stage of our work, we therefore have on the one hand the controllability criterion (to assess, in a risk management approach, the ability of an organization to control its data flows), and on the other hand a formal model to represent the semantics of a service contract (allowing to express a usage control policy, the purpose of a need, the required evidences, etc.). To link these two bricks, namely to calculate indicators to assess the degree of controllability from the metadata of the information exchanged and according to the terms of the contract, one of the elements that we have already discussed is the traceability of information (Section 4.2).

7. Conclusion and future work

In the age of the Internet, information systems are now open systems supporting collaboration in multi-organizational environments. Iot, mobile Edge/Cloud computing, cyber-physical-social systems,... are all new technologies that require the creation of next generation smart and intelligent systems and applications. Beyond the technical solutions (network, systems, interoperability,...), in this article we approached the collaboration in terms of the informational quality of the manipulated data. Following an information security risk management approach we have defined a new criterion, the controllability, to quantify the level of control of an organization in the information it handles.

Although this criterion is intended to be integrated into a risk analysis alongside the traditional criteria of confidentiality, integrity and availability, we argue that it would be interesting to develop a "lite" methodology focused on this criterion alone, like what has been done for the protection of personal data under the GDPR.

Our main perspective now is to couple this risk analysis aspect with a more technical aspect: the use of contract models and security policies with a proactive view of traceability. We briefly presented the guidelines for this work at the end of this paper in the Section 6. It would also be beneficial to study how these contractual relationships could be transcribed from a legal point of view, for example in contract law.

Finally, it would be very interesting to consider how our work joins other current concerns about collaborative digital systems such as trust management, loyalty and/or transparency of algorithms, as well as notions of evidence (in the sense of provability) and decision-making responsibilities.

References

[1] D. B. Laney, "Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage", 1st ed. USA: Routledge, 2017.

[2] D. Laney, "Infonomics", *Information Governance World*, vol. 1, no. 1, p. 50, 2018.

[3] ANSSI, “EBIOS 2010 - Expression des Besoins et Identification des Objectifs de Sécurité,” *Agence Nationale de la Sécurité des Systèmes d’Information (ANSSI)*, 2010.

[4] C. Alberts and A. Dorofee, “An introduction to the octave method”, *Pittsburgh, PA: Software Engineering Institute*, Carnegie Mellon University. http://www.cert.org/octave/methodintro.html, 2001.

[5] “ISO/IEC 27005:2011: Information technology – security techniques – information security risk management”, *International Organization for Standardization*, Geneva, Switzerland, Published, 2011.

[6] “ISO/IEC 27000:2018: Information technology – security techniques – information security management systems – overview and vocabulary”, *International Organization for Standardization*, Geneva, Switzerland, Published, 2018.

[7] “NIST IR 7298 Rev. 2 : Glossary of key information security terms”, Tech. Rep., May-2013.

[8] D. M. Strong, Y. W. Lee, and R. Y. Wang, “Data quality in context”, *Commun. ACM*, vol. 40, no. 5, pp. 103–110, May 1997.

[9] R. Y. Wang and D. M. Strong, “Beyond accuracy: What data quality means to data consumers”, *Journal of Management Information Systems*, vol. 12, no. 4, pp. 5–33, 1996.

[10] K.-K. Muniswamy-Reddy and M. Seltzer, “Provenance as first class cloud data”, *SIGOPS Oper. Syst. Rev.*, vol. 43, no. 4, pp. 11–16, Jan. 2010.

[11] P. Buneman, S. Khanna, and T. Wang-Chiew, “Why and where: A characterization of data provenance”, in *Database Theory — ICDT 2001*, J. Van den Bussche and V. Vianu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 316–330.

[12] M. Herschel, R. Diestelkämper, and H. Ben Lahmar, “A survey on provenance: What for ? what form ? what from ?”, *The VLDB Journal*, vol. 26, 10 2017.

[13] P. Buneman and W.-C. Tan, “Data provenance: What next ?”, *ACM SIGMOD Record*, vol. 47, no. 3, pp. 5–16, 2 2019.

[14] A. Alkhalil and R. Ramadan, “Iot data provenance implementation challenges”, *Procedia Computer Science*, vol. 109, pp. 1134–1139, 12 2017.

[15] D. H. McKnight and N. L. Chervany, “The meanings of trust”, 1996.

[16] O. E. Dictionary, “Oxford english dictionary”, Retrieved May, vol. 30, p. 2008, 2008.

[17] A. Jøsang, R. Ismail, and C. Boyd, “A survey of trust and reputation systems for online service provision”, *Decision support systems*, vol. 43, no. 2, pp. 618–644, 2007.

[18] S. Tadelis, “Firm reputation with hidden information”, in *Assets, Beliefs, and Equilibria in Economic Dynamics*. Springer, 2004, pp. 537–553.

[19] L. Mui, M. Mohtashemi, and C. Ang, “A probabilistic rating framework for pervasive computing environments”, in *Proceedings of the MIT Student Oxygen Workshop (SOW’2001)*, 2001.

[20] L. Mui, M. Mohtashemi, C. Ang, P. Szolovits, and A. Halberstadt, “Ratings in distributed systems: A bayesian approach”, in *Proceedings of the Workshop on Information Technologies and Systems (WITS)*, 2001, pp. 1–7.

[21] P. Resnick, K. Kuwabara, R. Zeckhauser, and E. Friedman, “Reputation systems”, *Communications of the ACM*, vol. 43, no. 12, pp. 45–48, 2000.

[22] A. van Dijk, “Contracting workflows and protocol patterns”, in *Business Process Management*, W. M. P. van der Aalst and M. Weske, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 152–167.

[23] D. D. Lamanna, J. Skene, and W. Emmerich, “Slang: A language for defining service level agreements”, in *Proceedings of the The Ninth IEEE Workshop on Future Trends of Distributed Computing Systems*, ser. FTDCS ’03. Washington, DC, USA: IEEE Computer Society, 2003, pp. 100–.

[24] B. Rochwerger, D. Breitgand, E. Levy, A. Galis, K. Nagin, I. M. Llorente, R. Montero, Y. Wolfsthal, E. Elmroth, J. Caceres et al., “The reservoir model and architecture for open federated cloud computing”, *IBM Journal of Research and Development*, vol. 53, no. 4, pp. 4–1, 2009.

[25] V. Tosic, K. Patel, and B. Pagurek, “Wsol—web service offerings language”, in *International Workshop on Web Services, E-Business, and the Semantic Web*. Springer, 2002, pp. 57–67.

[26] A. Andrieux, K. Czajkowski, A. Dan, K. Keahey, H. Ludwig, T. Nakata, J. Pruyne, J. Rofrano, S. Tuecke, and M. Xu, “Web services agreement specification (ws-agreement)”, in *Open grid forum*, vol. 128, no. 1, 2007, p. 216.

[27] Liu, X., Heller, A. & Nielsen, P.S. “CITIESData: a smart city data management framework”, *Knowledge and Information Systems*, vol. 53, 699–722 (2017).

[28] V. Lalanne, M. Munier, and A. Gabillon, “Information security risk management in a world of services”, in *2013 International Conference on Social Computing (SocialCom)*, vol. 00, Sept. 2013, pp. 586–593.

[29] F. Connes, “Maîtriser ses données dans le cloud computing”, *CNIS-Mag*, nov 2012.

[30] E. Jaramillo, “A semantic contract model and knowledge-driven process for supporting controllability in service-oriented approaches”, *Ph.D. dissertation, UPPA - ED211 Sciences Exactes et leurs Applications*, 12 2016.

[31] G. E. Jaramillo, M. Munier, and P. Aniorté, “From human collaboration control towards semantic service contracts for information security”, *Ingénierie des Systèmes d’Information*, vol. 22, no. 1, pp. 43–64, 2017.