Legal Issues about Metadata: Data Privacy vs Information Security

Manuel Munier, Vincent Lalanne, Pierre-Yves Ardoy, Magali Ricarde

Legal Issues about Metadata
Data Privacy vs Information Security

Manuel Munier[1], Vincent Lalanne[1], Pierre-Yves Ardoy[2], and Magali Ricarde[3]

[1] Univ Pau & Pays Adour, LIUPPA, Mont de Marsan, France
[2] Univ Pau & Pays Adour, CRAJ, Pau, France
[3] BackPlan Company, Project Communication Control, Pau, France

Abstract. For the purposes of our work we use the concept of metadata to implement enterprise digital right management mechanisms in an intelligent document environment. Such metadata allows us to firstly define contextual security rules and secondly to ensure the information traceability. However, its use may have legal implications, especially with regard to metadata that can be stored (see personal data, privacy), how it should be stored (see probative value in case of litigation, digital forensics) or computer processing in which it may be involved. Another topical issue is the storage and the processing of data using a service provider: the cloud. We must ensure, however, that this solution does not lead to a loss of information controllability for the company. This article aims to position our work with respect to these legal issues.

Keywords: privacy; metadata; information security; socio-economic issues;

1 Introduction

Whatever business areas, new information technologies (ADSL, laptops, smartphones, tablets,...) lead us to exchange and store more and more information. Their content has also evolved. Data is more and more complex (notions of structured documents, whole archives, or even complete projects). Nowadays, public data is sometimes combined with more confidential data (notion of access restriction). Moreover, we carry our data on usb stick or in our smartphones and share it via (possibly unsecured) wireless communications like 3G, wifi or bluetooth. In the information society, the reliability of the data we handle has become a major issue in terms of security.

Security criteria most commonly used are confidentiality (assurance that information is shared only among authorized persons or organizations), integrity (assurance that the information is authentic and complete), availability (assurance that the systems responsible for delivering, storing and processing information are accessible when needed, by those who need them) and traceability (ability to chronologically interrelate uniquely identifiable entities in a way that is verifiable).

Regarding confidentiality, usage control models introduced the notion of context in order to express dynamic security rules in a policy: temporal contexts,

spatial contexts, prerequisite contexts, provisional contexts,... To enable or disable these contexts the information system must collect and store various metadata: date, IP address used, user location,... This metadata can later be used for traceability purposes, including use as evidence in case of litigation. Some of this metadata can, however, be considered as personal data of the user and thus bring privacy concerns. Metadata is one of the important keys to the success of the data warehousing and business intelligence effort since the mid nineties of the last century. Data warehouses are designed to manage and store the data whereas the business intelligence focuses on the usage of the data to facilitate reporting and analysis [1,2]. The term metadata refers to "data about data". However, the concept of metadata is not (yet) a well-known concept of the law.

The remainder of this paper is organized as follows: Section 2 presents our motivations for the use of metadata and usage control mechanisms to enforce information security; this section also presents a concrete case study in the context of business documents; we developed this example with our partner company BackPlan™ [1]; in Section 3 we present two areas of our research activities related to information security using the concept of metadata and how metadata can be useful for usage control, traceability and information system monitoring; Section 4, using examples of jurisprudence, highlights the need to formalize the metadata necessary to enforce the security policy and the framework in which they can be used order to comply the regulations; Section 5 presents some socio-economic issues underlying the storage of data (and metadata) in the today information society; Section 6 concludes the paper and presents some of our perspectives.

2 Motivations

As we stated at the beginning of this article, information technology (IT) allows us to share more and more information in the form of documents whose structure becomes more complex. Whether in the context of a "simple" information dissemination (unidirectional communication from a content provider to the users) or a collaborative work environment (several actors interact to complete tasks with a common goal but possibly different objectives), it is therefore necessary to implement security mechanisms that go beyond a simple access control: usage control (how partners can use a document: obligations, workflows, delegation rules,...), information consistency management (e.g. some documents may reference others), traceability (monitoring of actions, metadata attached to information),...

We do not give here all the details of our case study. This section focuses on the use of metadata to improve information security (traceability, usage control) and legal issues that may arise.

[1] BackPlan™ , Project Communication Control
http://www.backplan.fr

2.1 Sample Application

Consider an Oil & Gas project as the construction of a pipeline or an oil installation. The information system consists of numerous documents, it has a central role, its structure and development evolve along with the progress of the project: the documentation must always precede action (design, work procedures). Documentation is a requirement at the closure of the project along with verifications (records, the minutes,...). The document evolves at the same rate as the project. These documents are specifications, drawings, records of expertise, procedures, records,...

Business Aspects Such a project obviously involves many partners and subcontractors. Here is a representative example of a project timeline (Fig. 1):

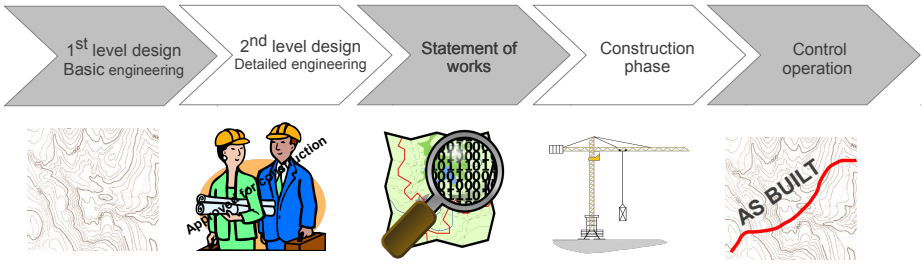

Fig. 1. Oil & Gas project timeline

- 1st level design: Basic Engineering. This step is performed by the land surveyor who makes a topographical survey of the site where the work will be done.
- 2nd level design: Detailed Engineering. This is the design phase of the project itself with the aim, in particular, to minimize the environmental and human constraints; it is performed by an engineering company that will plan the work and the construction will be launched from this plan. This level involves various partners (civil engineering, pipefitters, instrumentation engineering, utilities,...) who share many documents.
- Statement of works: numerous buried and air, public or private networks, go through the territory (water, electricity, gases, dangerous products, telecommunications, irrigation,...). Further to accidents, it is imperative to localize very exactly their position. So, during the realization of a project of construction, gas pipeline for example, companies have the obligation to question a centralized information system common to all the French territory.
- Construction phase: it is based on engineering documents and work procedures. It comes with many documents that are intended to demonstrate

compliance of the book in terms of quality and regulatory standards (e.g. multifluid standard, water code, capacity under pressure). As-builts will have to justify the differences for the administration.
– At the end of construction the land surveyor will come again, and verify the topographic survey: this is a control operation of a project to verify the differences from the planned location and update the data (to know where everything is). The engineering documents then pass status "As built". This operation can also update the geographic information system (GIS) of the place. As-builts must be attached to requests for authorization to operate sent to administrations (also signed by the legal representative).

As we have stated, the partners will handle many documents. Because of the nature of this type of project, a multitude of corporate associations has to work on the same documents. Where from requires it to manage the communications between these interfaces. Besides, it will be necessary to be able to guarantee that each works with documents "up to date" or that the last modifications in date were well taken into account before the "publication" of certain results (cf. usage control and collaborative work management). There can also occur unforeseen circumstances during the project.

Information Security Aspects We propose to improve the security of information in two directions: metadata management and usage control.

Metadata It could be used to "bind" reviews, certifications, good practice guides, standards, and other minutes to design documents and reports within the information system. The aim is to improve traceability, both in the design process (concept of workflow) in case of litigation (concept of proof of conformity, digital forensics). Take for example a phase control such as checking the welding of a pipe[2]; it would be interesting to use metadata to improve the traceability of the process for purposes of validation and/or evidence: photos geotagged (to certify checkpoints), metadata associated with the plans,... Since several partners are working on such a project, everyone could also attach some metadata to the information: confidence and trustworthiness indicators, impact risk of a change,... This metadata would permit to calculate various performance indicators for monitoring the partners' tasks or metadata to the information they produce.

In case of transfer of work this metadata would allow to improve the follow-up of the project towards the buyer: operations history, context decisions (standards, studies), how and why they have been taken, etc...

Usage Control Here are some examples of security rules that we would implement to control how partners use documents:

[2] The sections of pipe are welded every 12 to 15m. These welds should be checked: radiography, analysis by a certified individual, hydraulic tests. These controls are spread over time and generate many records that are, once made, a legal value.

- It is possible to write a deliverable of the project only if confidence in the various technical documents exceeds a certain threshold. It is a dynamic access control based on trust (whether in a document or a partner).
- The security rules may prohibit access to parts of the document based on location data. This prevents, for example, on a site (or train) an unknown person takes sensitive information over the shoulder of someone.
- A responsive access control may require a partner (or subcontractor), via a mechanism of pre-obligation to accept the terms of a contract (non-disclosure agreement, delegation of responsibility, deadlines,...) before accessing a plan and contribute to the design.
- A user control would be to require a partner to complete parts of design documents (e.g. inform the radii of curvature of the pipe, write the study of soil before drilling) before a deadline if he wants to stay a project member (notions of punishment and penalty).
- The usage control can also define collective obligations. For example, each partner must have reread each concept study in which they participate at least 7 days before the deadline for validation.

Legal Issues These are just some examples of opportunities. But the use of metadata is not easy to understand from a legal point of view ! First, regarding the collection and storage, some of this metadata can be in the domain of personal data (including geolocation). But their use to enable/disable contextual security policies (permissions, obligations,...) or calculate some indicators (e.g. trust in a partner, document quality) are automatically processes and are therefore subject to a number of regulatory frameworks. Add to that the concepts of accountability and sanctions mentioned above and it is obvious that metadata has now become essential elements of information systems. They should be considered as full data and be secured along with "classic" information.

Section 4 gives more details on requirements to formalize metadata necessary to the security policy and how metadata can be used. One of the objectives is that this metadata can be used as evidence in case of litigation (cf. probative value) while respecting the laws on privacy.

2.2 BackPlan™

BackPlan™ is a French company providing document management services and collaboration workflow applications to improve project communication, transparency across the project, ability to manage schedule and risks, reliable indicators and regulatory compliance. From the engineering phases to the construction phases, projects involve different companies. All of them will use the collaboration solution BackPlan™ to ensure consistency of information across the project and a complete audit trail of project communication. BackPlan™ document management services are currently provided on a server hosted in a data center: the document registry.

Using metadata (as data about data) would allow BackPlan™ to enhance existing services and to offer new services to their customers. During the course of the project, metadata would be used to calculate many indicators for project progress, compliance with deadlines, completion status of the various documents,... Once the project is completed, the company makes and delivers to its customers the case-file containing all the business documents for project traceability. In terms of risk management, in case of litigation this information can be used to identify those responsible for error or prove that during the construction phase of the project the standards in force at that time have been complied with.

3 Metadata & Information Security

Our research activities take place in the field of information system security. Issues related to the legal framework for the use of metadata were raised when we wanted to implement contextual security rules to perform usage control within two areas of research: the development of a secure autonomous document architecture, and the study of service oriented architecture security. This metadata will obviously be used during the life cycle of the document to perform usage control. But they can also be accessed later for traceability of actions performed and thus serve as evidence in case of litigation or for digital forensics.

This section focuses on the use of metadata to improve information security. In our work this "data about data" allows us to implement security mechanisms such as dynamic security policies, collaborative work management, calculation of confidence and trustworthiness indicators,... The use of metadata is not, however, trivial in terms of the law. As discussed in Section 4, these technological possibilities must not make us forget the legal issues. In addition, we will see in Section 5 that this metadata can be more important thant the data to which it is associated.

3.1 Intelligent Document Architecture

As part of our research we developed a multi-view model for secure data warehouse [3] and we proposed E-DRM[3] architecture based on secure autonomous documents [4]. While "traditional" information systems centralize all the data on a server which users must connect, we have chosen to define an approach where security mechanisms are relocated closer to the user. However, unlike "conventional" DRM architectures that require the use of a dedicated player (which is responsible for enforcing the security policy), we decided to embed these security mechanisms within the document following object-oriented approach: a document is an autonomous entity capable of ensuring by itself the security of the information it contains and controlling how this information is used. Such a document is a kind of information system on its own embedding both a data

[3] E-DRM: Enterprise Digital Right Management

warehouse and various security modules (access control, usage control, metadata,...). Users can thus exchange the document directly and safely without having to connect to a central site.

The core of our architecture, namely the security kernel, relies on the OrBAC model [5] to express security policies in terms of permissions, prohibitions and obligations between a subject, an object and an action. These security rules are dynamic, that is to say, they can be "adapted" to the context of the actions [6,7,8]: activation or deactivation of rules, the execution of an action triggers the insertion of an obligation,... OrBAC model supports various kinds of contexts: *temporal context* (depends on the time at which the subject is requesting for an access to the system), *spatial context* (depends on the subject (geo)location), *user-declared context* (depends on the subject objective or purpose), *prerequisite context* (depends on characteristics that join the subject, the action and the object) and *provisional context* (depends on previous actions the subject has performed in the system). The embedded information system must therefore provide the information required to check that conditions associated with the context definition are satisfied or not. To do this, either it has direct access to the host system (e.g. a global clock to check the temporal context) or it uses metadata carried by the actions and the nodes contained in the embedded database.

Our approach therefore relies heavily on the concept of metadata, both for collection (traceability) and for context activation (dynamic security rules) or, eventually, computing various indicators throughout the uses (confidence, trustworthiness) as works published in [9,10,11,12].

3.2 Service Oriented Architecture Security

Service Oriented Architectures (SOA) offer new opportunities for the interconnection of systems. Opening its Information System to the "world" is not insignificant in terms of security. Whether to use available services or provide its own services, new technologies have introduced new vulnerabilities and therefore new risks. Our work aims to propose an approach for risk management which is based on the ISO/IEC 27005 standard: we propose a development of this standard so that it can fully take into account the type "service" as web services and cloud services [13].

To develop a security model for communications in inter-organisational information systems we also use a usage control oriented approach as presented above. Again, the use of metadata for traceability of communications (via these services) allow us to compute indicators that will be used to monitor the information system [14]. Our goal is that companies can keep control over their information, despite the use of cloud technologies.

3.3 Data Privacy Concerns

Preliminary tests on our self-protecting document architecture, performed with predefined metadata, allowed us to implement dynamic security rules for usage

control. These experiments have also led us to concern ourselves with the privacy of metadata.

According to the ISO 8402 standard, traceability is the ability to trace the history, use or location of an entity by means of recorded identifications. In our architecture, an entity corresponds to an autonomous collaborative work document. History, identification, registration and use are relevant concepts. As we mentioned above, user localization can be an interesting metadata for expressing contextual security rules. But a document could then reveal the presence of a user at a given time in a certain place, or the various revisions which followed one another leading to the final of a contractual document ! So we must not only focus on the metadata that is collected, but also protect them from unauthorized use and control how it can be used (e.g. automatic computation of indicators). Basically, the problem of leakage and/or misuse of information is already known. However, in our approach, the "massive" collection of metadata "of any kind" can effectively exacerbate the problem.

A concrete example is presented in the article [15] entitled "Metadata: the ghosts haunting e-documents". This story demonstrates the risks of exchanging files with embedded data in negotiating a contract. During negotiations, partners used a common word processing program, Microsoft Word, to edit and propose revisions to the contract, and they utilized the program's track changes feature to allow the other side to see the specific changes proposed. They e-mailed the electronic draft, complete with embedded data, back and forth to each other between rounds of revisions. But without using anything but Microsoft Word's inherent functions, someone can reveal hidden internal comments from adverse party concerning terms of the contract, negotiating positions,... In doing so, a partner can be aware of confidential business information and use it to to pressure his opponent.

This article also raises an interesting question: is that partner (in this case a lawyer) bound by the same obligations that apply when documents in a misaddressed envelope are received or, conversely, is the partner free to use and review the embedded information ?

4 Metadata & Legal Issues

As it has been said from the beginning of this article, our works rely highly on the concept of metadata, both in the collection (traceability) for activation contexts (dynamic safety) and, in the course of time, the computation of various indicators over time (confidence rating or trustworthiness value). Note that, for the moment, the elements of study in this section relate only to the French law. For the scaling of our work, the specificities of each country should obviously be considered.

The concept of metadata is not a well-known concept of the law. Composed of the Greek prefix *meta-* referring to the reference to itself, the term metadata refers to data within data, data which defines limits or describes other data. That data is varied and can include durations, dates, places, elements to

identify people,... In a very precise domain, however law gives a definition to metadata. Article L.127-1, 6° of the Environmental Code actually specifies that it is the *"information describing sets and spatial data services, making possible their discovery, their inventory and their use"*. However, this definition, the first one put by a legal text, does not report the diversity which hides under the term of metadata.

Metadata raises three types of difficulties: its collection, its storage and its use. First of all, problems raise with regards to its collection. Very often, its collection is done unbeknown to the authors, at the least in the ignorance of the concerned entities. Concerning the law, this raises the question of the right of access to information contained in metadata and above all the question of the right to know the information is collected. Another difficulty arises straigh away: the entity who collects the metadata is not even aware he is making such a collection, it is only, subsequent to the collection, that the meditative elements are discovered and used.

Secondly, metadata raises the question of its storage. This storage implies guaranteeing conservation, not only of the authenticity of the metadata (which refers to the question of the collection and the reliability of the source), but also of its integrity, its stability. What is at the heart of the matter is ultimately to require the reliability of the metadata. It is not only important that it is stored in good conditions, it is also necessary as to ensure its availability, its accessibility, what is needed to be know, by whom, for how long and under which conditions.

Finally, metadata raises the problem of its usage. Such a usage may be done in good faith, as for example in the implementation of security mechanisms, but it can also in bad faith, the metadata being diverted from its original usage or be subject to falsification.

Potentially metadata affects all fields of the law and it is difficult or even useless to apprehend it in the abstract. It consists in identifying areas where the use of metadata is likely to raise specific questions without claiming completeness. Also note that this is a primary questioning limited to the French domestic law.

4.1 Evidence of Law

It is mainly in the field of the law of evidence that the rare decisions of the court of appeal which refers to metadata are found.

In civil law, proof of a legal act can be given by any means, nevertheless within certain conditions. It is in particular required that the evidence be reported fairly and that the type of evidence is reliable. Even if the rules are significantly different from criminal law, the reliability requirement is definetely common to both domains. This reliability requirement is obviously at the heart of the admissibility of an electronic document, proofwise. Article 1316-1 of the Civil Code thus provides that, for proof of legal acts *"the writing in electronic form is admissible in evidence equal to a written document on paper, provided that the person who issued the document and its establiment and storage are executed under the conditions so that its integrity can be duly identified"*. It is thus not a surprise that the rare decisions referring to metadata require its reliability.

This is foremost the case of the Ruling of the First President of the Paris Court of Appeal on October 25, 2011, Juris-Data N°2011-025553 and October 25, 2011 (unpublished, N°09/14462, 09/14501)[4]. In this specific case, a company filed a legal complaint against the search and seising in its premises authorized by local competent authorities with regards to the investigation of anticompetitive practices. Even if the legal problem did not directly address the issue of metadata, the Court nevertheless adopted the argument of the Competent Authority at the end of which *"the structure of a particular Outlook mail file and the obligation not to change the state of the computer visited nor the characteristics of a file (metadata in the file itself: title, author size, dates, location, signature,...) necessarily imply the complete seising of the mail file after verifying that it contains elements falling within the scope of the authorization"*. In all cases the emphasis is on the requirement that it should seize the entire Outlook messaging so as not to affect the reliability of the input by an alteration of metadata contained in the messages.

In a judgment of the Court of Appeal of Versailles, Ch.1, sect. 1 of September 30, 2010 (unpublished, N°09/03831), an applicant pressed on the metadata contained in photographs to prove his position as author of the photographs. The Court retained the metadata elements as evidence noting that these included the identity of the author, the date and time of shooting, the name of the manufacturer of the device, the model of the latter and the description of the camera settings. If metadata in itself may not demonstrate the source of the photographs, it is however the elements which can establish the paternity and can therefore be very useful in the case of a counterfeit lawsuit. Within the same context, the Labour Chamber of the Court of Appeal of Rennes retained, in a judgment of September 20, 2011 (unpublished, RG N°10/05183), that the production of metadata may be used to determine the creation date of an advertising brochure while noting that the reliability of the latter is not discussed.

It is of course conceivable that metadata can serve as evidence in a variety of areas. It can thus be used to prove the absence of an employee. In telework, metadata opens up the possibility to determine whether the employee has met with his working hours. Thus, arise questions of the distinction between personal time and working time, respect of privacy and where the bounderies of the employer's power are situated.

Beyond relations between employer and employees, metadata may be of interest in relationships between business partners or directors of a company. It could for example be used to demonstrate that this officer or that employee was responsible for the disclosure of a trade secret (Article L.621-1 of the Intellectual Property Code and L.1227-1 of the Labour Code).

Moreover, in the context of the mission of prevention and repression measures of illegal downloading which lie within the Hadopi law, metadata can be useful to show which person is actually responsible for downloading illegal copyright

[4] See also Orders of November 15, 2011 or August 31, 2012 from the First President of the Paris Court given in matters relating to anticompetitive practices.

protected work. In July 2013, French legislators struck down the heavy-handed Hadopi online copyright law. Under the law's "three strikes" rule, users who violated copyright restrictions three or more times could be punished by having their Internet connections cut. But Hadopi suffered great controversy when France's highest court, the Constitutional Council, declared access to the internet a basic human right. French legislators are now seeking policy reforms that will shift the focus of law enforcement towards commercial piracy issues.

These are obviously only a few possible usages of metadata, an exhaustive list does not seem possible or at least very delicate.

4.2 Privacy and Individual Liberties

Indeed, metadata may contain information about people, whether it concerns identification or location. For instance, a picture of a group of friends can be published on the Internet (we recall that The Internet is a public space) and may contain the names of the people in the picture, the place where it was taken, the date and time of the snapshot (tags on Facebook photos).

So these are issues that concern not only image rights (assuming the image was published without the agreement of different people), but also the respect of someone's privacy set by Article 9 of the Civil Code. For a number of reasons, in fact, a person may want a certain amount of information not to be disclosed. (such date, such time,...).

If the concept of metadata is just beginning to be understood by the law, it has on the other hand understood those of personal data. It concerns information that can directly or indirectly identify individuals (*Information Act N°78-17 of January 6, 1978, Directive 95/46/EC of October 23, 1995*). The concept of metadata does necessarely correspond to the concept of personal data insofar metadata may contain information which is not necessarily information which can be used to identify a person (information about location, time spent on a file,...). They may, however, raise questions which may be of interest to privacy.

As such, metadata collection is likely to be considered as a collection of personal data. In this sense it seems possible to read Deliberation of the CNIL N°2011-423 of December 15, 2011 authorizing the company GEOLSEMANTICS to implement, on a trial basis, as part of a research project, the treatment of personal data, necessary for the development of a tool, called SAIMSI (eng: follow adaptive inter-lingual and multi-source information). The metadata items concerned are those attached to the collected documents, that is to say, those *"corresponding of how the information was collected (if applicable: document URL source, date of registration, date, time and place of the issue body text and the source)"*.

4.3 Digital Protection and Intellectual Creations

The French DADVSI Act of August 1, 2006 introduced the ability to protect intellectual works by systems which limit or prohibit any copying. In French law, these devices are called Technical Protection Measures (MTP), better known by

the acronym DRM (Digital Right Management). The Intellectual Property Act has developed a complete,relatively complex system, designed to ensure that such DRM can not be used by producers or publishers aimed at anticompetitive purposes unrelated to the protection of copyrights. The presence of DRM should thus not prevent interoperability, that is to say the ability for the works to be read by the most diverse materials.

5 Metadata & Socio-Economic Issues

In the information society today, metadata becomes sometimes more important than the data which it is associated. Whether in the field of privacy (personal data) or professional (business data of a company), many companies have developed their business on it. Lately, the media focus on large multinational companies such as Google, Facebook or Microsoft regarding the protection of privacy and personal data. This is a hot topic that scares the public. In the professional world, same issues arise about critical data of companies (e.g. research & development, business strategy). This is the case for example of the BackPlan™ company whose business uses metadata on information exchanged between participants for project communication control.

The objective of this section is not to criticize the practices of a particular country or to denigrate the work of a particular company. We just want to highlight the socio-economic issues about information in today's society and the need to harmonize the laws of different countries to define an international legal framework.

5.1 Data is Future's Power

We live in a transitional period, the digitization of everything: people, society, organizations, knowledge, interactions,... Data is the basic building block of the information society. Its quantity is growing exponentially: we are talking about Big Data. The physical infrastructure of the information society, telecommunication systems, storage facilities and data processing, new online services, are industries experiencing unprecedented growth. Data per se offers tremendous potential that we begin to use to generate new knowledge.

Personal data, both that produced by the users (texts, photos, videos,...) and that generated by the systems we use often unknowingly, is the heart of the economy of the information society, and therefore the heart of the economy. Control of the data also allows control of certain markets, which currently are already using U.S. electronic commerce tools in some areas. Control of the information society gives power still difficult to evaluate and far beyond the areas of the market economy.

Data capture is the top priority in some countries such as the United States or China (which hold respectively 72% and 16% of the top 50 sites worldwide). In both countries, national data remain under control of the domestic industry. And both aspire to collect the data at the international level.

5.2 Data Location

Geographical location of the cloud provider can have a real impact on the protection and confidentiality of data.

Legal obligations Sensitive data can be stored using a cloud computing solution. But for a French company, for example, it is necessary to check that the provider undertakes to keep these documents in France. Otherwise, the company may be unable to ensure that the processing of personal data complies with the legislation in force for it (e.g. in France: duration of data retention, ability to modify and delete information,...).

Similarly, it is generally necessary to comply with certain legal tax obligations: prohibition to store account books outside the European Union, mandatory reporting to the tax authorities in order to store electronic invoices outside the national territory,...

The "USA PATRIOT Act" Dutch legal researchers have published a study [16] that highlights the importance for a European company to choose a European provider to outsource the processing of personal data or information vital to the company. Indeed, since the establishment of the USA PATRIOT Act, U.S. law allows security services to access all personal data [17]:

- data from U.S. companies, even if the data is physically stored on the European territory
- data from their subsidiaries, even if they are located in another country in the world
- data stored on servers that are hosted in the United States, even if the company that owns the servers is of another nationality

The U.S. government has now established a legal arsenal which allows personal data control of foreign citizens, including Europeans, by leveraging its major companies such as Facebook, Google or Microsoft. At the end of 2012, the European Parliament's Committee on Civil Liberties, Justice and Home Affairs ("LIBE") released a study titled "Fighting cyber crime and protecting privacy in the cloud" [18]. Authors denounce the "Foreign Intelligence and Surveillance Act" (FISA). This amendment expressly authorizes U.S. intelligence agencies (NSA, CIA,...) to wiretap (without judicial authorization) U.S. citizens communicating with foreigners suspected of terrorism or spying. Shortly, a secret tribunal is now able to issue a warrant, secret too (the "secret" for actions may be required for an indefinite period), forcing American companies to deliver to U.S. intelligence agencies the private data of foreign users. Therefore your information may be duplicated, stored and disclosed to third parties without notifying you... In December 2012 the amendment was extended until 2017.

Our aim is not to pass judgment on the "USA PATRIOT Act" and other secret projects such as Riot or PRISM (since 2007 and revealed in june 2013 by Edward Snowden). We simply point out that in the current legislation, a European company with strong constraints on the information confidentiality

must therefore be vigilant when choosing a service provider (data location and nationality of the provider).

5.3 Towards a European CNIL

The European Parliament has made good progress on the reform of the EU legislation on data protection proposed almost a year ago by the Commission [19]. The United States, which are reforming their own legislation, call for transatlantic regulatory convergence, noting that they are just as demanding as the Europeans in this area. The EU has the ambition to become the global standard for data protection, suggesting (according to the authors) that the United States are more lax.

One of the most controversial issues is the requirement of equivalent standards to allow European data transfer to a third country for processing. The problem also arises for global corporations (e.g. Google), whose processing practices should be approved by the Union, while the United States would continue to use their codes of conduct.

In January 2013 the European Parliament presented its preliminary report on the future reform of the EU Directive on the protection of personal data in response to proposals from the European Commission. Wishing to strengthen the protection of data of its citizens, Europe is about to start revising the measures that came into force in 1995 ensuring wanting to replace Directive 95/46/EC on the protection of data by a European regulation that all Member States should apply without discussion. This reform will require the creation of an independent administrative authority, ie a European CNIL, which will enforce the rules on data protection, which could take the form of an independent agency.

However, for the French Data Protection Authority ("CNIL"), the text proposed by the European Justice Commissioner Viviane Reding "presents considerable progress" but also "elements of concern". The President of the CNIL, Isabelle Falque-Pierrotin acknowledges that it has the "major advantage" to submit to the European law all data processing on a European resident by a company not established in Europe: in other words it is the European law that would apply to a French victim of abuse by an American internet company, for example. But, says the President of the CNIL, the European text raises the problem of the concept of "principal place of business", according to which the competent regulatory authority in the event of a dispute with a European citizen is that of the place where the company and not the complainant.

5.4 Synthesis

The political and economic authorities have become aware of the need to establish an international legal framework to control the collection, storage and use of data. Metadata associated with data is also included. It is indeed value of the highest importance for companies whose business is the management of information.

6 Conclusion

Confidence in the data that we handle every day is one of the major challenges of the information society. There are many mechanisms that allow us to collect, store and process huge amounts of data and especially data on this data: metadata. Metadata is an essential tool for information security: usage control for document sharing and cloud security, digital forensics, evidence in case of litigation,...

Technological possibilities must not however make us forget the legal issues. The objective being to implement a security policy and to ensure information traceability, it is essential to respect existing regulations regarding the metadata that can be stored (see personal data, privacy), how it should be stored (see probative value) and computer processing in which it may be involved.

Through this article we want to raise awareness of potential abuses related to the use of such metadata. Some work has already been done to preserve privacy. An example is the anonymization of data [20,21]. These are not always suitable for our problem of usage control where precisely some indicators should not be anonymous. In the context of E-DRM we talk about business projects between partners. Thus our approach is rather to formalize this "collaboration agreement". For the IT community, this will be in terms of language as specification for metadata to be collected, by what means, how it is stored and what will be the use. For the legal community, it must first qualify the metadata: should it be treated as "traditional" data or should it receive a specific legal regime ? Once defined the legal framework, we can study together under what conditions it is possible to use metadata and, in the other way, what are the metadata necessary to apply certain laws. For instance, in the Oil & Gas case study described in Section 2.1, it will now be necessary to include in the contract between companies ("collaboration agreements") the insertion or the deletion of this metadata. For instance, should metadata appear within documents delivered at the end of the project ?

Finally, in Section 5 we discussed some socio-economic issues underlying the mass storage of data (and metadata) in the today information society. Beyond collaboration between partners on a project, we must also study the use of service providers on the "cloud" (storage or processes). These technologies have become unavoidable for companies although they introduce new vulnerabilities for the information security (loss of information controllability). These threats are not just technical (hardware, software, network). They can also be political, which requires the definition of an international legal framework for data protection.

References

1. Inmon, W.H.: Tech topic: What is a data warehouse ? Prism Solutions **1** (1995)
2. Kimball, R., Ross, M., Thornthwaite, W., Mundy, J., Becker, B.: The Data Warehouse Lifecycle Toolkit. 2nd edn. Wiley Publishing (2008)
3. Munier, M.: A multi-view approach for embedded information system security. In: CRiSIS, IEEE (2010) 65–72

4. Munier, M., Lalanne, V., Ricarde, M.: Self-protecting documents for cloud storage security. In: TrustCom, IEEE (2012) 1231–1238
5. Kalam, A.A.E., Benferhat, S., Miège, A., Baida, R.E., Cuppens, F., Saurel, C., Balbiani, P., Deswarte, Y., Trouessin, G.: Organization based access control. In: POLICY, IEEE Computer Society (2003) 120–131
6. Elrakaiby, Y., Cuppens, F., Cuppens-Boulahia, N.: From contextual permission to dynamic pre-obligation: An integrated approach. In: ARES, IEEE Computer Society (2010) 70–78
7. Cuppens, F., Cuppens-Boulahia, N.: Modeling contextual security policies. Int. J. Inf. Sec. **7**(4) (2008) 285–305
8. Cuppens, F., Miège, A.: Modelling contexts in the or-bac model. In: ACSAC, IEEE Computer Society (2003) 416–427
9. Bertino, E., Lim, H.S.: Assuring data trustworthiness - concepts and research challenges. In Jonker, W., Petkovic, M., eds.: Secure Data Management. Volume 6358 of Lecture Notes in Computer Science., Springer (2010) 1–12
10. Zheng, X., Maillé, P., Le, C.T.P., Morucci, S.: Improving the efficiency of collaborative work with trust management. In Agoulmine, N., Bartolini, C., Pfeifer, T., O'Sullivan, D., eds.: Integrated Network Management, IEEE (2011) 1172–1179
11. Xingyu Zheng and Patrick Maillé and Cam Tu Phan Le and Stephane Morucci: Trust mechanisms for efficiency improvement in collaborative working environments. In: MASCOTS, IEEE (2010) 465–467
12. Le, C.T.P., Cuppens, F., Cuppens-Boulahia, N., Maillé, P.: Evaluating the trustworthiness of contributors in a collaborative environment. In Bertino, E., Joshi, J.B.D., eds.: CollaborateCom. Volume 10 of Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering., Springer (2008) 451–460
13. Lalanne, V., Munier, M., Gabillon, A.: Information security risk management in a world of services. In: PASSAT. (2013)
14. Jaramillo, E., Munier, M., Aniorté, P.: Information security in business intelligence based on cloud: A survey of key issues and the premises of a proposal. In: WOSIS. (2013)
15. Hricik, D., Scott, C.E.: Metadata: The ghosts haunting e-documents. In: FindLaw. (March, 2008)
16. Van Hoboken, J., Arnbak, A., Van Eijk, N.: Cloud computing in higher education and research institutions and the USA PATRIOT Act. Social Science Research Network Working Paper Series (November 2012)
17. Lee, L.T.: USA PATRIOT Act and telecommunications: Privacy under attack. Rutgers Computer & Tech. LJ **29** (2003) 371
18. EU: Fighting cyber crime and protecting privacy in the cloud. EU Parliament (2012)
19. EU: Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protecting Regulation). Comm. European Communities, Bruxelles (2012)
20. Guarda, P., Zannone, N.: Towards the development of privacy-aware systems. Inf. Softw. Technol. **51**(2) (February 2009) 337–350
21. Zhou, B., Pei, J., Luk, W.: A brief survey on anonymization techniques for privacy preserving publishing of social network data. SIGKDD Explor. Newsl. **10**(2) (December 2008) 12–22