ILNAS

uni.lu
UNIVERSITÉ DU LUXEMBOURG

White Paper

DATA PROTECTION AND PRIVACY IN SMART ICT

SCIENTIFIC RESEARCH AND TECHNICAL STANDARDIZATION

Version 1.0 · October 2018

ANEC

SNT
securityandtrust.lu

Avec le support de :

LE GOUVERNEMENT
DU GRAND DUCHE DE LUXEMBOURG
Ministère de l'Économie

White Paper

DATA PROTECTION AND PRIVACY IN SMART ICT

SCIENTIFIC RESEARCH AND TECHNICAL STANDARDIZATION

Version 1.0 · October 2018

ILNAS
Institut Luxembourgeois de la Normalisation, de l'Accréditation, de la Sécurité et qualité des produits et services

ANEC
Agence pour la Normalisation et l'Economie de la Connaissance

uni.lu | SnT
securityandtrust.lu

Avec le support de :

LE GOUVERNEMENT
DU GRAND DUCHE DE LUXEMBOURG
Ministère de l'Économie

Today, the Smart ICT domain is becoming ubiquitous, making an impact across economic sectors and everyday lives. Moreover, it is contributing towards the competitiveness of societies globally. The wide spread adoption of technologies such as Cloud computing, Big data and Internet of Things have resulted in a situation where the amount and variety of data that are being generated and processed are higher than ever before.

Luxembourg has already high-quality Internet services and is actively embracing this digital revolution. In order to realize our vision of a data-driven economy, data must be treated and protected like a high value asset. Thus, Luxembourg is setting-up frameworks and platforms to provide a secure environment for handling data, spanning the entire lifecycle from data generation, collection, storage, processing, analysis and disposal. For instance, the new version of the national cybersecurity strategy (2018-2020) defines objectives and implementation guidelines for strengthening public confidence in digital environments, infrastructure protection, and promotion of the economy. These initiatives will not only make Luxembourg a trusted place for businesses and citizens but also be in line with the European regulatory framework on cybersecurity, data protection and privacy.

In this context, ILNAS, the national standards body and the Interdisciplinary Centre for Security, Reliability and Trust (SnT) of the University of Luxembourg have established a partnership in order to converge their expertise in technical standardization and research respectively for building secure, reliable and trustworthy Smart ICT systems and services. On the one hand, the SnT conducts internationally competitive research that has high impact. Its research expertise include in particular topics such as security, data management, satellite systems and Cloud computing. On the other hand, ILNAS is in charge of implementing the national technical standardization strategy and leads a strong policy concerning Smart and Secure ICT domains. Considering the importance of the protection of data in the digital world, ILNAS and the SnT have developed this white paper “Data Protection and Privacy in Smart ICT” as a first outcome of their partnership.

First, this white paper clarifies the fundamental concepts as well as data protection and privacy challenges in Smart ICT. Secondly, it provides a model describing how data serves as a common thread among Smart ICT topics and the relevant state-of-the-art from two perspectives: scientific developments and technical standardization. Finally, it concludes by highlighting the common points between scientific developments and technical standardization.

Luxembourg considers technical standardization and research particularly concerning security, privacy and data protection in Smart ICT as a force multiplier for the economy and for its competitiveness. In this sense, with associated research and education initiatives, this white paper represents an example of a project that ILNAS and the SnT are carrying out on a common basis to develop the necessary related culture about ICT technical standardization within the Smart Secure ICT framework at the national level.

Etienne Schneider

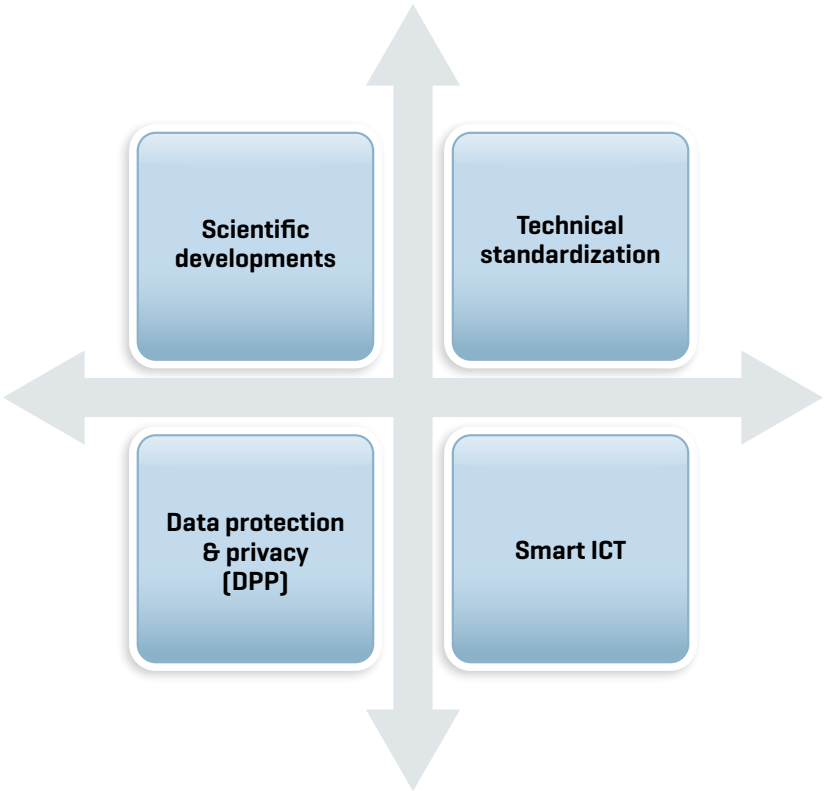
Deputy Prime Minister
Minister of the Economy

Table of contents

List of Figures

List of Tables

1 Introduction



1. Introduction*

Today, modern computing techniques are capable of easily storing, processing and analyzing large amounts of and a variety of data. These techniques could correlate diverse datasets in order to create individual's profiles, to gain insights, and to offer new digital services. Among others, three technologies that could play a major role in gaining these capabilities are Cloud computing, Internet of Things (IoT) and Big data. Although developed independently, the integration of these three technologies (referred to as **Smart ICT)** has accelerated the growth of data-driven applications and has unleashed numerous opportunities for businesses, individuals and the society at large [1].

At the same time, numerous organizations are collecting information about individuals, and people are disclosing personal information (either voluntarily or being unaware) to a multitude of institutions more than ever [2]. For example, individuals are disclosing their geographical locations, life-events and pictures of themselves, friends and family on social media, information such as tax returns (e.g., via online forms) to authorities, policy details and claims to health insurance companies, and so on. This implies that personal information about individuals is widely spread and many third parties are involved in handling (e.g., collecting and processing) this information.

The data collected and queried by third parties (e.g., Cloud computing service providers) hence pose a constant risk for the individuals who can be identifiable. Personal **data thus needs to be protected** carefully and processing of personal data must, among others, ensure legitimacy (have justifiable reasons to process personal data), purpose (personal data must be used only for a given purpose), fairness (treatment of data must be clearly communicated to the data owner) as well as security and **privacy** [2]. The latter is important since there is often a loss of control over personal data in the way it is treated, shared, and used by third parties, breaching individual's privacy [3].

The goal of this white paper is to provide a holistic view of **privacy and data protection in Smart ICT**. To this aim, a review of the state-of-the-art highlighting existing challenges and proposed solutions is presented from two different viewpoints: **scientific developments and technical standardization**, so that the readers of this white paper could broadly answer to questions such as:

- Is data the common thread in Smart ICT? If so, what is the data model?
- What is the scientific state-of-the-art concerning privacy and data protection in Smart ICT?
- What are the recent developments in technical standardization related to privacy and data protection?

In the remainder of this chapter: Section 1.1 will introduce the notion of Smart ICT by providing the definitions and characteristics of Cloud computing, Internet of Things and Big data, and by highlighting how ubiquitous applications combine Smart ICT domains as well as interaction between them. Thereafter, Section 1.2 will introduce the aspects of privacy and data protection in Smart ICT, while Section 1.3 will explain the need and role of technical standardization in this context. Finally, Section 1.4 will summarize the organization of the white paper for the sake of readability.

---

* This white paper is a joint work between ILNAS and the SnT of the University of Luxembourg, with the support of the Ministry of the Economy, as part of the research program (https://smartict.gforge.uni.lu/) initiated in 2017. The **lead authors** of this white paper are the **PhD students** who are working within the research program. Each student is focusing on a specific Smart ICT topic: Ms. Saharnaz Dilmaghani (Big data), Mr. Chao Liu (Cloud computing) and Mr. Nader Samir Labib (Internet of Things).

1.1 Introduction to Smart ICT

Recent years have witnessed major innovations in the Information and Communication Technologies (ICT) in the form of Cloud computing, IoT and Big data. These technologies have changed the way of computational resources are utilized by individuals, the role that data plays in ICT applications and services, and the scope of technologies in our lives. The interaction between these three technologies (Smart ICT) is also creating several new opportunities. For better understanding, this section briefly presents Cloud computing, IoT and Big data paradigms, and introduces the notion of Smart ICT.

1.1.1 Cloud computing

According to ISO/IEC 17788:2014 [4], the Cloud computing paradigm enables ubiquitous, on-demand network access to a shared pool of configurable computing resources (e.g., storage, processing, network, applications and services), that can be provisioned and released rapidly, with minimal management effort or interaction[1]. Today, many large commercial Cloud computing service providers virtually make unlimited storage and processing capabilities available to their users over the Internet and follow a competitive pay-per-use business model, thus offering on-demand low-cost virtualized computing resources with high elasticity and flexibility, providing both technical as well as economic benefits. For instance, Cloud computing offers a number of technical benefits including energy efficiency, optimization of hardware and software usage, performance isolation and high availability, to name a few. Similarly, it remains an attractive model for businesses since it significantly reduces the need to invest in in-house computing infrastructure, decreases operating costs, and transfers business risks – to some extent – towards service providers by means of Service Level Agreements (SLA).

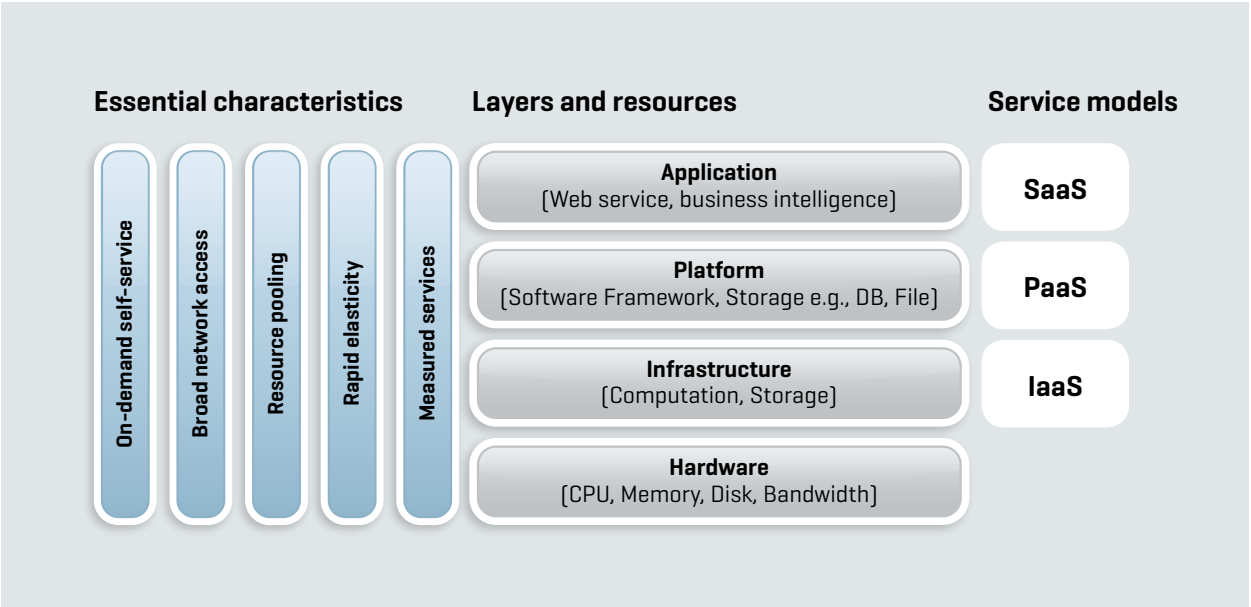


Figure 1: Overview of the Cloud computing paradigm [5]

[1] A similar definition of Cloud computing is also provided by the National Institute of Standards and Technologies (NIST) [6]

To summarize, as shown in Figure 1, Cloud computing is designed to possess the following characteristics [5] [6]:

- **On-demand self-service:** On the one hand, users can order and manage services offered by the service provider without human interaction (e.g., using a web portal and management interface). On the other hand, at the service provider's side, the provisioning and de-provisioning of services and associated resources occur automatically.
- **Ubiquitous network access:** Cloud services could be accessed via a network (usually the Internet), using standard mechanisms and protocols.
- **Resource pooling:** The infrastructure used to provide Cloud services is shared between all its users.
- **Rapid elasticity:** Resources can be scaled up and down rapidly and elastically.
- **Measured service:** Resource/service usage is constantly metered. These metrics are used to optimize resource usage, are reported to the customer, and are used as input for the pay-per-use business models, among other things.

As shown in Figure 1, a Cloud environment typically consists of four layers: hardware (datacenter), infrastructure, platform and software. Each layer provides a service by itself and acts as a service for the layer above it. The hardware layer (usually datacenters) act as the backbone, based on which three categories of Cloud computing services are typically offered:

- Infrastructure-as-a-Service (IaaS) provisions computational resources that allow users to store, process and manage their data and applications;
- Platform-as-a-Service (PaaS) provides a set of middleware tools that simplify application development and deployment;
- Software-as-a-Service (SaaS) refers to the provisioning of applications (e.g., web services) running on Cloud environments that are accessible via a web/client browser.

This layered and service-oriented architecture of Cloud computing has not only resulted in a rich ecosystem of innovative services but also triggered the rapid growth of other major ICT paradigms such as the Internet of Things (see Section 1.1.2) and Big data (see Section 1.1.3). For instance, Cloud computing not only provides the computational infrastructure required to store and process massive amounts of data but also offers services that enable faster and scalable ways to integrate, analyze, transform, and visualize various types of structured, semi-structured, and unstructured data in real time, thus providing a means for realizing the full potential of Big data.

1.1.2 Internet of Things (IoT)

Internet of Things refers to a network of interconnected objects that are uniquely addressable, built on standard communication protocols, and whose point of convergence is the Internet [7]. The fundamental idea behind the notion of IoT consists in connecting the objects that people use in everyday life, thus making a pervasive presence and enabling a wide range of services that were otherwise infeasible to be realized. The applications of IoT are already visible in several sectors including (but not limited to) smart cities, industrial services and healthcare [7].

The growth in the number of Internet-connected devices and the increasing variety of such devices, spanning everyday activities, is contributing towards the IoT vision. The number of such objects/devices has grown exponentially in the recent past. Although studies are making different estimates about this trend for the near future, they all expect continuing growth [7]. Gartner [8] for instance estimates that about 11.2 billion devices connected to the Internet will be in operation worldwide in 2018 (about 33% higher than in 2017) and this number will reach around 19 billion by 2019. Predictions of several organizations provide a wide range of estimates of the total number of IoT devices, from a low of 19 billion to a very optimistic prediction of up to 40 billion [7] [9] [10].

Note that devices – referred to as *things* in the IoT context – not only include mobile phones or electronic appliances but also relate to objects such as clothing, food containers, furniture, artworks, sensors in buildings etc. The large number and variety of these things suggest that IoT devices could serve as one of the primary sources for data acquisition in the Big data paradigm (see Section 1.1.3).

The basic components of IoT are as follows:

- A **device/thing**, that is hardware and software, which interacts with the world. Typically, devices connect to a network to communicate with each other or to centralized applications. They connect directly or indirectly to the Internet. There are two main types of devices:
  - **Sensors** are a type of devices that gather information from the environment.
  - **Actuators**, on the other hand, are devices that reach out and act on the world.
- Things are connected using wireless and wired technologies, standards and protocols, to provide pervasive **connectivity**.
- Given the heterogeneity of devices, their limited storage and processing capabilities, and a variety of applications, **middleware** plays a key role in abstracting the functionalities and communication capability of devices. Middleware not only connect components such as things, people and services, but also enable access to devices, ensure appropriate installation and behavior of devices, in addition to facilitating interoperability between local networks, Cloud or other devices.

A simplified IoT architecture could be viewed as a composition of four layers [11]:

- **Sensing and actuation layer:** this bottommost layer comprises a wide range of devices/things (e.g., sensors, actuators, gateways).
- **Transmission and communication layer:** contains networking and transport capabilities (e.g., support of a set of communication protocols).
- **Storage and processing layer:** comprises components to store the data generated and for processing the acquired data.
- **Application layer:** this topmost layer contains the IoT application user interface.

As a transversal layer, security and other management capabilities and functions are often realized. A detailed IoT architecture is presented in [7]. The scientific developments related to security, privacy and data protection in IoT are summarized in Chapter 3 along these layers.

1.1.3 Big data

The term Big data was introduced in 1997 to refer to large volumes of scientific data that was mined for better visualization [12]. Currently, Big data is often defined in terms of V-characteristics. The first three Vs of Big data were introduced in 2001 as follows [13]:

- **Volume** refers to the quantity of data being generated.
- **Velocity** refers to the speed at which data is collected and how fast it is processed.
- **Variety** refers to the diversity of data types.

Over the years, other concepts have also been attributed to Big data [14], as discussed below:

- **Veracity** refers to the quality or trustworthiness of the data.
- **Variability** refers to the changes in data structure, semantics, quality, etc. over time.
- **Volatility** refers to a limited time span in which data values remain relevant for a particular analysis.
- **Visualization** refers to the presentation of data in a way that it is understandable by user.

- **Value** refers to the monetary Return on Investment (ROI) over the cost of building, using and maintaining the data processing system.

Today, the notion of Big data comprises activities starting from data generation to the point where hidden knowledge is uncovered using data mining, machine learning and/or artificial intelligence algorithms. This process is also referred to as Big data value chain and has been defined in different studies (e.g., [15]). Figure 2 illustrates simple Big data value chain comprising three main classes of activities.

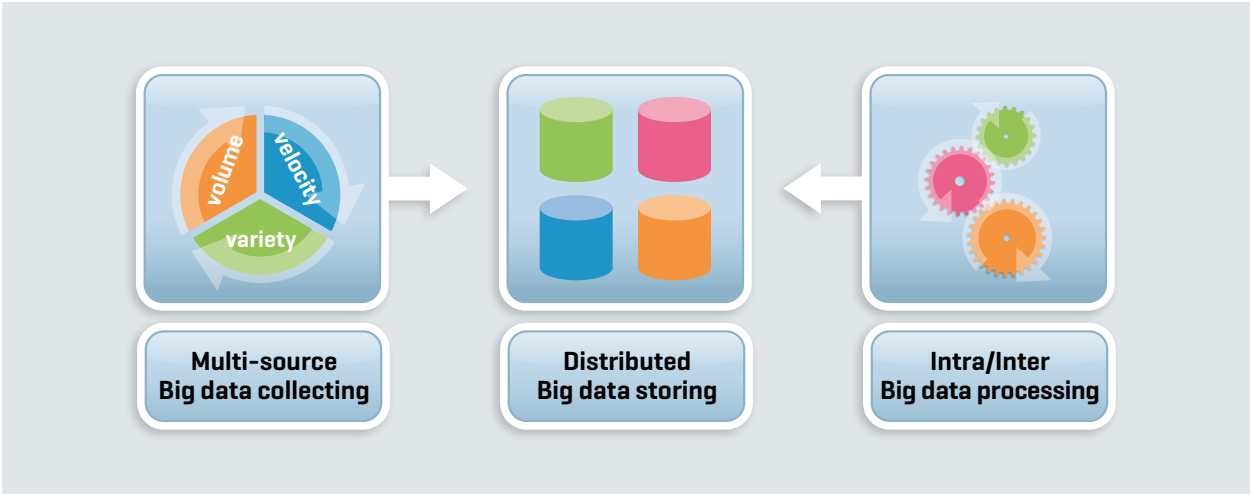


Figure 2: Big data value chain and general architecture of Big data analytics [15]

- **Big data collection** refers to the process of gathering (from various sources), categorizing, and cleaning data before it is stored. The challenges within this process (depending on the application context) arise due to infrastructural requirements. For instance, while some applications require low and predictable latency in capturing data and in executing queries, others require capabilities to handle high transaction volumes. Support for distributed environments, dynamic data structures, and a combination of the above are often requirements of state-of-the-art applications.
- **Big data storage** refers to persistence and management of data in a scalable manner such that it satisfies the needs of applications such as fast access/retrieval of information.
- **Big data processing and analysis** concerns making the acquired raw data amenable to use in decision-making as well as domain-specific usage. The task of data analysis involves exploring, transforming, and modelling data with the goal of highlighting relevant data, synthesizing and extracting useful hidden information with high potential from a business point of view. Data analysis is closely associated to areas such as data mining, machine learning and artificial intelligence.

Data usage involves activities that need access to a large amount and variety of data and its analysis in order to enhance business's competitiveness through costs reduction, addition of value-added services, or in building new and innovative applications. The value of data grows significantly once it is analyzed, the analytics results used and ROI calculated. However, Big data could also be seen in the context of data value chain with a perspective of potential value delivery. From business standpoint, data value chain is at the center of the future knowledge economy, and has the potential to bring the opportunities of the digital developments to the more traditional sectors (e.g. transport, financial services, health, manufacturing, retail) [16]. This transformation could create numerous business opportunities such as [17]:

- Data monetization (collecting and selling data).
- Developing technical tools for data storage, processing, analysis, etc.
- Offering services like recommendation systems, optimized information research, deep and correlated insights, augmented reality, etc. for end users.

Table 1 outlines some domains where Big data could bring benefits, based on [17] (note that this is not an exhaustive list; it is provided as an example for better understanding).

Domain	Applications of Big data
Environment	Environmental data is required to understand climate change, evolution of the planet and the impact of human activities on the planet. Reliable and up-to-date information on environmental changes could help governments to elaborate the sustainable environmental policies. Satellite images could provide such information and Big data would allow optimized image processing.
Mobility, transport and logistics	Logistics sector could benefit from Big data, especially in setting of multimodal urban transportation. Combining traffic information, Machine-to-Machine (M2M) communication between vehicles, sensor information from the environment, etc. would allow optimization in the logistical processes and deliver higher-quality services. This could in turn result in economic and environmental savings as well as better user satisfaction.
Manufacturing and production	Introducing sensor-equipped machinery into the manufacturing and production chain could help human workers and bring efficiency gains. The product customization is easier than in the past, thanks to Big data frameworks. Industry 4.0 is the new form of manufacturing that is based on Smart ICT.
Healthcare	Diagnosis of illnesses could be made more efficient using Big data, even for rare diseases. However, since such healthcare applications process sensitive data, measures must be in place to protect patients' data as well as privacy.
Financial services	The huge volumes of available data could make it possible to efficiently detect fraud, evaluate and reduce risks, analyze customer behavior, trade efficiently, prevent cyber-attacks on sensitive services, to list a few.
Retail	Consumers expect personalized services with high levels of availability. With Big data, marketing campaigns could address specific needs of the consumers. In addition, external data such as competitors' prices and weather conditions could be used for demand forecasting and pricing.

Table 1: Applications and importance of Big data in different domains [17]

Active management of data over its life cycle is necessary in order to ensure data quality requirements, thus making the overall Big data application trustworthy and fit for purpose (see Chapters 3 and 4 for details).

1.1.4 Smart ICT – convergence of Cloud computing, IoT and Big data

Cloud computing, IoT and Big data are major technologies contributing to the notion of Smart ICT. Although developed independently, these technologies are converging in new ways, unlocking each other's true potential, and transforming the overall technological landscape. Following [1], Figure 3 illustrates a Smart ICT system with different components and their interactions. IoT involves numerous devices that communicate with each other and/or to a centralized application. These connected devices interact with the real world, collect a range of information, and transmit the collected data to the Cloud by means of a gateway service. Similarly, IoT devices also receive data from the Cloud (typically, about the actions to be performed by an actuator) and acts on the real world accordingly.

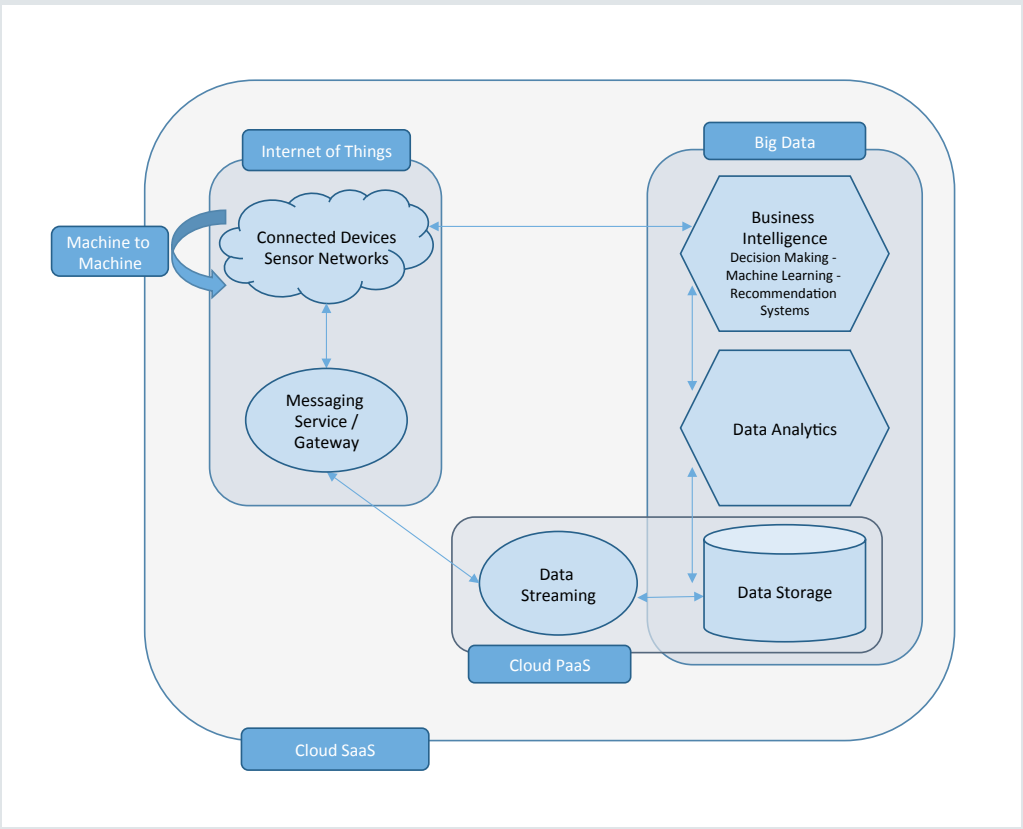


Figure 3: Smart ICT Components and their interactions [1]

The large number of IoT devices, their heterogeneity as well as limited storage and processing capabilities, make Cloud computing services' support for IoT applications ideal (see Section 2.1). For instance, using Cloud computing services, IoT applications could benefit from the economies of scale, and realize sensor-centric data-driven complex applications that are otherwise infeasible to implement.

As discussed above, IoT generates data involving numerous sources, possibly in different formats, originating at varying frequencies and volumes. This implies that the data of an IoT application could be characterized as Big data that is generated in the form of a continuous real-time data stream (see Section 2.3). To handle this data stream, the IoT application could benefit from Cloud services that could store stream data, filter, aggregate and transform it for suitable use for analysis (see Section 2.2). The transformed stream data can then be queried using analytical tools and integrated with traditional business intelligence solutions, thus providing a holistic Smart ICT application (see Section 2.4).

1.2 Data protection and privacy in Smart ICT

The advancements in Smart ICT, while allowing users to easily access high quality innovative applications and services, introduce a range of privacy risks of improper information disclosure and dissemination [2] [1] [3]. Ensuring proper data protection and privacy of information stored, transmitted, processed, and published by Smart ICT applications as well as of users who leverage Smart ICT is a major challenge.

Consider the Cloud computing context for instance: it is challenging to ensure that sensitive data remain properly protected and that users maintain control over who could access what part of their data stored on an external Cloud server. In other words, a risk arises due to the loss of control among Cloud users, as compared to traditional in-house systems. Cloud Security Alliance[2] has identified major security threats to Cloud computing that need to be addressed in order to mitigate risks for Cloud users and providers [18]. These threats range from data breaches (e.g., sensitive, protected or confidential data being disclosed to unauthorized entities), ineffective identity, credential and access management, to advanced persistent threats, denial of service and shared technology vulnerabilities [18].

On the other hand, IoT presents a unique set of information security challenges, specifically due to its highly distributed nature, involvement of a large number of diverse devices, and a very large attack surface. Consequently, privacy and data (e.g., personally identifiable information – PII[3]) protection remains a significant concern for IoT systems. The potential disclosure or misuse of PII handled by an IoT system, that could cause harm to the people identified by the information is not only a technological but also reputational and a trust concern [19]. In fact, as the sensing, actuation, communication and control methodologies are becoming sophisticated and as IoT systems are handling more security-critical and privacy-sensitive data, they are becoming attractive targets for attackers.

The security of Cloud computing and IoT systems are essential to protect the (Big) data from unauthorized disclosure. The former provides the infrastructure to store and process (Big) data and the latter are the main source of data within a Smart ICT system. Moreover, globally, each jurisdiction defines laws and regulations (e.g., European General Data Protection Regulation – GDPR[4]) that systems must comply with. These regulations are an opportunity to create a trusted environment. However, there could be difficulties in implementing a compliant system. For example, in Cloud environments, challenges may arise for users to keep track of resources that are being used and their physical location. The responsibilities of parties involved may be ambiguous in certain contexts. This situations exacerbates when an IoT system acquires/uses PII or sensitive data (e.g., related to the health of an individual) and the need for holistic approaches to privacy and data protection in Smart ICT become of paramount importance.

The Big data paradigm allows the data captured using IoT devices to be associated with other data sources (e.g., social networks and healthcare applications) in order to gain deeper insights. In fact, the essence of Big data analytics is to provide inferences between large volumes of extremely varied data, which do not necessarily have any direct correlation at the time of processing. To achieve this, Big data analytics focus on data maximization, while one of the fundamental principles to ensure privacy and data protection is data minimization [20]. This possibility of combining and analyzing information from several data sources (including IoT devices handling PII) has intensified the challenges in ensuring privacy. There is need for a wide range of solutions, for instance, to empower users in understanding the privacy risks to which they are exposed and, to maintain control over their own information, while benefiting from the Smart ICT services.

[2] https://cloudsecurityalliance.org

[3] PII refers to any information that: i) might be directly or indirectly linked to an individual or ii) could be used to identify the individual to whom such information relates [140].

[4] http://data.europa.eu/eli/reg/2016/679/oj

The abovementioned challenges have attracted considerable attention of the research community as well as technical standardization organizations in the recent years. Several research activities have been performed, proposing novel solutions for protecting privacy at various stages of data lifecycle (e.g., collection, storage and analysis). Chapter 3 provides a holistic view of security, privacy and data protection challenges as well as recent research developments in Cloud computing, IoT and Big data contexts.

1.3 Technical standardization

The increasing demand for transparency, coherence, and effectiveness in the development of Smart ICT domains have resulted in a huge demand for careful study and development of relevant technical standards. In this context, technical standards serve as a reference for products, services and processes, and among other benefits, ensure:

- **Transparency** in the sense that all essential information (e.g., recommendations and guidelines on using a Smart ICT solution) is accessible to all interested parties.
- **Coherence** in order to avoid duplication of work; in this respect, technical standards are developed by cooperating and coordinating with other relevant Standards Developing Organizations (SDOs).
- **Effectiveness and relevance** implying that international standards remain relevant and respond to regulatory and market needs effectively as well as take into account scientific and technological developments. Such standards promote fair competition and innovation.

The benefits listed above are applicable in general to any technology since they are based on the fundamental principles for developing technical standards (see Chapter 4).

Technical standards also address specific challenges of Smart ICT domains and focus on different aspects such as terminology, reference architecture, interoperability, security etc. These benefits improve confidence and trust in adopting Smart ICT domains. Consider Big data as an example; this domain could benefit from the following set of standards (note that this list is not exhaustive and is only an example):

- Standards are essential for improving **trust** in Big data, e.g., with respect to Cloud computing, by enabling interoperability between various applications and consequently preventing vendor lock-in.
- Similarly, standards providing guidelines and good practices for using analytics techniques (e.g., machine learning) could ease adoption and help in building **trust** in Big data.
- In general, analytics designers tweak a reference model repeatedly in order to fit the data in their application context. This practice could result in interpreting noise or randomness as truth. Technical standards could help in **preventing** such **over-fitting** in Big data.
- Big data systems should be designed with **security** in mind (e.g., to ensure confidentiality of data). If there is no global perspective on security, then solutions may become fragmented and they will offer only a partial sense of safety rather than full security.
- Standards could play an important role in **data quality** and **data governance** by addressing the veracity and value of data.

As a response to these needs and benefits, standardization bodies at international as well as European levels have initiated a range of activities that could increase market confidence in Smart ICT and relevant data protection and privacy topics. For instance, an international standard related to security controls for Cloud computing has been published (ISO/IEC 27017:2015 [21]) and a new one regarding security and protection of PII in Cloud SLAs (ISO/IEC 19086-4 [22] [23]) is currently under development. Similarly, a standard concerning Big data security and privacy is currently under development.

Besides Smart ICT, a large set of technical standards that addresses security, privacy and data protection issues in general have been developed and some are widely adopted. For instance, the standards ISO/IEC 27001:2013 *Information technology – Security techniques – Information security management systems (ISMS) – Requirements* [24] and ISO/IEC 27002:2013 *Information technology – Security techniques – Code of practice for information security controls* [25] are used by a number of organizations and stand as a reference for ISMS compliance. Similarly, ISO/IEC 29100:2011 *Information technology – Security techniques – Privacy framework* [26] is applicable to natural persons and organizations involved in specifying, procuring, architecting, designing, developing, testing, maintaining, administering, and operating information and communication technology systems or services where privacy controls are required for the processing of PII. This standard: i) specifies a common privacy terminology, ii) defines the actors and their roles in processing PII, iii) describes privacy safeguarding considerations, and iv) provides references to known privacy principles for information technology.

Chapter 4 provides a detailed overview of various developments in technical standardization in the areas of security, privacy and data protection in general as well as in all three Smart ICT domains.

1.4 Outline of the white paper

This white paper focuses on privacy and data protection in Smart ICT. It describes the relevant state-of-the-art from research as well as technical standardization perspectives and attempts to understand the links between these two lines of work. The rest of this white paper is thus organized as follows:

- **Chapter 2:** This chapter introduces a Smart ICT data model. The interaction between different Smart ICT domains is presented and a model analyzing how data serves as the common thread to all three Smart ICT domains is provided. This data model is the conceptual and terminological unification and enables high-level understanding of integrative Smart ICT components.
- **Chapter 3:** This chapter summarizes the literature (research and scientific developments) concerning security, privacy and data protection in Smart ICT.
- **Chapter 4:** While the previous chapter analyzes research results, this chapter provides an overview of various developments in the areas of technical standardization. After providing some background about standardization, including an overview of standards developing organizations as well as their working principles, this chapter provides details about technical committees that focus on security, privacy and data protection, and outlines their most relevant projects. Finally, details about standardization activities in Cloud computing, Big data and IoT are provided respectively.
- **Chapter 5:** This chapter builds on the results of Chapters 3 and 4 and highlights some links between research results and standardization developments.

2

Data model

2. Data model

As discussed in Chapter 1, Cloud computing, IoT and Big data are interacting in new ways, unlocking each other's true potential, and transforming the overall technological landscape. As illustrated in Figure 4, this chapter first analyzes how Cloud computing and IoT integrate with each other, where the data collected by numerous IoT devices are transmitted to the Cloud, complexities and functionalities abstracted, and decision-making is effectively performed in order to act on the real world. Having established that Cloud computing and IoT could complement each other, this chapter then provides insights on the flow and components of a typical Big data architecture, integrated with Cloud computing. The third part interprets IoT from a data perspective and analyzes its relationship with Big data. The result of the above is that **data** acts as a **common** thread between individual **Smart ICT** domains.

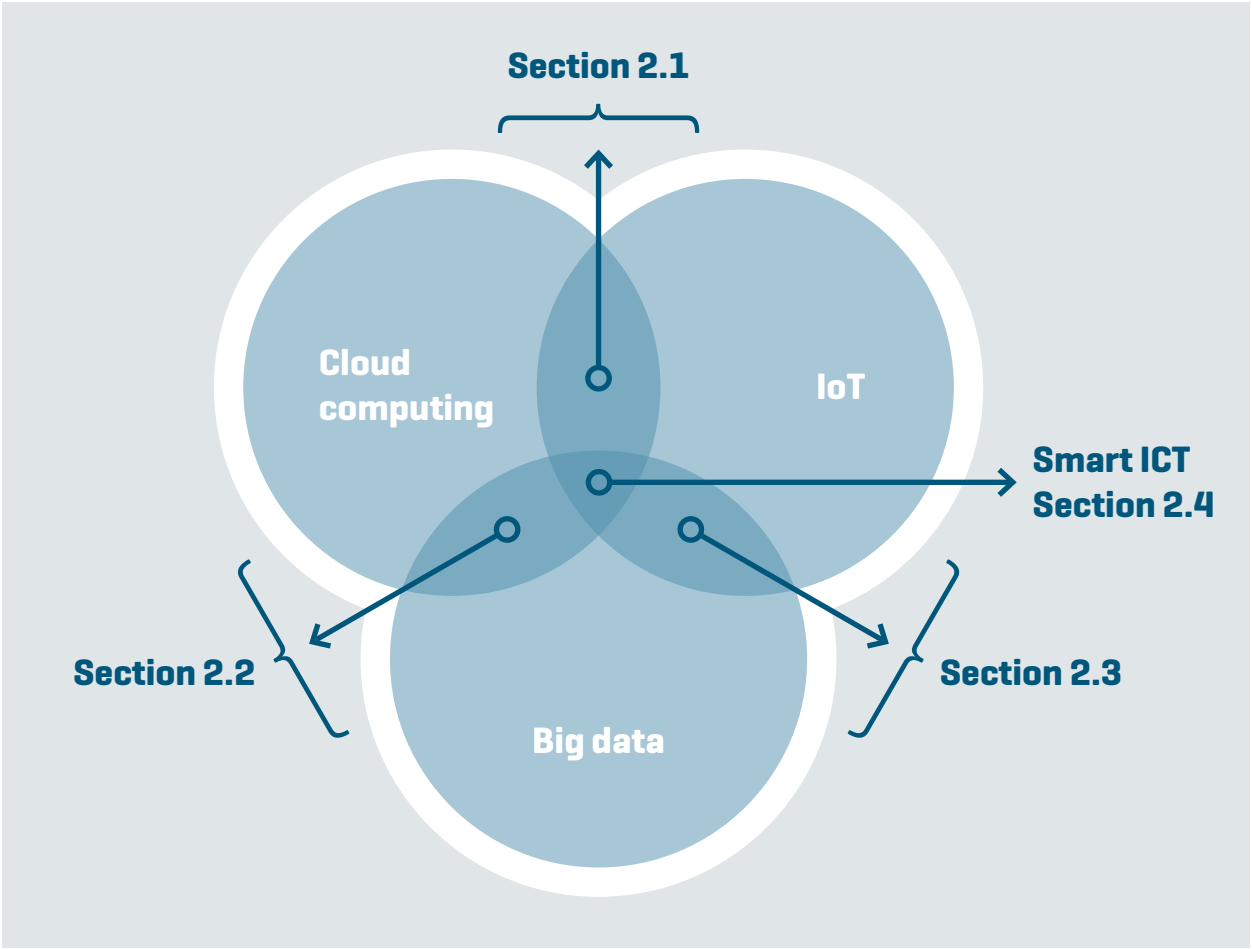


Figure 4: Overview of Chapter 2

In summary, within a large-scale Smart ICT application, numerous IoT devices generate and collect data (typically in different formats, varying frequencies and volumes) that could be characterized as Big data. To handle generated data, the Smart ICT application could benefit from Cloud computing services where, depending on the context, data could be stored, filtered, aggregated and transformed for analysis. The transformed data could then be queried using analytical tools and integrated with traditional business intelligence solutions, thus providing wide-ranging and deep insights on the data.

2.1 Cloud computing and IoT

The characteristics of Cloud computing and IoT could be interpreted as complementing each other and their integration could benefit several application scenarios. A comparison of the characteristics of these two technologies is provided in [5] and is summarized in Table 2. In general, to overcome its technological constraints (e.g., limited storage and processing capabilities), IoT can benefit from the easily accessible and virtually unlimited resources that Cloud computing provides. On the contrary, Cloud computing could enlarge its scope by integrating with real world things that would enable it to create a more distributed and dynamic framework, offering new services connected to real-life scenarios.

Characteristic	Cloud computing	IoT
Displacement	Centralized	Pervasive
Reachability	Ubiquitous	Limited
Components	Virtual resources	Real world things
Computational capabilities	Virtually unlimited	Limited
Storage	Virtually unlimited	Limited or none
Role of the Internet	Means for delivering services	Point of convergence
Big data	Means to manage data	Source of data

Table 2: Comparison between Cloud computing and IoT characteristics [5]

To gain a deeper understanding on how Cloud computing and IoT could complement each other consider the following basic aspects as **examples**:

- **Storage:** By definition, IoT involves numerous sources of information, each providing different sets of non-structured or semi-structured data, which originate at varying frequencies and volumes. The data handled in a typical large-scale IoT application could be considered as **Big data**. Since IoT devices have limited storage capacities by themselves, they require support to effectively store and utilize the generated data [28]. On the other hand, given that Cloud computing offers virtually unlimited, low-cost and on-demand storage, it serves as the most convenient solution to address the storage requirements of IoT applications. The combination of Cloud computing and IoT could generate opportunities for data aggregation, sharing and integration as well.
- **Computation:** IoT devices have limited processing and energy resources. Consequently, instead of on-device processing, collected data is typically transmitted to devices that are more powerful, where aggregation and processing could be performed in a timely manner. However, the involvement of a huge number of devices often result in scalability issues particularly when the underlying infrastructure is not well-equipped [5]. In this context, Cloud computing could address IoT's limitations by providing processing capabilities and by allowing IoT to achieve near real-time data analysis [29], to implement scalable, collaborative and sensor-centric complex applications [30], with possibilities for task offloading and energy savings [31].
- **Communication:** Cloud computing allows IoT to implement effective and cheap solutions to connect, track, and manage devices from anywhere, at any time, using customized portals and built-in services. Moreover, effective monitoring and control of remote things, their coordination, communication and real-time access to data becomes feasible. Note that, although Cloud computing can significantly improve communication capabilities for IoT applications, limitations could arise while transferring huge amounts of data from devices onto the Cloud [32].

In an integrated Cloud and IoT system, Cloud computing can be viewed as an intermediate layer between the IoT devices and final (IoT) applications, which abstracts complexities and functionalities inherent to implementing the latter. This flexibility comes with cost-effectiveness in data collection, processing as well as rapid setup and integration of new things. As a result, data-driven decision-making can be effectively realized, providing a means for developing innovative services (see Section 2.2).

If the risks associated with these technologies (e.g., scalability, interoperability, reliability, privacy and security) are carefully addressed, this will increase the potential to further transform the ICT sector and, in a more general sense, the society itself. For instance, in a given application, security issues concerning IoT devices (sensors and actuators) and communication between them and with the Cloud need to be addressed in order to protect the data and user's privacy [7].

2.2 Cloud computing and Big data

Similarly to IoT, Cloud computing is foundational/complementary to Big data applications and provides a wide range of relevant services including storage and processing of data (taking into account data formats, volume and generation velocity, etc.) as well as tools to filter, aggregate and transform data for suitable analysis.

Note that Big data handles the collection, processing, and analysis of large amounts of datasets that are complex/ infeasible for traditional databases. Organizations adopt Big data approaches at different thresholds (e.g., some may use Big data tools for hundreds of gigabytes of data while others may do so for several terabytes of data[5]) based on their capabilities and needs. The reducing costs of storage, ever-increasing means of data collection and availability of a wide range of advanced tools have accelerated the growth of Big data over the years [33]. Accordingly, though this term originally referred to the size of the data, today it increasingly relates to the value that could be harnessed from available datasets.

Figure 5 illustrates the three main logical steps in line with the Big data architecture described in Section 1.1.3 and corresponding components as examples. While most applications may involve all or a subset of these components, depending on data sources and goals, components tailored for specific purposes may also be integrated. Similarly, new logical steps (e.g., a presentation layer on top of analysis) could be included. Note that many of these components from storage models to analytical tools and orchestration methods are in fact Cloud computing services [33].

[5] https://www.mongodb.com/big-data-explained

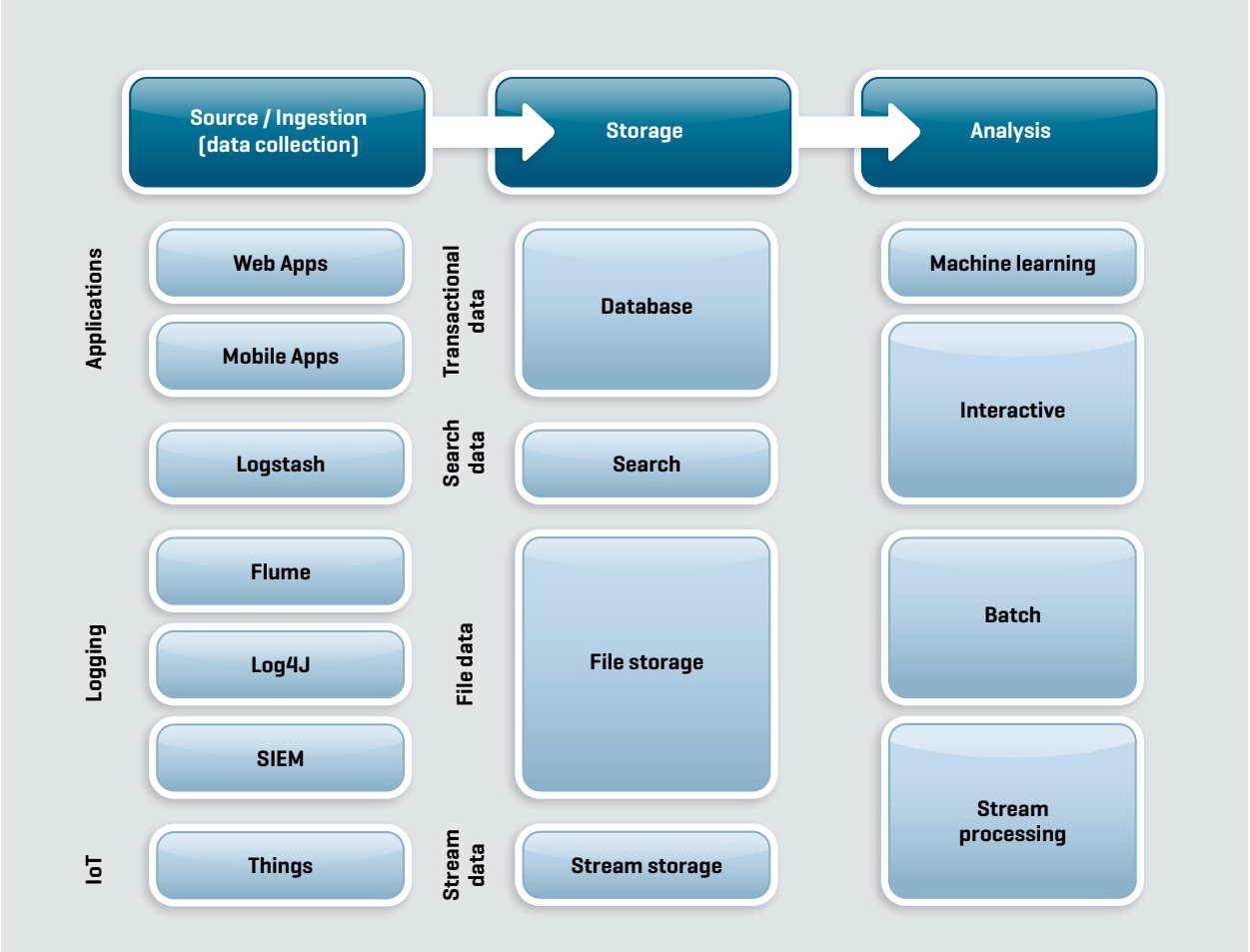


Figure 5: Convergence of Cloud computing and Big data [33]

Availability and integration of one or more sources of data is typically the starting point in understanding a Big data solution. Data sources may include, but are not limited to, application data store (e.g., relational database), static files (e.g., log information), transactional data (e.g., from mobile or web applications) and real-time data streams (e.g., arising from IoT devices) [33]. Based on **data characteristics**, it is **stored** using appropriate Cloud services. For instance, data for **batch processing** operations is often stored in data lakes (distributed file store that can hold high volumes of large files in various formats), and given the large sizes of datasets, they are processed using long-running batch jobs that filter, aggregate, and prepare the data for analysis. In general, the output of batch processing is structured data, which is saved on a new output file.

On the other hand, if an application involves real-time data sources (e.g., IoT devices), approaches to capture and store such data for **stream processing** is necessary. While a simple data store (e.g., where incoming messages are stored in a folder) could be used, a more efficient approach consists in having a message ingestion store that acts as a buffer for messages, and supports scale-out processing, reliable delivery and message queuing semantics [34]. Data streams are complemented with processes for filtering, aggregating, and for transforming the data so that it is more suitable for analysis. Similar to batch processing, the processed stream data typically takes a structured format that can be queried using analytical tools. A variant of a relational data warehouse (e.g., traditional business intelligence solutions), or a low-latency NoSQL technology, or an interactive database that provides a metadata abstraction over data files in the distributed data store could be used for analytics-ready stream data.

To summarize, Big data solutions typically involve one or more of the following types of workflows:

- Batch processing of Big data sources at rest;
- Real-time processing of Big data (streams);
- Interactive exploration of Big data;
- Analytics and predictive insights using machine learning techniques.

Given that the goal of most Big data applications is to generate insights from available data, Cloud services that provide data analysis algorithms (e.g., predictive analysis using machine learning techniques) and reporting are available. MapReduce [35] is one of the most popular programming models to process large amounts of data. Hadoop[6] is a widely adopted open source MapReduce implementation and several Cloud providers have made Hadoop available to their users. Hadoop allows to partition and replicate data sets across multiple nodes, such that when running a MapReduce application, data that is stored on the node could be accessed where it is executing, thus providing a highly distributed file system. Hadoop not only provides a set of APIs that allows developers to implement MapReduce applications but also support the use of high-level query languages such as Hive[7]. A survey about the features, benefits, and limitations of MapReduce for parallel data analytics is presented in [36]. Similarly, for applications that require repeated data processing operations as part of their workflows (e.g., to iteratively transform data, to transmit data for multiple processes, or to provide analytics results as a report), orchestration techniques are being made available. Orchestration services are designed to handle workflows, synchronize parallel executions (meta-scheduling) and to provision and manage computational resources.

[6] http://hadoop.apache.org/

[7] https://hive.apache.org/

2.3 IoT and Big data

One of the goals of Internet of Things is to create a pervasive presence where everyday objects are interconnected and provide new and improved services. Given its multi-faceted presence, IoT is being studied from **different viewpoints**, and some examples are provided here [7] [37] [38]:

- **Services:** From this viewpoint, IoT is defined as a world where things can automatically communicate to computers and each other providing services to the benefit of the human kind.
- **Connectivity:** captures the notion that IoT moves from anytime, anyplace connectivity for anyone, to having connectivity for anything.
- **Communication:** From this viewpoint, IoT refers to a worldwide network of uniquely addressable interconnected objects that are built on standard communication protocols.
- **Networking:** This IoT viewpoint refers to the change in the scope of the Internet from being a network of interconnected computers to a network of interconnected objects.

In line with the primary focus of this white paper, **IoT** could be viewed from a **data perspective** as follows [37] [38]: in contrast to the traditional Internet environments where the primary data producers and consumers are human beings, in IoT the majority of data producers and consumers are the things. In this context, computers obtain information and solve real world problems for humans using the data generated by things.

To understand the data perspective of IoT, details about data models and data handling approaches are summarized in this section. A number of data models describing the devices, their attributes and associated schema are being studied [39]. These data models are typically industry-specific. For example, while OPC Foundation's data model[8] is generally considered for industrial automation applications, other implementations such as smart homes use entirely different modeling and schema standards [39].

In a Smart ICT context, IoT devices capture and periodically transmit data records to the **Cloud**. The process of uploading data records into Cloud is called **ingestion** [40] [39]. The transmitted data usually takes the form of continuous data streams – often arising from multiple sources – and comprises messages, events, telemetry and alerts.

While the terms **messages** and **events**, used interchangeably, refer to the data **generated** by connected devices, **telemetry** refers to the data collected by devices (i.e., the messages carrying data reported by device sensors). For instance, a telemetry data record might refer to the current temperature captured by the sensor on a device. These records may comprise individual or multiple data points [40]. For example, a device with a humidity and a temperature sensor might send both measurements either in the same message or as separate messages.

Similarly, given that devices could have multiple sensors, telemetry may comprise either measurements reported by all sensors or only the values changed since the previous telemetry. The value of a data point in a telemetry record becomes the last known **state**. Note that **state information** describes the status of the device and not of the environment. For the sake of consistency and synchronization, telemetry may occasionally include a full snapshot of all sensors values (generally called a **key frame**) [39] [40] [41] [42].

Each source of telemetry results in a **channel**. Depending on the data and application, telemetry data could be stored and processed as a stateful variable on the device or in the Cloud, and analyzed against a set of rules. The results of an analysis may generate a new data type, commonly referred to as an **alert**. Note that, although each device periodically sends only a single data point, the large number of devices involved and the heterogeneity of telemetry data, motivates the **use of Big data strategies** and methodologies.

[8] https://opcfoundation.org/

Finally, a recommended practice for IoT solutions consists in including **metadata** in each record. **Metadata** for instance could contain information about a device such as device identifier, class or type, model, date of manufacture and firmware update information. Given the vast variety of IoT data, their sources and applications, there is no specific standard defining IoT data taxonomy. Based on the study [41] [43], Table 3 attempts to categorize and provide a taxonomy of IoT data.

Category	Characteristics	Description
Data generation	Velocity	IoT data could be generated at different rates. For example, for traffic monitoring sensors in a smart city, signal-sampling frequency could vary from every few seconds, minutes, or half an hour. It is challenging to handle very high sampling rates (e.g., processing power) as well as to deal with low sampling rates (e.g., important information may be lost).
	Scalability	As the number of devices generating continuous stream of data further increases, IoT applications need to be designed for extreme scalability. This aspect is in line with the current Big data trend.
	Dynamics	IoT is a highly dynamic system and so is the IoT data. For instance: • Several devices could be mobile. This implies that data will be generated from different locations at different times and from different environments. • Devices could be fragile and are prone to failures. • The connections between devices could be intermittent, creating another source of dynamics in IoT data processing.
	Heterogeneity	In line with the goal of IoT to have pervasive presence, more and more devices are becoming "smart" and being connected to the Internet. These devices generate data in different formats and using different semantics.
Data quality	Uncertainty	Uncertainty could arise from different sources. For instance, from the sensing precision or accuracy of devices.
	Redundancy	IoT data could have redundancy due to various factors. For instance, a group of sensors of the same type deployed in a nearby area could produce similar sensing results. Similarly, due to high sampling rates, redundant sensing data could be produced.
	Ambiguity	IoT data will inherently have large amounts of ambiguity. Other devices or data consumers could interpret the data generated by assorted devices in different ways.
	Inconsistency	Inconsistency is also prevalent in IoT data. When multiple sensors are monitoring the same environment and reporting sensing results, due to precision and accuracy issues, data inconsistency could be observed.
Data interoperability	Incompleteness	In order to process IoT data, being able to detect and react to events in real-time, it is important to combine data from different types of data sources to gain situational awareness. However, since this process relies on the data generated by mobile and distributed devices, incompleteness could easily be observed.
	Semantics	Semantics within IoT data could play an important role in the process of enabling things/machines to understand and process IoT data by themselves.

Table 3: IoT data taxonomy [38] [41]

2.4 Data as the common thread in Smart ICT

Building on the insights provided above, Figure 6 illustrates an overall Smart ICT system and its components. One of the core components are the IoT devices that interact with the real world, have the ability to register with the Cloud, and provide connectivity for sending and receiving data to and from the Cloud. The data from each device is connected (sent and received) with the Cloud by means of gateway services. In this context, a gateway describes a class of devices that process data on behalf of a group or cluster of devices, have device management capabilities (including command and control), facilitate telemetry and event ingestion, and finally, that enable devices that are not directly connected to the Internet to reach Cloud services.

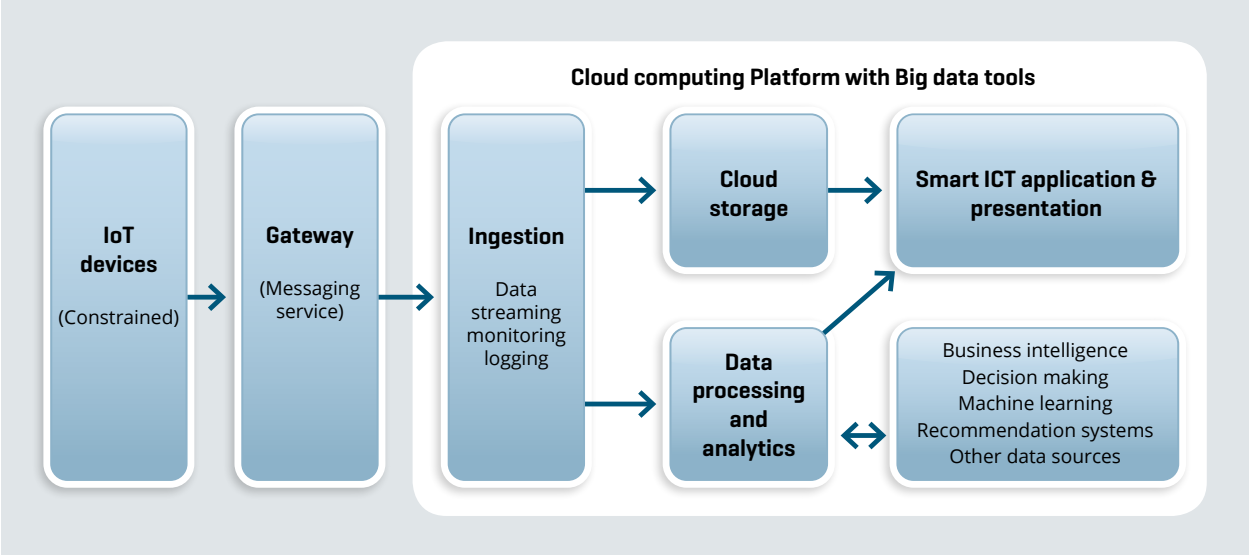


Figure 6: Integration of Cloud computing, Big data and IoT

In certain contexts, once the data from a sensor is collected, the device may process the data before sending it to the Cloud. Such processing may include (as examples):

- Validating data to ensure that predefined rules are satisfied.
- Summarizing data to reduce the volume as well as to eliminate unnecessary details.
- Enhancing data to append additional relevant information.
- Consolidate data into batches.

Irrespective of whether or not the data is processed by devices, data is sent to the Cloud where **stream processors** consume that data, integrate it with other **business processes**, and store the data using appropriate **storage services** (as discussed in Section 2.2). Given that this data has all the characteristics of **Big data**, relevant approaches are to be employed [1] [7] [33] [39] [40].

- **Stream processors** process large streams of data records and evaluate rules for those streams. Note that, depending on the complexity of the rules, different Cloud stream processing services are available. For example, if the application comprises complex rule processing, services such as Azure Stream Analytics or Amazon Kinesis could be used. Similarly, for simple rules, Azure IoT Hub Routes or Amazon IoT Core could be employed.
- During stream processing, the data (for example, telemetry) from the devices could be **integrated** with **business processes** (e.g., including storage of informational emails, alarms, sending emails etc.), and this practice facilitates easier execution of application-level actions and better insights.

- Depending on the type of data, different **Cloud storage services** could be used. For instance, for data that needs to be stored for longer term and that is used for batch processing, cold path (e.g., Azure Blob Storage or Amazon S3) is used. In contrast, if the data that is required to be available for reporting and visualization immediately from devices, warm path (e.g., Azure CosmosDB or Amazon DynamoDB) is used.
- Other components of Smart ICT applications include:
  - A **user interface** to visualize telemetry data and facilitate device management.
  - A **data transformation** module that allows restructuring, combination, or transformation of telemetry data sent from devices.
  - **User management** to divide the functionality amongst different roles and users (Chapter 3).

Similarly to typical Big data applications, predictive algorithms (including machine learning methods and artificial intelligence techniques) as well as business intelligence solutions can be integrated over historical telemetry data in order to gain deeper insights and enable predictive maintenance. These results and the overall Smart ICT paradigms open a range of opportunities that have the potential to not only address real world challenges but also to transform the society as a whole.

3

Scientific developments – Data protection and privacy in Smart ICT

3. Scientific developments – Data protection and privacy in Smart ICT

Smart ICT domains are rapidly advancing and becoming widely adopted given the promises of flexible resources management, energy efficiency, improved scope for innovation, etc. However, since the resulting (Smart ICT) applications often collect and process large amounts of sensitive data, research efforts have focused on addressing the security, privacy and data protection concerns in order to make Smart ICT trustworthy. This chapter summarizes the outcomes of such research efforts for Cloud computing, IoT and Big data respectively.

3.1 Cloud computing

Cloud computing has many advantages which have been introduced in Chapter 1 (e.g., cost effectiveness, on demand access to computational resources, elasticity etc.). However, aspects of security, privacy and data protection remain a major concern [18]. Table 4 summarizes data protection and privacy challenges and corresponding potential solutions (based on [44]) in three main categories:

- Security, privacy and data protection **controls** in the Cloud (identity management; authentication and authorization; access control policy management).
- Inherent **properties of Cloud** computing (virtualization, secure service provisioning and composition).
- Data **stored** and processed in the Cloud (sensitivity of information; confidentiality, integrity and availability of data; data storage and transfer locations).

The following subsections describe relevant research efforts following these categories.

Cloud computing aspect	Security, privacy and data protection aspect	Challenges	Potential solutions
Security and privacy controls in the Cloud	Identity management, authentication and authorization	• Exporting users' identities • Securely transferring identity attributes	• Federated identity management [45] [46] [47] • Efficient credentials management [48] • Multi-factor authentication [49] • MiLAMob: a SaaS authentication middleware [50] • A user-centric approach for platform-level authorization [51]

Cloud computing aspect	Security, privacy and data protection aspect	Challenges	Potential solutions
Security and privacy controls in the Cloud	Access control	• Provide access only to authorized users • The risks of information leakage	• RBAC (Role-based access control) [52] • An integrated solution which combines trust with cryptographic RBAC [53] • An authorization-as-a-service approach [54] • Multi domain access control policies: a comprehensive policy management framework [55] [56] • A heuristic solution to find an RBAC state [57]
	Policy management	• Auditing and proof of compliance	• A scalable distributed monitoring system [58]
Inherent properties of Cloud computing	Virtualization, secure service provisioning and composition	• In multi-tenancy an attacker having access to a virtual machine deployed on a given physical machine could compromise other VMs hosted on the same physical machine • Service providers and integrators are required to collaborate in order to provide newly composed services to customers	• The Open Services Gateway Initiative service platform [59]
Data stored and processed in the Cloud	Sensitivity of information	• Lack of users' control over Cloud resources	• Enabling users to define transparency policies over their data [60]
	Confidentiality, integrity and availability of data	• Security and privacy of data • Frequent outages reported on well-known CSPs [61]	• Using verifiable proofs of violation by external third parties [62] • Fuzzy authorization for Cloud storage [63]
	Data storage and transfer locations	• The highly distributed nature of Cloud infrastructures • Certain data protection and privacy laws also apply in specific jurisdiction	(e.g., EU's General Data Protection Regulation – GDPR [64])

Table 4: Summary of privacy and data protection challenges and corresponding solutions in Cloud computing [44]

3.1.1 Identity management, authentication and authorization

To improve the resilience of their applications, Cloud users are increasingly provisioning computing resources via heterogeneous (and multiple) service providers. This application architecture, in addition to the design goals of making Cloud applications accessible from multiple locations (e.g., home, office, public places), requires flexibility in exporting users' digital identities and in securely transferring identity attributes to various devices. Furthermore, Cloud users must be able to integrate their Cloud identity management solutions with the enterprise's existing identity management framework.

To satisfy the above-mentioned requirements, **federated identity management** solutions for Cloud environments are being proposed in the literature [45] [46] [47]. Federated identity management aims at establishing the trust among the different service levels of Cloud computing in order to reduce the complexity and risks of security management. For most federated identity solutions, identity is federated only for one service level. In this situation, if a SaaS provider wants to track a user's actions for auditing purposes, to implement the user identification at a lower level (PaaS, IaaS) the user could not use the federated identity solutions because of the unawareness at the lower levels. To cope with this scenario, [45] proposed a new architecture for integral federated identity management. In this structure, a third party identity provider is introduced to authenticate the transparent translation of high-level identities to lower-level identities. This solution brings various security advantages for accounting, auditing, and facilities for access control.

Another line of work consists in providing user-centric identity management approaches for handling private and critical identity attributes [65]. In this approach, identifiers or attributes help identify and define a user. Such an approach facilitates users to control their digital identities and masks the complexity of identity management from the enterprises, thereby allowing them to focus on their own functions [66].

Given that an identity is fundamentally a collection of credentials that uniquely define a given user, solutions for efficient credentials management, including strategies to evaluate the complexity of Cloud ecosystems, have been proposed in the literature. One such solution [48] is based on an innovative credentials classification approach.

Prominent applications of identity and credentials management solutions are authentication and authorization. Accordingly, a design model for multi-factor **authentication** in Cloud computing along with an analysis of potential security threats has been proposed in [49]. Another authentication solution called the MiLAMob [50] provides a SaaS authentication middleware for mobile consumers of IaaS Cloud applications. This proposal (MiLAMob) handles real-time authentication events on behalf of consumer devices with minimal HTTP traffic [44]. On the other hand, [51] proposed a user-centric approach for platform-level **authorization** of Cloud services using the OAuth2 protocol to allow services to act on behalf of users when interacting with other services, thus avoiding sharing usernames and passwords across services [44].

3.1.2 Access control

The concepts of authentication and authorization (of Cloud users using predefined identification schemes) provide the basis for access control mechanisms i.e., ways of ensuring that access is provided only to authorized users. Access control is a critical component in the Cloud computing paradigm. For instance, given that the data are trusted to a third party (Cloud Service Provider – CSP) for handling and/or storage, precautions must be taken to ensure uninterrupted and full control over the data to Cloud users. Similarly, since users of the same Cloud share the data processing and storage infrastructure, they are exposed to the risks of information leakage (either accidentally or intentionally) thus requiring efficient access control mechanisms [67].

In general, role-based access control (RBAC) is one of the most promising models because of its simplicity, flexibility in capturing dynamic requirements, support for the least privilege principle and efficient privileges management [52]. Building on these benefits, [53] proposed a solution for ensuring trust while securely sharing data in the Cloud using cryptographic RBAC techniques.

However, since CSPs do not usually know their users in advance, it is difficult to assign users to roles directly in access control policies. This RBAC limitation remains an open challenge within intensely service-oriented environments such as Cloud computing. In this context, many solutions currently employ credentials-based or attribute-based policies. For instance, [44] and [54] discussed collaborative access control properties and an authorization-as-a-service approach using a formal multi-tenancy authorization system and by providing administrative control over enhanced fine-grained trust models. The goal of this solution is to ensure agility, homogeneity, and outsourcing trust [44].

Going beyond simple access control models, researchers have focused on multi domain access control policies and policy integration issues, which can be adopted to build a comprehensive policy management framework for Cloud computing [55] [56]. Similarly, new approaches for role engineering have been proposed. In the Cloud computing context, changes to the existing role-set could cause disruptions to the organization and prevent it from functioning properly. This implies that role mining should look for a set of roles as close as possible to both the existing and optimal sets of roles. The StateMiner approach [57] is one such solution since it introduces new measures for optimality and presents a heuristic solution to find an RBAC state with the smallest structural complexity that is as similar as possible to both the existing and optimal state.

3.1.3 Security and privacy policies management

Security and privacy **policies management** aims at defining and enforcing rules to enforce certain actions such as auditing and proof of compliance. In some scenarios, CSPs need to collaborate to offer various application services to clients. Different service providers implement different security and privacy strategies. As a result, the heterogeneity among different policies is a big challenge. It is necessary to build mechanisms to make sure that such a collaboration is monitored effectively with security guarantee.

Security violations can still occur during the integration progress even though individual domain policies have been verified [68]. Consequently, the CSPs should ensure that policy integration does not lead to any security breach. To satisfy the continuous monitoring of the Cloud infrastructure to ensure compliance with consumer security policies and auditing requirements, [58] presents a scalable distributed monitoring system for Clouds using a distributed management tree that covers all the protocol-specific parameters for data collection [44]. Data acquisition is performed through specific handler implementations for each infrastructure-level data supplier. Data suppliers provide interoperability with Cloud software, virtualization libraries and OS-level monitoring tools [44].

3.1.4 Virtualization, secure service provisioning and composition

CSPs typically use virtualization technologies, which decouple application services (SaaS) from the underlying infrastructure (IaaS and PaaS). This approach gives rise to the notion of multi-tenancy that refers to the sharing of physical devices and virtualized resources between multiple independent users. However, in multi-tenancy an attacker having access to a virtual machine (VM) deployed on a given physical machine could compromise other VMs hosted on the same physical machine.

Another dimension of using virtualization technology in Cloud computing requires service providers and service integrators to collaborate in order to provide newly composed services to customers [66]. This sort of activity requires automatic service provisioning and composition frameworks that allow CSPs and service integrators to describe services with unified standards to introduce their functionalities, discover existing interoperable services, and securely integrate them to provide services. Such frameworks must include a declarative language to describe services, features, and mechanisms to provision and compose appropriate services [66]. The Open Services Gateway Initiative's service platform provides an open, common architecture for service providers, developers, software vendors, gateway operators, and equipment vendors to cooperatively develop, deploy, and manage services [59]. The challenges of such collaboration systems include dynamic access control to resources shared by agents and controlling collaborative actions that are geared towards a collaboration goal [66].

3.1.5 Data security, privacy and protection

Sensitivity of information: Cloud computing provides a wide range of services to store users' data (as discussed in Chapter 2). In this section, sensitive information mainly refers to the following:

- **Personally identifiable information (PII):** any information that could be used to identify an individual (e.g., name, address, credit card number, Internet Protocol address).
- **Sensitive information:** any private information (e.g., personal financial information, religion, health, etc.) or business secret.
- **Usage data:** behavioral information of data usage (e.g., recently visited history).

However, storing sensitive information (that is confidential or valuable to the user such as a company asset) in the Cloud could increase risks since there is a possibility that the Cloud platform or service provider could exploit this information (for example by sharing it with user's competitors). These risks further increase due to the lack of users' control over Cloud resources.

Users typically have limited control over how their data and applications are managed in the Cloud. Due to the loss of control for the procedure of sensitive data transportation, there could arise risks when a Cloud transaction is performed. In case of a malicious environment, this could result in data exposure to third parties or the Cloud provider itself.

One solution to this issue is to leverage data containers[9], which ensures that each user has more control over its deployed resources and data. Driven by the excellent improvements in security and performance aspects, containers are becoming more and more popular as an part of Cloud computing infrastructures among CSPs [69] (e.g., Microsoft, Google, Amazon Web Services).

Another approach, proposed in [60], consists in having usable, transparent data processing in the Cloud. This solution enables users to define transparency policies over their data [44]. They identify the requirements for transparent policy management in the Cloud based on two aspects: user demands and legal aspects of transparent data processing [44].

Confidentiality, Integrity and Availability: Data protection is one of the major concerns of Cloud computing users. As a part of personal data protection, data security is a fundamental principle of all activities related to the field of data protection, which refers to the confidentiality, integrity and availability of data.

For the Cloud users, every authorized person should be able to access user's data in the Cloud from anywhere. However, in some cases, Cloud users and service providers may manipulate data simultaneously. Consequently, it is necessary to guarantee data integrity to protect all the data in Cloud. The availability of data means that no matter when the Cloud users access their data, all of it should be available in expected format.

[9] A data container is a container that aims to store and manage data in resource-isolated processes

Encryption is a core mechanism for maintaining the confidentiality of all data, whether it consists of business, personal or sensitive information, and it could also be used to establish the integrity of various transactions, code and data [67]. Encryption is considered as a primary security control, especially for maintaining confidentiality and integrity. The uses of encryption in accessing services in the Cloud are similar to data protection in conventional technologies. Many publicly offered services are provided via an HTTPS-protocol connection to a web service, which is based on the concept of Transport Layer Security or Secure Socket Layer protection mechanisms.

For the availability, frequent outages reported on well-known CSPs increase users' concerns [44] [61]. A number of solutions have been proposed to address this issue. For instance, [62] proposed CloudProof as a secure storage system to guarantee confidentiality, integrity and write-serializability using verifiable proofs of violation by external third parties. Fuzzy authorization for Cloud storage [63] is another flexible and scalable approach to enable data to be shared securely among Cloud participants. This approach ensures confidentiality, integrity and secure access control by utilizing secret sharing schemes for users with smartphones who are using the Cloud services [44].

Data transfers to different locations: In Cloud computing, data location plays an import role to improve security. Nevertheless, one of the top security threats concerns location transparency. Due to the highly distributed (and global) nature of Cloud infrastructures, users' data may be stored on datacenters that are geographically located in multiple legal jurisdictions. On the one hand, this improves fault tolerance and availability of data and, on the other hand, poses regulatory risks. Data location should be transparent to customers, as CSPs usually do not usually specify where clients' data is stored. The provisions within current data protection regulations might affect and violate in some regions, without knowing the specific location of datacenter.

Privacy protection: Security issues in the Cloud could also lead to a number of privacy concerns. For instance, PII stored and processed in the Cloud could be prone to adversarial attacks revealing sensitive information.

Cloud computing users (e.g., companies) need to ensure that privacy commitments made to their stakeholders (customers, employees, third parties, etc.) are enforced even if they store and process relevant stakeholder data using Cloud services. In fact, privacy and data protection regulations also apply in specific jurisdictions across the world (e.g., EU's GDPR) [64]. Such regulations often expand the definition of personal data protection to cover any information related to the people who are the subjects of the data, irrespective of whether the information is private, public or professional in nature [44]. Regulations (among others) also include definitions of new roles related to handling data and propose restricting the transfer of data to third party countries that do not guarantee adequate levels of protection [44].

3.2 Internet of Things

The exponentially increasing amount of data that is being collected using IoT devices and rapidly growing number of such connected devices has underlined the importance of addressing security, privacy and data protection issues [70]. This need is exacerbated because of the high impact of potential failures and attacks. A critical failure for instance could lead to user dissatisfaction, monetary loss and even loss of life.

Note that IoT systems generate, process and exchange large amounts of security-critical and privacy-sensitive data, making them attractive targets of attacks. A given breach could potentially result in data protection risks and direct violations of regulations such as the GDPR [7]. This section describes the state-of-the-art security, privacy and data protection challenges and solutions that are being studied by the scientific community. For the sake of clarity, according to [7], the state-of-the-art is presented following the layers within an IoT architecture, as previously explained in Chapter 1 involving sensing and actuation, transmission and communication, processing and storage, and finally the application layer.

Several security concepts are closely related to privacy and data protection solutions and are therefore studied in this white paper. Table 5 – proposed originally in [71] – summarizes seven main requirements and properties of a secure network-connected device.

Property	Description
Hardware-based Root of Trust	Unforgeable cryptographic keys generated and protected by hardware. Physical countermeasures resist side-channel attacks.
	Does the device have a unique, unforgeable identity that is inseparable from the hardware?
Trusted Computing Base	Private keys stored in a hardware-protected vault, inaccessible to software. Division of software into self-protecting layers.
	Is most of the device's software outside the device's trusted computing base?
Defense in Depth	Multiple mitigations applied against each threat. Countermeasures mitigate the consequences of a successful attack on any one vector
	Is the device still protected if the security of one layer of device software is breached?
Compartmentalization	Hardware-enforced barriers between software components prevent a breach in one from propagating to others.
	Does a failure in one component of the device require a reboot of the entire device to return to operation?
Certificate-based Authentication	Signed certificate, proven by unforgeable cryptographic key, proves the device identity and authenticity.
	Does the device use certificates instead of passwords for authentication?
Renewable Security	Renewal brings the device forward to a secure state and revokes compromised assets for known vulnerabilities or security breaches.
	Is the device's software updated automatically?
Failure Reporting	A software failure, such as a buffer overrun induced by an attacker probing security, is reported to Cloud-based failure analysis system.
	Does the device report failures to its manufacturer?

Table 5: Security properties of secure network connected devices [71]

3.2.1 Sensing and actuation

The perceptual layer of sensing and actuation forms the basis or root of trust in an IoT ecosystem. This layer enables interaction between the physical and digital worlds, and given the increasing integration of everyday objects and traditional technologies with IoT, security, privacy and data protection in this IoT layer has become of paramount importance. The state-of-the-art security, privacy and data protection approaches related to this IoT layer are outlined in this section, particularly along the following lines:

- Hardware and security of IoT devices – hardware forms the basis and root for trust in IoT infrastructure.
- Integrity of IoT devices – privacy and personal data protection are at the core of IoT and the integrity of the connected devices.
- PII protection and location information and communication privacy – form the main challenge hindering the development of trust within the technology.

Hardware and security of IoT devices: The relationship between hardware and security, particularly in the context of IoT devices, is bidirectional. On the one hand, it is important to ensure the security of the hardware, and on the other hand, several hardware features could be used to increase the security of IoT devices and the overall system.

Firstly, though strong and secure hardware is acknowledged as a fundamental requirement to ensure IoT security, recent works have shown the possibility of **hardware-level trojans** through malicious components or instruction sequences, that when triggered bypasses security guarantees. One well-known example of attacks discussed in [72] and [73] is the **fabrication-time attacks**. Such attacks work by adding a cell as a capacitor to the processor's blueprint, a tiny electric charge can be stored each time a malicious program is run by a user or from a website. The more the command is executed, the more energy the capacitor stores, until it reaches a predetermined threshold. Once the threshold is met, the capacitor discharges its energy, switching on a logical function deploying a malicious payload giving a hacker elevated privileges or root access to the device. Since IoT components are often fabricated by third country manufactures and given that IoT devices typically have simple sensors, an attack like the fabrication-time attack could have severe impact.

Secondly, since hardware forms the root of trust, leveraging certain capabilities of the hardware (of IoT devices) to improve security, it has also received research attention [74]. In the seven properties of highly secure devices recently discussed in [71], two properties directly concern hardware security techniques, namely a **hardware root of trust** and **hardware supported software isolation**. The ideas of using hardware mechanisms to securely store cryptographic keys for example using **trusted platform modules** or one-time fuses and to create isolation units like Memory Management Units (MMU) and Software Guard Extensions (SGX) enclaves are similar to those in classic information technology research [11]. However, many challenges arise when applying these methods of hardware security to IoT systems, mainly due to the limited energy and computational capabilities of most IoT devices. These limitations can affect higher-layer security primitives. For instance, some IoT devices may not have very precise real-time clocks making it harder to implement even the most basic of network security protocols that assume the presence of reliable clocks [11]. To address this issue, an approach to utilize the natural decay rate of Static Random Access Memory **(SRAM)** as a **timekeeper** for embedded devices without clocks (e.g., smart cards) is presented in [75].

Integrity of IoT devices: To address data protection and privacy challenges, it is important to ensure that sensitive data does not leak due to unauthorized manipulation or handling of hardware and software in IoT devices [74]. For example, an intruder can reprogram a surveillance camera in a smart city in such a way that it sends data not only to the legitimate server, but also to the intruder. This implies that robustness and tamper-resistance of IoT devices that gather sensitive data are especially important. Recent years have witnessed research efforts to address these challenges by means of device integrity validations, tamper-resistant modules and trusted execution environments [76].

PII protection and location information and communication privacy: An interesting line of research works focuses on addressing challenges related to:

- **Non-identifiability** in order to protect the identification of the exact nature of the device. For instance, [77] recommends adding randomness or noise, having synchronous CPUs and blind values in calculations for privacy.
- **Protecting the personal information** in case of device theft/loss and resilience-related challenges, specifically by avoiding side channel attacks. In this context, [78] proposed use of QR codes (Quick Response Code).
- The **location privacy** of device holders. For instance, [79] uses algorithms such as **multi-routing random** walk to address location privacy in IoT and mainly wireless devices such as wireless sensors.

3.2.2 Transmission

This IoT layer is responsible for transmitting the data collected by the physical (sensing and actuation) layer to different devices (e.g., IoT devices and hubs, gateway to Cloud). It incorporates a diverse range of connectivity protocols such as Bluetooth Low Energy (BLE), Near Field Communication (NFC), Transmission Control Protocol (TCP) and Universal Unique Identifier (UUID). Since each protocol has its own notion of how devices communicate and exchange data, a functional and security challenge consists in managing the network layer. In fact, existing works such as [71] and [80] highlight the negative impact of the lack of standardized protocols and regulations.

To gain a deeper insight on the transmission layer challenges, consider the report on the seven properties of highly secure devices [71]. The report highlights that a common IoT system architecture connects multiple devices, usually with different communication protocols, to a hub. However, given that WiFi is becoming pervasive, a number of IoT devices are supporting only communication protocols related to WiFi and the system architecture is treating WiFi routers as IoT hubs. This configuration poses new security challenges that WiFi was not designed to support [11]. For example, it is very difficult to ensure that only a WiFi-enabled presence detector should affect a door lock. Such an isolation boundary is useful because there could be multiple devices on a network, some of which might be malicious or compromised through bugs. The isolation unit would serve as defense in depth against such a situation [11].

Software updates are essential for bugs and new security updates, however, drawbacks or challenges as some IoT devices are not designed with clear update channels, a central hub like a WiFi router can be in a good position to apply updates in the form of filters for known malicious traffic patterns. The design of a WiFi home hub that can perform such security functions is discussed in [81].

Additionally, as defense in depth, detecting misbehaving devices on the network is a common and well-deployed security practice in many computing areas known as anomaly detection. The main challenge in obtaining useful results from anomaly detectors is tuning it to produce a low number of errors, either raising a flag for benign behavior or not raising a flag for malicious behavior [11]. This challenge arises due to the fundamental complexity of the devices we typically connect to a network such as personal laptops, mobile phones, desktops, and servers. These devices perform multiple functions, and lead to complicated network traces that make it difficult to characterize “normal” behavior. In contrast, IoT devices are simple and have a single purpose. This can translate to simpler network dynamics, and hence easier to model behaviors ultimately leading to a lower number of errors in anomaly detectors [11]. Recent work in the context of industrial control systems show promising results. For example, [82] shows how predictable network characteristics of relays and circuit breakers can be used to reliably fingerprint them.

Finally, the most common approach to ensure data confidentiality during the transmission of the data consists in applying encryption techniques. Some encryption schemes allow securely adding data to network packets, which

provides a way to introduce sequence numbers, timestamping, IPsec parameter index etc. This appended data could be used in multiple ways such as analysis of traffic flow, audit of protocol's results, to name a few. Several studies have employed security and transmission protocols to address privacy and data protection challenges in the IoT context:

- **Secure Communication Protocol (SCP)** could be the suitable approach according to [79].
- According to [83] during the communication pseudonyms can be replaced for encryption in case it is not feasible to the devices identity or users in order to decrease the vulnerability.
- Another widely used approach is the **Temporary Mobile Subscriber Identity (TMSI)**. This approach enforces devices to initiate communication if and only if when there is a need to, hence derogating privacy disclosure induced by communication.
- In 3GPP machine type communications, devices detach from the network after a certain period of inactivity in order to avoid unnecessary collection of location information by the network [78] [83].

3.2.3 Storage and processing

For protecting privacy of information in storage, only least possible amount of information needed should be stored and in the case where it is mandatory to store personal information, data within IoT devices should conceal the real identity tied with it using de-identification techniques (e.g., pseudonymization, anonymization) described in Section 3.3. The approach presented in [77] consists of adding random noise to databases for differential privacy to preserve individual privacy. This formal mathematical approach allows quantifying the degree of individual privacy in a statistical database.

For protecting privacy while processing, according to [83], two approaches could be adopted. First, personal data must be treated while respecting the intended purpose of data collection. Second, without explicit acceptance and the knowledge of the data owner, their personal data should not be disclosed or retained to third parties. This correlates to the requirements set in the GDPR.

In addition to the above, the storage and processing layer of the IoT system also builds on the underlying hardware to establish trust and isolation. This IoT layer consists of firmware, OS code and privileged system application frameworks, and employs several privacy enhancing methods including:

- **Process isolation** is a set of different hardware and software technologies designed to protect each process from other processes on the operating system. To overcome the limitations of most IoT devices, [84] discusses process isolation without the need for MMU where language-based isolation features are explored to provide process isolation abstraction.
- **Access Control** is a security technique that can be used to regulate who or what can view or use resources in a computing environment by limiting connections to computer networks, system files and data [71] [85]. Access control is like a gatekeeper – once code obtains access to sensitive resources, access control does not provide any further protection [76].
- **Information Flow Control (IFC)** extends on access control as it is concerned with the control of how information is propagated by computing systems. Classically, access control was the main means of preventing information from being disseminated; however, as the name implies, access control verifies that the program's access rights at the point of access, and either grants or denies the program access.

  Once the program has been given access to information no further effort is made to make sure that the program handles the accessed information correctly. In other words, access control forces an all-or-nothing choice of either fully trusting the program not to compromise information or not allowing access to this information altogether. On the other hand, information flow control tracks how information propagates through the program during execution to make sure that the program handles the information securely.

[76] analyzed a set of smart home platforms, as a common IoT scenario, and found that current platforms only use access control. IFC is a promising technique to control how (untrusted) code uses its access to sensitive resources. Although IFC is not a new concept, as evidenced by the multitude of proposed systems for various domains, the challenge lies in applying it meaningfully to a specific domain [11]. For example, FlowFence is a recent proposal for consumer IoT frameworks that enables a data-flow-graph approach to IFC due to the structure of IoT apps [11] [76]. Furthermore, confidentiality properties for environments such as homes are well studied; however, integrity properties, which are important in IoT, are less well studied [11] [71] [86].

- **Authentication** is one of the more important security methods covered in literature. Passwords are currently the most widely used mechanism to authenticate users to their IoT devices but they are also a major point of concern due to weak passwords. Recent large DoS attacks were facilitated through such weak passwords, an alternative would be activity-based biometrics alternative, however, IoT devices tend to be limited in I/O in turn limiting authentication method [87] [76]. For protecting privacy of information in storage, only the least possible amount of information needed should be stored and in the case where it is mandatory to store personal information, data within IoT devices should conceal the real identity tied with it through pseudonymization and anonymization techniques as described in Section 3.3.

3.2.4 Application

The final and the topmost layer in an IoT system comprises the end user applications. This application layer is similar to other computing paradigms – it simply runs customized code for end-user scenarios. In this section, we consider two ways based on [11] in which IoT application behavior can affect security, privacy and data protection: i) physical co-relations and ii) machine learning and control of physical processes.

Physical co-relations: The behavior of physical devices could be continuously monitored and could serve as a feedback channel to IoT platforms. For instance, physical processes could be monitored for deviations from expected behavior, and accordingly potential system failures or security issues could be determined. The study [81] models a simple If-This-Then-That rule that closes a garage door, and uses the acoustic pattern for a specific amount of time when the door closes as a feedback channel, in order to detect deviations.

Machine Learning (ML) and control of physical processes. In recent years, machine learning (and deep learning) found wide applicability in many domains of computing [85]. However, [82] explains that as more physical processes come under the control of machine learning algorithms, their vulnerabilities in adversarial settings will become pressing security and safety issues. They show that classic IT security has often applied ML to security problems (e.g., malware detection), however, only recently has work begun on securing the ML algorithms themselves [11]. Therefore, building robustness into ML algorithms against such attacks is an active area of research [88].

Recent research work has shown that deep learning algorithms are susceptible to adversarial manipulations of their inputs – attackers can design inputs that look indistinguishable from benign inputs to humans but can be interpreted in a completely different way by machines [11]. For example, tampered images that are fed into a vision algorithm running on an autonomous vehicle can make the vehicle believe a stop sign was a yield sign, causing a possible crash at an intersection. With the ever-rising number of deployed UAVs in cities, the threat is greater as these drones fly at very low altitude and in highly populated areas, multiplying the risks. Therefore, building robustness into ML algorithms against such attacks is an active area of research [88].

3.3 Big data

While the Big data paradigm provides many advantages such as new types of digital services and discovery of hidden patterns and insights, privacy challenges have also increased. The privacy of an individual could be at risk at several stages: from the point when data is created and released to another party, to the time when the data would be analyzed and processed. Several research efforts have focused on discovering the security and privacy challenges of Big data and proposing privacy preserving techniques as solutions. The purpose of this section is to describe security, privacy and data protection challenges in different stages of Big data cycle and to summarize corresponding solutions proposed in the literature. In particular, the three main layers considered are:

- Data collection layer;
- Data storage layer;
- Data analysis and AI layer.

Table 6 shows an overview of the privacy challenges and existing potential solutions for each Big data layer. These challenges and techniques are discussed in detail in the following subsections.

Big data phase	Privacy Challenge	Potential solutions
Collection	• Protecting individuals' sensitive information • Control over the sensitive information	• Access Restrictions • Data Falsification • K-anonymity • L-diversity • T-closeness
Storage	• Protecting individuals' sensitive information from compromising storage system • Privacy protection in distributed databases • Data sharing with multi parties	• Anonymization • Attribute Based Encryption • Homomorphic encryption
Processing	• Extracting meaningful information of the data without violating the privacy of the individuals • Ensure the efficiency	• Homomorphic encryption • Differential privacy • Association Rule Mining • Classification, Clustering

Table 6: Summary of privacy challenges and potential solutions for each Big data layer

3.3.1 Data collection

Sensitive data breach is a potential attack while providing and collecting Big data from various sources. Data could be produced by the activities of a user through either an online service (i.e., online social networks, Internet browsing) or by an IoT device (i.e., smart watch, wireless sensor networks) which monitor the environment for any activities. In another way, data could be provided by the data owner itself who gives the information to a third party such as a patient who is registering to e.g., an e-health system of a healthcare institute and providing his personal information to benefit from the services. In the literature [89], the first model of data generation is called passive data generation and the latter is active data generation.

In both cases, privacy-preserving techniques are required to reduce the private information before releasing a dataset to other parties for storage and also for data processing and analysis. Data protection in this level also consists of hardware protection such as for sensors that are monitoring the data of a user.

The simplest solution might be to not disclose the information; however, this would hinder the process of data analysis, which is the main purpose of collecting Big data. Hence, the main concern is to assure the high utility of data for the processing and analysis phases while it reveals no sensitive information.

Online services and browsing are one of the threats for an individual's sensitive information that might be accessed unwillingly by other parties who are not authorized to have the access for that data. There exists some tools and solutions for protecting the data that one could benefit from. These solutions are classified as access restriction and data falsification methods and tools. In the following, these techniques and popular tools that are available online as a solution are described.

- **Access Restriction:** Limiting the access to data is one of the main solutions for protecting sensitive information of an individual, which is not supposed to be shared. This way the sensitive data will be safe and covered. There are different tools and techniques to apply the access restriction solution on data. Some of these tools are as follows [89]:
  - **Anti-tracking extensions**. There exist some extensions to block the tracking of users' movements on the Internet. Some of the popular anti-tracking extensions are uBlock Origin, Disconnect, and Privacy Badger.
  - **Advertisement/script blockers**. There are other extensions to block the advertisements of websites and prevent scripts from sending the information about users. Adblock Plus, Brave Browser, and NoScript are some of the extensions for blocking the advertisements and scripts.
  - **Encryption tools**. These tools assure a private and secure online surfing. Tor Browser is a tool that provides high levels of privacy for online communications by using the Tor network.
  - **Anti-malware and anti-virus software**. There are various anti-virus software that could protect the communication and block malicious activities.

  Benefiting from the abovementioned security tools, individuals' data could be protected from unauthorized access hence, the access is restricted.
- **Data falsification:** Manipulating research data with the intention of giving a false impression. In this case, before sending the data to any third party, the data is distorted using a wide range of tools. Different operations can be done during this process such as manipulating images (e.g. micrographs, gels, and radiological images), removing outliers or "inconvenient" results, changing, adding or omitting data points, etc. Hence, if the data is distorted, the true information could be protected against being published easily [90]:
  - **Creating fake identity**. This approach is based on deceiving the adversary by providing wrong, however, plausible-looking information. By this method, some fake data of a user's identity is produced along with the user's true information. Hence, an attacker cannot recognize the true identity which buried under the fake data. One of the popular techniques in this area is called Honey Encryption [91],

which is widely used as a privacy preserving solution in different areas, such as protecting credit card information, a secure storage of genomic data [92], and natural language processing [93].

- **MaskMe** is one of the tools that is beneficial to hide individual's identity and sensitive data such as bank account information through an online shopping.
- **Socketpuppet** is a fake identity created to promote someone or something through blogs, wikis, forums or social networking sites. Benefiting from Socketpuppet each specific individual's data will be considered as the data of different people. In this case as the third party does not have enough knowledge, relating different Socketpuppets to an individual becomes challenging and hence, the data is expected to remain safe and secure.

The abovementioned methods are useful to protect individuals' security and privacy; however, they are not sufficient by themselves when it comes to sharing Big data of users' information for data analysis. A dataset, which is gathered by a large number of users' personal information for a specific purpose, i.e., personal health records, if attacked by an adversary could reveal sensitive information of an individual. On the other hand, this data has a great value in the case of data analysis in order to provide various health services i.e., fast and precise diagnosis and treatment for patients. To address the challenge of publishing data while preserving the privacy of the data owner is usually referred as privacy-preserving data publishing (PPDP). [94] have discussed a systematic overview over different approaches to PPDP along with the practical challenges and solutions. Moreover, in another study [95] the authors provide an overview on how the data owner can modify the data and later how this modified data could protect sensitive information and preserve privacy of individuals.

In a more specific study, [96] has focused on providing an efficient solution for the privacy-utility tradeoff in data publishing by addressing the challenge of high-dimension data, high-correlation data and powerful attack models. In [97], personalized anonymity is proposed for data publishing to meet the personalized privacy preservation requirement.

De-identification: One of the first beneficial techniques for privacy-preserving data mining is de-identification mechanisms [98] [99], in which the purpose is to encrypt or remove the PII from the datasets in order to make the respected individuals anonymous. As one of the essential tools that could be converted to privacy-preserving Big data analytics is the de-identification technique. The two main functions as the basic tools for de-identification are **generalization** (replacing the data with a less precise information [100]) and **suppression** (removing the identifiers from the data or replacing them with tags [101]), that are applied to data before publishing it.

Pseudonymization is a data management and de-identification procedure by which artificial identifiers replace PII of a data record. A single pseudonym for each replaced field or collection of replaced fields makes the data record less identifiable while remaining suitable for data analysis and data processing [98].

- **K-anonymity:** A release of data is said to have the k-anonymity property if the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appear in the release [102] [103] [104].
- **L-diversity** is a form of group-based anonymization that is used to preserve privacy in data sets by reducing the granularity of a data representation. This reduction is a trade-off that results in some loss of effectiveness of data management or mining algorithms to gain some privacy. The L-diversity model is an extension of the k-anonymity model, which reduces the granularity of data representation using techniques including generalization and suppression such that any given record maps onto at least k-1 other records in the data [102] [104]. However, a knowledge regarding the distribution of an attribute could reveal sensitive information about the data. The next method is defined to address this limitation.

- **T-closeness** is based on decreasing the granularity of sensitive data. A set possess T-closeness, if the distance between the sensitive attributes is smaller or equal to a threshold T. Although this reduction might cause some loss of result in data mining algorithms, it is a trade-off to gain privacy [105].

An attacker could gain further information in Big data in order to help for de-identification, hence, Big data could increase the possibility of re-identification as well. Therefore, using de-identification techniques by themselves are not enough to protect data from privacy attacks. Table 7 provides a simple analysis of pseudonymization techniques.

Privacy enhancing techniques	Advantages [+] and disadvantages [-]	Applications and use cases
K-anonymity	+ simple definition + a lot of existing algorithms - the whole data should be presented and considered - limited to the cases where each record presents an individual	• Wireless sensor networks [106] • E-Health [107] • Cloud [108] • Location based services [109]
L-diversity	+ the diversity of the sensitive information is taken into consideration - does not consider the distribution of the sensitive values	• E-Health [107] • Location based services [110]
T-closeness	+ the distribution of the sensitive values is considered - increasing the privacy leads to loss of effectiveness of data analysis algorithms	• Location based services [111]

Table 7: Pseudonymization techniques, their advantages and limitations, and example use cases

3.3.2 Data storage

Cloud computing is one of the main platforms for Big data storage, as discussed in Sections 1.1, 2.2 and 3.1. In this section, the focus is on the techniques that have been widely used for security, privacy and data protection of Big data. The main techniques, which are used to ensure privacy of Big data storage, are mostly based on encryption methods. Benefiting from these procedures, a data sender encrypts the data by using a public key (and private key) to transfer the data. The receiver, on the other hand could only decrypt the data with a valid key. Encryption based methods to increase the security, privacy, and data protection could be classified as Attribute Based Encryption (ABE), Homomorphic Encryption (HE) and Identity Based Encryption (IBE). Table 8 provides an overview of these encryption techniques.

- **Attribute Based Encryption:** ABE is an encryption technique that ensures end-to-end Big data privacy in a Cloud storage system. In ABE, access policies are defined by data owner and data are encrypted under those policies [112].
- **Homomorphic Encryption:** Operations can run over encrypted data while protecting individual's privacy. This method allows a third party such as Cloud, to perform operations on encrypted data without learning any information about the data [113].
- **Identity Based Encryption:** IBE is an alternative to public-key encryption which is proposed to simplify key management in a certificate-based public key infrastructure (PKI) by using human identities like email address or IP address as public keys [89].

Privacy techniques	Advantages [+] and disadvantages [-]	Applications and use cases
Attribute Based Encryption	+ Reduce the communication overhead of the Internet - Technical implementation efforts	• Broadcast encryption [114]
Homomorphic Encryption	+ Privacy guaranteed on the third party (i.e. Cloud) + Data can be decrypted less often - Slow procedures over cipher text - Does not verify the computation	• Bitcoin [115] • E-health [116]
Identity Based Encryption	+ Less vulnerable to spam + The keys expires and there is no need to revoke them - Needs a centralized server	• Vehicle network and distributed systems [117]

Table 8: Overview of privacy preserving techniques for Big data storage

3.3.3 Data analytics

The key component of Big data analytics is Big data processing, as it mines new knowledge for economic growth and technical innovation. Because Big data processing efficiency is an important measure for the success of the Big data paradigm, the privacy requirements of Big data processing become more challenging. On the one hand, privacy should not be sacrificed for a big efficiency, and on the other hand, the purpose is not only to protect privacy but also to ensure efficiency at the same time.

The main privacy preserving consideration during Big data analytics could be discussed in two general stages [118]:

- The first stage is the protection of sensitive raw data, which should be trimmed out and modified from the main database; various relevant techniques have been discussed in Section 3.3.1.
- The second stage considers extracting knowledge from the raw data benefiting from machine learning and AI algorithms that might contain sensitive information of a user without violating the privacy of individuals. These techniques are classified as Privacy-Preserving Data Mining (PPDM), which are designed to ensure a specific level of privacy while allowing for an effective data analysis by maximizing the utility of the data.

The term PPDM was introduced in [119] and [120] in the context of privacy in partitioned dataset across different private enterprises. In particular, [119] proposed a cryptographic protocol for decision tree algorithm over a partitioned dataset. Although, later on, following the same methodology, they improve the protocol [121] by simplifying it and by making it more efficient as compared to [119]. On the other hand, [120] introduced randomization approach that is now one of the typical algorithms among the perturbation techniques in centralized data analysis while maximizing the data protection from any the disclosure of users' valuable information. These studies are seminal works and have been studied by various PPDM research works.

There are several categorizes for PPDM approaches that divide the techniques based on different dimensions [122] such as data distribution (centralized or distributed data models), data modification (i.e. perturbation, blocking), data mining, privacy preserving techniques (i.e., heuristic-based techniques, cryptography-based techniques), etc.

4

Technical standardization – Data protection and privacy in Smart ICT

4. Technical standardization – Data protection and privacy in Smart ICT

The rapid technological advancements in Smart ICT, their widespread adoption, and the constantly evolving data protection and privacy landscape (e.g., introduction of new regulations) have resulted in a huge demand for careful study and development of relevant technical standards. For instance, a technical standard defining privacy framework is necessary to not only establish a common language but also to clarify the responsibilities and approaches for ensuring privacy protection.

Standardization bodies at international as well as European levels have initiated a range of activities that could increase market confidence in Smart ICT and relevant data protection and privacy topics. This chapter provides an overview of various developments in the areas of technical standardization. After providing some background about technical standardization (Section 4.1), including an overview of standards developing organizations (SDOs) as well as their working principles, Section 4.2 provides details about technical committees that focus on security, privacy and data protection, and outlines their most relevant projects. Finally, Sections 4.3.1- 4.3.3 provide details about standardization activities in Cloud computing, IoT and Big data respectively, with an emphasis on security, privacy and data protection projects.

4.1 Background on technical standardization

Technical standardization is a long-standing tool that is widely recognized for its ability to provide technical or qualitative referential for products, services or processes [123]. Technical standards are developed by organizations that bring all interested stakeholders together and that follow well-accepted principles (e.g., defined by the World Trade Organization) as described in Section 4.1.2.

- At the **international level**, the three recognized standardization organizations are the:
  - International Organization for Standardization (ISO).
  - International Electrotechnical Commission (IEC).
  - International Telecommunication Union's Telecommunication Standardization Sector (ITU-T).
- In the **EU**, the three recognized European Standardization Organizations are the:
  - European Committee for Standardization (CEN).
  - European Committee for Electrotechnical Standardization (CENELEC).
  - European Telecommunications Standards Institute (ETSI).

Finally, **at national level**, countries have national standards body (NSB) that protects the country's interests in international (and European) standardization organizations, where applicable. In **Luxembourg, ILNAS** is the NSB and is a member of the following European and international standardization organizations: CEN, CENELEC, ETSI, ISO, IEC and ITU-T.

4.1.1 Cooperation between Standards Developing Organizations (SDOs)

To ensure transparency and to avoid duplication of standards, agreements have been established between international and European SDOs [123]. Main objectives of these agreements are to provide a: i) framework for optimal use of resources and expertise available for standardization work, and ii) mechanism for information exchange between international and European standardization organizations (ESOs) to increase the transparency of ongoing work at international and European levels [124]. For instance, ISO and CEN signed the Vienna Agreement[10] in 1990 based on the following guiding principles:

- Primacy of international standards and implementation of ISO Standards at European level (EN ISO);
- Work at European level (CEN), if there is a lack of interest at international level (ISO);
- When a given project undergoes parallel development, procedures are in place ensuring standardization documents of common interest are approved by both (ISO and CEN) organizations.

Similarly, CENELEC and IEC signed the Dresden Agreement[11] in 1996 with the aim of developing intensive consultations in the electrotechnical field. The Frankfurt Agreement signed in 2016 replaced this agreement, and has simplified the parallel voting processes and increased the traceability of international standards adopted in Europe, thanks to a new referencing system. These agreements are signed on the following guiding principles:

- Development of all new standardization projects within IEC (to the extent possible);
- Work at European level (CENELEC), if there is no interest at international level (IEC);
- When a given project undergoes parallel development, ballots for relevant standardization documents are organized simultaneously at both (IEC and CENELEC) organizations.

Under these agreements, 33% of all European standards ratified by CEN and 72% of standards ratified by CENELEC are identical to ISO or IEC standards respectively[12], indicating that European and international SDOs avoid duplicate work.

Several other agreements that aim at improving the standards development processes, cooperation between SDOs, and wider recognition of standards have been reached. Some of these agreements are listed here as examples:

- ITU-T and ETSI have signed a Memorandum of Understanding (MoU)[13] that enhances cooperation. Moreover, it paves way for **standards** developed by **ETSI** to be **recognized internationally**.
- ISO and IEC have the possibility to sign conventions to create Joint Technical Committees (JTC) or Joint Project Committees (JPC) when the area of work overlap their competences. One of the most prominent Joint Technical Committee created in this framework is the **ISO/IEC JTC 1** on Information Technology[14].
- ISO, IEC and ITU have established the World Standards Cooperation in 2001, a high-level collaboration system intending to strengthen and advance the voluntary consensus-based international standards system and to resolve issues related to the technical cooperation between the three organizations[15]. Similarly, the cooperation between CEN and CENELEC aims to create a European standardization system that is open, flexible and dynamic.

[10] Agreement on technical co-operation between ISO and CEN (Vienna Agreement). Accessible at: https://isotc.iso.org/livelink/livelink/fetch/2000/2122/3146825/4229629/4230450/4230458/01__Agreement_on_Technical_Cooperation_between_ISO_and_CEN_%28Vienna_Agreement%29.pdf?nodeid=4230688&vernum=-2

[11] IEC-CENELEC Agreement on Common planning of new work and parallel voting (Frankfurt Agreement). Available at: ftp://ftp.cencenelec.eu/CENELEC/Guides/CLC/13_CENELECGuide13.pdf

[12] CEN-CENELEC Quarterly Statistical Pack (accessed for 2018 Q2). Available at: https://www.cencenelec.eu/aboutus/InFigures/Pages/default.aspx

[13] https://www.itu.int/en/ITU-T/extcoop/Documents/mou/MoU-ETSI-ITU-201605.pdf

[14] https://www.iso.org/committee/45020.html

[15] http://www.worldstandardscooperation.org/

4.1.2 Objectives and principles for developing technical standards

As stated in the Regulation (EU) N°1025/2012 on European standardization, and according to the World Trade Organization (WTO)[16], formal standards bodies follow a set of fundamental **principles for developing technical standards** [123]:

- **Transparency:** All essential information regarding current work programs as well as on proposals for standards, guides and recommendations under consideration and on the results should be made easily accessible to all interested parties.
- **Openness:** Membership of an international standards body should be open on a non-discriminatory basis to relevant bodies.
- **Impartiality and Consensus:** All relevant bodies should be provided with meaningful opportunities to contribute to the elaboration of an international standard so that the standard development process will not give privilege to, or favor the interests of, a particular supplier, country or region. Consensus procedures that seek to take into account the views of all parties concerned and to reconcile any conflicting arguments should be established.
- **Effectiveness and Relevance:** International standards need to be relevant and to effectively respond to regulatory and market needs as well as scientific and technological developments in various countries. They should not distort the global market, have adverse effects on fair competition, or stifle innovation and technological development. In addition, they should not give preference to the characteristics or requirements of specific countries or regions when different needs or interests exist in other countries or regions. Whenever possible, international standards should be performance based rather than based on design or descriptive characteristics.
- **Coherence:** In order to avoid the development of conflicting international standards, it is important that international standards bodies avoid duplication of, or overlap with, the work of other international standards bodies. In this respect, cooperation and coordination with other relevant international bodies is essential.
- **Development dimension:** Constraints on developing countries, in particular, to effectively participate in standards development, should be taken into consideration in the standards development process. Tangible ways of facilitating developing countries participation in international standards development should be sought.

Technical standards enable its users to pursue various **objectives** such as:

- Management of the diversity;
- Convenience of use;
- Performance, quality and reliability;
- Health and safety;
- Compatibility;
- Interchangeability;
- Security;
- Environmental protection;
- Product protection;
- Mutual understanding;
- Economic performance;
- Trade;
- Etc.

[16] Source: Second triennial review of the operation and implementation of the agreement on technical barriers to trade – Annex 4: Decision of the committee on principles for the development of international standards, guides and recommendations

4.2 Overview of data protection and privacy standardization

This section first provides an overview of standardization committees that focus on security, privacy and data protection topics. Then, fundamental privacy and data protection terms from different technical standards are outlined.

4.2.1 Relevant standardization committees from different SDOs

Several international and European SDOs have established committees focusing on privacy and data protection aspects, often in association with topics related to security, due to the natural proximity of these topics. The foremost of these committees are the following:

- ISO/IEC JTC 1/SC 27 - IT Security techniques.
- ISO/PC 317 - Consumer protection: privacy by design for consumer goods and services (created in 2018).
- CEN/CLC JTC 13 - Cybersecurity and data protection.
- CEN/CLC JTC 8 - Privacy management in products and services.
- ETSI/TC CYBER - Cybersecurity.
- ITU-T SG 17 - Security.

An overview of each of these committees, highlighting Luxembourg's involvement and listing relevant projects in these committees is provided below. Considering the maturity and widespread acceptability, developments within ISO/IEC JTC 1/SC 27 are specifically discussed at a greater detail.

4.2.1.1 ISO/IEC JTC 1/SC 27

This is one of the largest committees, which has published 179 standards (including updates) and another 75 are currently under development[17]. ISO/IEC JTC 1/SC 27 has 50 Participating members including Luxembourg and 27 Observing members[18]. 31 delegates from Luxembourg are actively participating in this committee by voting and commenting on the proposals, contributing towards the drafting of standards/reports, and by participating in the National Mirror Committee (NMC) and/or plenary meetings.

Scope: This committee focuses on the development of standards for the protection of information and ICT, including generic methods, techniques and guidelines to address both security and privacy aspects such as [125]:

- Security requirements capture methodology;
- Management of information and ICT security; in particular information security management systems, security processes, and security controls and services;
- Cryptographic and other security mechanisms, including but not limited to mechanisms for protecting the accountability, availability, integrity and confidentiality of information;
- Security management support documentation including terminology, guidelines as well as procedures for the registration of security components;
- Security aspects of identity management, biometrics and privacy;

[17] At the time of writing this part – August 2018.

[18] P-members are required to participate actively in the work of the Committee by voting on all official ballots (e.g., at various stages of standards development) as well as by participating in plenary meetings. O-members on the other hand can simply observe the standards that are being developed and contribute to the work, albeit without formal obligation. O-members have less impact on voting results and cannot participate in any WG of the Committee.

- Conformance assessment, accreditation and auditing requirements in the area of information security management systems;
- Security evaluation criteria and methodology.

Structure: This committee has established several working-groups, as described below [125]:

- ISO/IEC JTC 1/SC 27/**WG 1 Information Security Management Systems (ISMS).**
  WG 1 covers all aspects of standardization related to ISMS: requirements, methods and processes, security controls, sector and application specific use of ISMS, governance, information security economics and accreditation, certification and auditing of ISMS.
- ISO/IEC JTC 1/SC 27/**WG 2 Cryptography and security mechanisms.**
  WG 2 covers both cryptographic and non-cryptographic techniques and mechanisms including confidentiality, entity authentication, non-repudiation, key management and data integrity (e.g., message authentication, hash-functions, digital signatures, etc.).
- ISO/IEC JTC 1/SC 27/**WG 3 Security evaluation testing and specification.**
  This WG covers aspects related to security engineering, with particular emphasis on, but not limited to standards for IT security specification, evaluation, testing and certification of IT systems, components, and products. The following aspects may be distinguished: security evaluation criteria, methodology for application of the criteria, security functional and assurance specification of IT systems, components and products, testing methodology for determination of security functional and assurance conformance, accreditation schemes, administrative procedures for testing, evaluation and certification.
- ISO/IEC JTC 1/SC 27/**WG 4 Security controls and services.**[19]
  WG 4 is developing and maintaining International Standards (IS), Technical Specifications (TS) and Technical Reports (TR) for information security in the area of security controls and services, to assist organizations in the implementation of the ISO/IEC 27000-series of ISMS International Standards and technical reports. Also the scope of WG 4 includes evaluating and developing International Standards for addressing existing and emerging information security issues and needs and other security aspects that resulted from the proliferation and use of ICT and Internet related technology in organizations (such as multinationals corporations, SMEs, government departments, and non-profit organizations).
- ISO/IEC JTC 1/SC 27/**WG 5 Identity management and privacy technologies.**
  This WG is responsible of the development and maintenance of standards and guidelines addressing security aspects of identity management, biometrics and privacy.
- Joint ISO/TC 307 and ISO/IEC JTC 1/SC 27 WG.

Interesting projects: ISO/IEC JTC 1/SC 27 through its WGs (specifically, WG 5) has initiated a number of projects related to **privacy**. Table 9 outlines the projects (standards) that have been published already. The projects that are currently under development are then listed below.

[19] The Convenor of ISO/IEC JTC 1/SC 27/WG 4 – Mr. Johann Amsenga is a Luxembourg delegate.

Published standards		
Identifier	**Title**	**Description (scope/abstract)**
ISO/IEC 29100	Privacy framework	This standard provides a privacy framework which [26]: • Specifies a common privacy terminology; • Defines the actors and their roles in processing personally identifiable information (PII); • Describes privacy safeguarding considerations; • Provides references to known privacy principles for information technology. ISO/IEC 29100:2011 is applicable to natural persons and organizations involved in specifying, procuring, architecting, designing, developing, testing, maintaining, administering, and operating information and communication technology systems or services where privacy controls are required for the processing of PII. This standard is freely available.
ISO/IEC 29101	Privacy architecture framework	ISO/IEC 29101:2013 [126] defines a privacy architecture framework that: • Specifies concerns for information and communication technology systems that process personally identifiable information. • Lists components for the implementation of such systems. • Provides architectural views contextualizing these components. This standard is applicable to entities involved in specifying, procuring, architecting, designing, testing, maintaining, administering and operating ICT systems that process PII. It focuses primarily on ICT systems that are designed to interact with PII principals.
ISO/IEC 29134	Guidelines for privacy impact assessment	ISO/IEC 29134:2017 [127] gives guidelines for: • A process on privacy impact assessments; • A structure and content of a PIA report. It is applicable to all types and sizes of organizations, including public companies, private companies, government entities and not-for-profit organizations. ISO/IEC 29134:2017 is relevant to those involved in designing or implementing projects, including the parties operating data processing systems and services that process PII.

Identifier	Title	Description (scope/abstract)
ISO/IEC 29151 **ITU-T X.1058**	Code of practice for personally identifiable information protection	ISO/IEC 29151:2017 [128] establishes control objectives, controls and guidelines for implementing controls, to meet the requirements identified by a risk and impact assessment related to the protection of PII. In particular, this Recommendation / International Standard specifies guidelines based on ISO/IEC 27002, taking into consideration the requirements for processing PII that may be applicable within the context of an organization's information security risk environment(s). ISO/IEC 29151:2017 is applicable to all types and sizes of organizations acting as PII controllers (as defined in ISO/IEC 29100), including public and private companies, government entities and not-for-profit organizations that process PII.
ISO/IEC 29190	Privacy capability assessment model	ISO/IEC 29190:2015 [129] provides organizations with high-level guidance about how to assess their capability to manage privacy-related processes. In particular, it: ● Specifies steps in assessing processes to determine privacy capability. ● Specifies a set of levels for privacy capability assessment. ● Provides guidance on the key process areas against which privacy capability can be assessed. ● Provides guidance for those implementing process assessment. ● Provides guidance on how to integrate the privacy capability assessment into organizations operations.
ISO/IEC 29146	A framework for access management	● This standard [130] defines and establishes a framework for access management and the secure management of the process to access information and ICT resources, associated with the accountability of a subject within some context. ● It provides concepts, terms and definitions applicable to distributed access management techniques in network environments. ● It also provides explanations about related architecture, components and management functions.
ISO/IEC 29191	Requirements for partially anonymous, partially unlinkable authentication	This International Standard [131] provides a framework and establishes requirements for partially anonymous, partially unlinkable authentication.
ISO/IEC 27018	Code of practice for protection of personally identifiable information (PII) in public Clouds acting as PII processors	See Section 4.3.1 for details.

Table 9: ISO/IEC JTC 1/SC 27 projects related to privacy

In addition to the standards listed in Table 9, JTC 1/ SC 27 has the ISO/IEC 24760 series of standards in its program of work. This series – defining a framework for identity management – includes:

- ISO/IEC 24760-1:2011 - Terminology and concepts [132] defines terms for identity management, and specifies core concepts of identity and identity management and their relationships. It is applicable to any information system that processes identity information.
- ISO/IEC 24760-2:2015 - Reference architecture and requirements [133] provides guidelines for the implementation of systems for the management of identity information, and specifies requirements for the implementation and operation of a framework for identity management. It is applicable to any information system where information relating to identity is processed or stored.
- ISO/IEC 24760-3:2016 - Practice [134] provides guidance for the management of identity information and for ensuring that an identity management system conforms to ISO/IEC 24760-1 and ISO/IEC 24760-2.

Relevant ISO/IEC JTC 1/SC 27/WG 5 projects under development:

- ISO/IEC PRF TS 19608 - Guidance for developing security and privacy functional requirements based on ISO/IEC 15408.
- ISO/IEC AWI 20547-4 - **Big data** reference architecture - Part 4: **Security and privacy.**
- ISO/IEC FDIS 20889 - Privacy enhancing data de-identification terminology and classification of techniques.
- ISO/IEC AWI 27030 - Guidelines for **security and privacy in Internet of Things** (IoT).
- ISO/IEC PDTR 27550 - Privacy engineering.
- ISO/IEC CD 27552 - Enhancement to ISO/IEC 27001 for privacy management – Requirements.
- ISO/IEC AWI TS 27570 - Privacy guidelines for smart cities.
- ISO/IEC CD 29184 - Guidelines for online privacy notices and consent.
- ISO/IEC FDIS 29101 - Privacy architecture framework.

Other ISO/IEC JTC 1/SC 27 standards: In addition to the SC 27/WG 5 projects discussed above, several technical standards developed by SC 27 are widely adopted. For instance, the standards **ISO/IEC 27001:2013** *Information technology – Security techniques – Information security management systems – Requirements* and **ISO/IEC 27002:2013** *Information technology – Security techniques – Code of practice for information security controls* are foremost and the number of organizations becoming compliant with the ISMS requirements defined in ISO/IEC 27001 are increasingly numerous.

It is important to note that SC 27 works in liaison with many other JTC 1/SCs and develops standards related to security and privacy of specific domains. For example, SC 27 has published an International Standard related to the security of Cloud computing and a new one regarding security and privacy aspects in Cloud SLAs is currently under development in liaison with ISO/IEC JTC 1/SC 38. Similarly, a standard concerning Big data security and privacy is currently under development in JTC 1/SC 27, in close collaboration with ISO/IEC JTC 1/SC 42: ISO/IEC AWI 20547-4 *Information technology – Big data reference architecture – Part 4: Security and privacy* (see Section 4.3 for details).

4.2.1.2 CEN/CLC JTC 13 Cybersecurity and data protection

CEN/CLC JTC 13 [135] was created in 2017 based on the recommendation of the CEN/CLC Cyber Security Focus Group (CSCG), which identified cybersecurity, including privacy and data protection, as an essential need to realize EU's Digital Single Market strategy.

Scope: Development of standards for data protection, information protection and security techniques with specific focus on cybersecurity covering all concurrent aspects of the evolving information society, including [135]:

- Organizational frameworks and methodologies, including IT management systems.
- Data protection and privacy guidelines – processes and products evaluation schemes.
- ICT security and physical security technical guidelines.
- Smart technology, objects, distributed computing devices, data services.

This includes identification and possible adoption of standards already available or under development, which could support the EU Digital Single Market and different standardization requests and/or EC Directives/ Regulations. If required these standards will be augmented by Technical Reports (TRs) and Technical Specifications (TSs). Special attention will be paid to ISO/IEC JTC 1 standards, but will not be limited to this. Other SDOs and international bodies will also be taken into account, such as ISO, IEC, ITU-T, IEEE, NIST or industrial fora.

For the relevant standards, different options will be considered:

- Identical adoption as a European standard using for example Vienna/Frankfurt agreements.
- Adoption as European standard with additional/complementary requirements, for example in order to fulfil European legal requirements.

In this context, during a recently held plenary meeting (July 2018), JTC 13 – contingent upon approval via ballots – has decided to adopt the following international standards unmodified:

- ISO/IEC 27006 - Requirements for bodies providing audit and certification of ISMS.
- ISO/IEC 27007 - Guidelines for information security management systems auditing.
- ISO/IEC 27008 - Guidelines for auditors on information security controls.
- ISO/IEC 27010 - Information security management for inter-sector and inter-organizational communications.
- ISO/IEC 27011 - Code of practice for information security controls based on ISO/IEC 27002 for telecommunications organizations.
- ISO/IEC 27017 - Code of practice for information security controls based on ISO/IEC 27002 for Cloud services.
- ISO/IEC 27018 - Information technology -- Security techniques -- Code of practice for protection of personally identifiable information in public Clouds acting as PII processors.
- ISO/IEC 27019 - Information security controls for the energy utility industry.
- ISO/IEC 15408 - Evaluation criteria for IT security.
- ISO/IEC 18045 - Methodology for IT security evaluation.
- ISO/IEC 19790 - Security requirements for cryptographic modules.
- ISO/IEC 30111 - Vulnerability handling processes.
- ISO/IEC 29147 - Vulnerability disclosure.
- ISO/IEC 19608 - Guidance for developing security and privacy functional requirements based on ISO/IEC 15408.

Structure: This newly formed joint technical committee has created six working-groups:

- WG 1 Chairman advisory group;
- WG 2 Cybersecurity Management Systems;
- WG 3 Security evaluation and assessment;
- WG 4 Cybersecurity services;
- WG 5 Data protection, privacy and identity management;
- WG 6 Product security.

Interesting projects: CEN/CLC JTC 13 has also decided to establish feasibility studies on the following subjects:

- Small Business Standards (SBS) Guide on the implementation of ISO/IEC 27001 for Small and Medium sized Enterprises (SMEs);
- Standards on basic requirements for cybersecurity in products;
- Adoption of ISO/IEC 24760 - Framework for Identity Management: Parts 1-3;
- Adoption of ISO/IEC 29100 - Privacy framework;
- Adoption of ISO/IEC 29101 - Privacy architecture framework;
- Adoption of ISO/IEC 29151 - Code of practice for personally identifiable information protection;
- Lightweight and non ISO/IEC 15408 evaluation methods.

Currently, one delegate from Luxembourg is participating in the activities of CEN/CLC JTC 13.

4.2.1.3 CEN/CLC JTC 8 Privacy management in products and services

In 2014, CEN and CENELEC created a new Joint Working Group (JWG) whose main task was to provide the response to the new EC standardization request on "Privacy management in the design and development and in the production and service provision processes of security technologies"[20]. The request aims at the implementation of Privacy-by-design principles for security technologies and/or services lifecycle. The new standardization deliverables are intended to define and share best practices balancing security, transparency and privacy concerns for security technologies, manufacturers and service providers in Europe.

In 2017, the JWG was transformed in a new joint technical committee CEN/CLC JTC 8 [136] that met for the first time in July of the same year. Since then this committee has begun work on the development of a new European Standard setting out requirements on **Privacy-by-design** principles in the design and implementation of security technologies and services in response to a request from the European Commission (M/530). The committee will also begin work on two Technical Reports with specific guidelines for the application of Privacy-by-design principles for video-surveillance and for biometrics for access control including facial recognition[21]. In addition, the committee has voted for the adoption of the ISO/IEC 29134 (Guidelines for privacy impact assessment) as a European standard in order to use it a basis for their work, and has initiated two feasibility studies:

- Development of interface for interchange of personal data between applications;
- Definition of data protection professional profiles.

Scope: The scope of the CEN/CLC JTC 8 is to cover privacy and personal data protection in products and services.

Structure: The joint technical committee is currently organized in two working groups:

- WG 1 Privacy management in products and services;
- WG 2 Video surveillance and access control.

Two delegates from Luxembourg are participating in the activities of this committee.

[20] http://ec.europa.eu/growth/tools-databases/mandates/index.cfm?fuseaction=search.detail&id=548

[21] Source: CEN and CENELEC Work Programme 2017

4.2.1.4 ETSI/TC on Cybersecurity

ETSI/TC CYBER [137] mainly works on cross-domain cybersecurity while addressing more specific domains or security tools and techniques complementing other ETSI groups. TC CYBER serves as a center of expertise and offers security advice and guidance to users, manufacturers and network and infrastructure operators as well as ETSI committees.

ETSI standards developed by TC CYBER – and relevant to the topics of this white paper – are as follows:

- ETSI TS 103 532: "Attribute Based Encryption for Attribute Based Access Control".
- ETSI TS 103 458: "Application of Attribute Based Encryption for PII and personal data protection on IoT devices, WLAN, Cloud and mobile services – High-level requirements".
- ETSI TR 103 304: "PII Protection in mobile and Cloud services".
- ETSI TR 103 456: "Implementation of the Network and Information Security (NIS) Directive".
- ETSI TR 103 306: "Global Cyber Security Ecosystem".
- ETSI TS 102 165-1: "Methods and protocols; Part 1: Method and pro forma for Threat, Vulnerability, and Risk Analysis (TVRA)".

4.2.1.5 ITU-T SG 17 Security

ITU-T Study Group 17 [138] coordinates security-related work across all ITU-T Study Groups. This SG works in cooperation with other SDOs and various ICT industry consortia, and deals with a broad range of standardization issues. For instance, SG 17 is currently working on cybersecurity, security management, identity management, the protection of personally identifiable information, security of applications and services for IoT, Big data analytics, Cloud computing, to name a few. This work is largely undertaken by the following working parties (WP):

- WP 1/17 Telecommunication/ICT Security.
- WP 2/17 Cyberspace security.
- WP 3/17 Application security.
- WP 4/17 Identity management and authentication.

The Recommendations that are currently approved for development include projects such as:

- X.1080.0: Access control for telebiometrics data protection.
- X.1362: Simple encryption procedure for Internet of Things environments.
- X.1550: Access control models for incident exchange networks.
- X.1603: Data security requirements for the monitoring service of Cloud computing.
- X.1092: Integrated framework for telebiometric data protection in e-health and telemedicine.
- X.1171: Threats and requirements for protection of personally identifiable information in applications using tag-based identification.
- X.1275: Guidelines on protection of personally identifiable information in the application of RFID technology.
- Y.IoT-IoD-PT: Identity of IoT devices based on secure procedures and ensures privacy and trust of IoT systems.

4.2.2 Basic data protection and privacy terms from different ISO standards

This section provides the terms and definitions of some basic privacy and data protection concepts, as defined in different ISO standards. ISO's Online Browsing Platform[22] is used for this purpose.

- **Anonymity:**
  - Lacking individuality, distinction, and recognizability within message exchanges (ISO/TR 17427-4:2015).
  - Characteristic of information that does not permit a personally identifiable information principal to be identified directly or indirectly (ISO/IEC 29100:2011 and ISO/IEC 19286:2018).
  - Condition in identification whereby an entity can be recognized as distinct, without sufficient identity information to establish a link to a known identity (ISO/IEC 24760-1:2011).
- **Anonymization:**
  - Process by which personally identifiable information (PII) is irreversibly altered in such a way that a PII principal can no longer be identified directly or indirectly, either by the PII controller alone or in collaboration with any other party (ISO/IEC 29100:2011, ISO/IEC 38505-1:2017, ISO/IEC 27038:2014, ISO 27237:2017).
  - Process that removes the association between the identifying data set and the data subject (ISO/TS 17975:2015, ISO/TS 14441:2013).
  - Process of removing, obscuring, aggregating, or altering identifiers with the aim of preventing the identification of individuals to whom data originally related (ISO 19731:2017).
  - Process whereby the association between a set of recorded information and an identifiable individual is removed where such an association may have existed (ISO/IEC 15944-8:2012).
- **Anonymized data:**
  - Data that has been produced as the output of a personally identifiable information anonymization process (ISO/IEC 29100:2011, ISO 25237:2017).
  - Personal data modified in such a way that direct reference to data subjects is eliminated (ISO 5127:2017).
- **Identifiability:**
  - Condition that results in a personally identifiable information (PII) principal being identified, directly or indirectly, on the basis of a given set of PII (ISO/IEC 29100:2011).
- **Personally identifiable information (PII):**
  - Any information that (a) can be used to identify the PII principal to whom such information relates, or (b) is or might be directly or indirectly linked to a PII principal (ISO/IEC 19086-1:2016, ISO/IEC 19944:2017, ISO/IEC 19286:2018, ISO/IEC 29100:2011, ISO/IEC 27018:2014, ISO/IEC 38505-1:2017, ISO/IEC 27038:2014, ISO/IEC 17789:2014).
  - Any information:
    - that identifies or can be used to identify, contact, or locate the person to whom such information pertains,
    - from which identification or contact information of an individual person can be derived, or
    - that is or might be directly or indirectly linked to a natural person (ISO/IEC 24745:2011).

[22] https://www.iso.org/obp/ui

- **PII controller:**
  - Privacy stakeholder (or privacy stakeholders) that determines the purposes and means for processing personally identifiable information (PII) other than natural persons who use data for personal purposes (ISO/IEC 19086-1:2016, ISO/IEC 19944:2017, ISO/IEC 29100:2011, ISO/IEC 27018:2014).
  - Person designated by an organization to control access to PII (ISO/TR 18638:2017).
- **PII principal:**
  - Natural person to whom the personally identifiable information (PII) relates (ISO/IEC 19086-1:2016, ISO/IEC 19286:2018, ISO/IEC 19944:2017, ISO/IEC 29100:2011, ISO/IEC 27018:2014, ISO/IEC 38505-1:2017).
  - Person who granted/entrusted an organization with the ability to manage his/her PII (ISO/TR 18638:2018).
- **PII processor:**
  - Privacy stakeholder that processes personally identifiable information (PII) on behalf of and in accordance with the instructions of a PII controller (ISO/IEC 19086-1:2016, ISO/IEC 19944:2017, ISO/IEC 29100:2011, ISO/IEC 27018:2014).
- **Privacy controls:**
  - Measures that treat privacy risks by reducing their likelihood or their consequences (ISO/IEC 29100:2011).
  - Technical and organizational measures aimed at mitigating risks that could result in privacy breaches (ISO/TS 14441:2013, ISO/TS 17975:2015).
- **Privacy-enhancing technology (PET):** privacy control, consisting of information and communication technology (ICT) measures, products, or services that protect privacy by eliminating or reducing personally identifiable information (PII) or by preventing unnecessary and/or undesired processing of PII, all without losing the functionality of the ICT system (ISO/IEC 29100:2011).

4.3 Smart ICT standardization

Section 4.2 described standardization projects related to security, privacy and data protection, initiated by different SDOs. This section highlights such standardization activities concerning each Smart ICT domain.

4.3.1 Cloud computing and technical standardization

As mentioned in Table 9, the international standard ISO/IEC 29100:2011 defines a privacy framework that specifies a common privacy terminology, defines the actors and their roles in PII processing, and describes privacy safeguarding considerations. Based on the set of privacy principles described in this standard, supplementary controls in the context of Cloud computing are provided in ISO/IEC 27018 (details below). Similarly, ISO/IEC 29151 (see Table 9) establishes the code of practice for PII protection that could be enhanced for CSC (Cloud service customers) in line with their obligations with respect to collecting, storing and processing of PII. To facilitate this process further, the CSC could properly catalogue the data and identify its sensitivity and the risk to the business in case of data leakage, loss or corruption, using ISO/IEC 27002 as a reference.

Note that the notion of data protection assumes a new dimension in Cloud computing. An organization can opt to store its data in a Cloud service but then data protection responsibility and accountability needs to be agreed upon clearly (e.g., by means of SLAs). Ideally, it should be CSC's responsibility to secure the data before it is stored on a Cloud computing system. However, Cloud service providers are often accountable for data tampering or theft.

Several approaches have been adopted to address the privacy and data protection challenges specific to Cloud computing. For instance, encryption based techniques are used to ensure only authorized access to data in the Cloud. However, this solution requires specially designed key management services. To address the specificities of Cloud computing and to develop relevant technical standards, multiple SDOs have initiated numerous projects. **ISO/IEC JTC 1/SC 38 Cloud computing and distributed platforms** is one of the most prominent committees in this domain.

ISO/IEC JTC 1/SC 38 committee has published 13 standards so far and 9 are currently under development. Luxembourg is a P-member in SC 38 and actively participates in its projects with 16 national delegates. Among SC 38 projects, the ones that are closely related to the topics of this white paper are summarized in Table 10.

Cloud computing aspect	Standardization committee	Project		
		Identifier	Title	Current status
Cloud data storage & processing	ISO/IEC JTC 1/SC 38	ISO/IEC 17826:2016	Cloud data management interface	Published
		ISO/IEC 19944:2017	Data and its flow across devices and Cloud services	Published
		ISO/IEC 19941:2017	Interoperability and portability	Published
Service Level Agreements (SLA)	ISO/IEC JTC 1/SC 38	ISO/IEC 19086-1:2016	Overview and concepts	Published
		ISO/IEC FDIS 19086-2	Metric model	Under development
		ISO/IEC 19086-3:2017	Core conformance requirements	Published
	ISO/IEC JTC 1/SC 27	ISO/IEC FDIS 19086-4	Components of security and of protection of PII	Under development
Security controls	ISO/IEC JTC 1/SC 27	ISO/IEC 29151:2017 ITU-T X.1631	Code for practice for information security controls based on ISO/IEC 27002 for Cloud services	Published
	ETSI CYBER	ETSI TS 103 532 V1.1.1 (03/2018)	Attribute Based Encryption for Attribute Based Access Control	Published
Personally Identifiable Information (PII) protection	ISO/IEC JTC 1/SC 27	ISO/IEC 27018:2014	Code of practice for protection of PII in public Clouds acting as PII processors	Published
	ETSI CYBER	ETSI TR 103 304 v1.1.1 (07/2016)	PII Protection in mobile and Cloud services	Published
		ETSI TS 103 458 v1.1.1 (06/2018)	Application of Attribute Based Encryption for PII and personal data protection on IoT devices, WLAN, Cloud and mobile services – High-level requirements	Published
Trust	ISO/IEC JTC 1/SC 3	ISO/IEC 23186	Framework of trust for processing multi-sourced data	Under development

Table 10: Cloud computing technical standardization projects

4.3.1.1 Cloud data storage and processing

ISO/IEC 17826:2016 - *Information technology – Cloud Data Management Interface (CDMI)* [139]. This international standard specifies the interface to access Cloud storage and to manage the data stored therein. It applies to developers who are implementing or using Cloud storage, and documents how to access Cloud storage and to manage the data stored there.

ISO/IEC 19944:2017 - *Information technology – Cloud computing – Data and its flow across devices and Cloud services* [140]. This standard describes the various types of data flowing within the devices and Cloud computing ecosystem. It describes the impact of connected devices on the data that flow within the Cloud computing ecosystem. ISO/IEC 19944:2017 provides foundational concepts, including a data taxonomy and identifies the categories of data that flow across the CSC devices and Cloud services. The standard is applicable primarily to CSPs, CSCs, and Cloud Service Users, but also to any person or organization involved in legal, policy, technical or other implications of data flows between devices and Cloud services.

ISO/IEC 19941:2017 - *Information technology – Cloud computing – Interoperability and Portability* [141]. Portability is the ability of a CSC to move their data or their applications between two different Cloud services at a low cost and with minimal disruption. Portability is important since CSCs are interested in avoiding lock-in when they choose to use Cloud services. The critical considerations for portability discussions are the porting cos, the risk associated with the porting and how to control the costs and risks compared to the expected benefits. The international standard specifies Cloud computing interoperability and portability types, the relationship and interactions between these two crosscutting aspects of Cloud computing and common terminology and concepts used to discuss interoperability and portability, particularly relating to Cloud services. ISO/IEC 19941:2017 describes engineering solutions required by data portability, interoperability and application portability.

4.3.1.2 Service level agreements

According to the ETSI's Cloud Standards Coordination Group, SLAs should facilitate Cloud Service Customers (CSCs) in understanding the claims behind the Cloud service, and in relating such claims to their own requirements. On the other hand, NIST and the European Commission suggest the use of Cloud SLAs to develop better assessments and perform informed customer decisions, and ultimately to improve trust and transparency between Cloud stakeholders. To comply with global privacy regulations, organizations need to ensure that their CSPs implement technical and administrative controls to protect their data. This implies that the contracts with Cloud service providers should not only define data protection standards but also establish SLAs that outline security and privacy measures.

Ensuring that the data belonging to the CSC is well protected by the CSP that provides Cloud service used by the CSC is a fundamental key capability in Cloud. In the relationship between the CSC and the CSP, the responsibilities and related measures needed to ensure that the CSC's data is secure and protected is described in the Cloud SLAs. In that context, one of the standards that could be used to define Cloud SLAs is ISO/IEC 19086-x Series.

ISO/IEC 19086 – series - *Information technology – Cloud Computing – Service Level Agreement (SLA).* This family of standards seeks to create common Cloud SLA building blocks (concepts, terms, definitions, contexts) that could be used to establish unambiguous SLAs. It is built on ISO/IEC 17788 [4] and ISO/IEC 17789 [142], and with the goal of providing a common terminology, composed of the following four parts:

1. **ISO/IEC 19086-1:2016** - *Information technology – Cloud computing – Service level agreement (SLA) framework and technology – Part 1: Overview and concepts* [22]. This standard defines base terminology and concepts related to Cloud SLAs. This includes service level objectives and lifecycle. It is for the benefit and use of both CSPs and CSCs. The aim is to avoid confusion and facilitate a common understanding between CSPs and CSCs.

2. **ISO/IEC FDIS 19086-2** - *Information technology – Cloud computing – Service level agreement (SLA) framework – Part 2: Metric model*. This standard currently is **under development**. It has reached the *Final Draft International Standard* stage[23]. It provides the capability to assess the delivered characteristics through measurements. ISO/IEC FDIS 19086-2 proposes a technical model for specifying Cloud SLAs metrics. The metrics are used to set the boundaries and margins of error limitations, and can be used for automation, CSP comparison, service monitoring.

3. **ISO/IEC 19086-3:2017** - *Information technology – Cloud computing – Service level agreement (SLA) framework – Part 3: Core conformance requirements* [23]. The ISO/IEC 19086-3:2017 international standard specifies the core conformance requirements for SLAs for Cloud services based on ISO/IEC 19086-1:2016 and guidance on the core conformance requirements. It describes specific requirements for each SLA component.

4. **ISO/IEC FDIS 19086-4** - *Information technology – Cloud computing – Service level agreement (SLA) framework and technology – Part 4: Components of security and of protection of PII*. The standard is **under development in ISO/IEC JTC 1/SC 27** - and reached its *Final Draft International Standard* stage (FDIS). ISO/IEC FDIS 19086-4 represents a major achievement in the area of Cloud SLAs. It specifies the security and privacy aspects of SLA for Cloud services including requirements and guidance. This standard will facilitate common understanding between CSPs and CSCs. It acknowledges the importance of developing common service level objectives and metrics for security SLAs. The standard specifically provides guidance on information security risks associated with the use of Cloud services and managing those risks effectively, and responding to risks specific to the acquisition or provision of Cloud services.

4.3.1.3 Cloud security controls

ISO/IEC 27017:2015/ITU-T X.1631 - *Information technology – Security techniques – Code for practice for information security controls based on ISO/IEC 27002 for Cloud services* [21] is an international standard that aims to enhance the set of controls in ISO/IEC 27002 to cover all the security and privacy aspects of operating a Cloud service. ISO/IEC 27017:2015/ITU-T X.1631 "provides guidelines supporting the implementation of information security controls for CSCs and CSPs. These guidelines guide the CSPs to assist the CSCs in implementing the controls, and guide the CSC to implement such controls. Selection of appropriate information security controls, and the application of the implementation guidance provided, will depend on a risk assessment as well as any legal, contractual, regulatory or other Cloud-sector specific information security requirements".

4.3.1.4 Personally identifiable information (PII) protection

One approach to address the privacy concerns in Cloud computing consists in proper and consistent collection, processing, communication, use and disposition of PII in relation to Cloud services. In fact, several jurisdictions have defined strict rules and regulations that Cloud services that store and process PII must comply.

ISO/IEC 27018:2014 - *Information technology – Security techniques – Code of practice for protection of personally identifiable information (PII) in public Clouds acting as PII processors* [143] is an international standard that focuses on protection of privacy of PII that is processed in the Cloud, specifically by a public CSP.

This standard establishes commonly accepted control objectives, guidelines for implementing measures to protect PII in accordance with the privacy principles in ISO/IEC 29100 for the public Cloud computing environment. To achieve this, it provides a set of controls, supplementing ISO/IEC 27002, aimed at CSPs who act as PII processors on behalf of a PII controller. Furthermore, the standard specifies guidelines based on ISO/IEC 27002, taking into consideration the regulatory requirements for the protection of PII that might be applicable within the context of the information security risk environment(s) of a provider of public Cloud services. The idea of the standard is that a CSP can have an ISMS audited using the ISO/IEC 27001 requirements, where the auditor verifies that the risk

[23] ISO's harmonized stage codes: https://www.iso.org/stage-codes.html

management process and subsequent ISMS implementation has properly taken into account the supplementary set of controls in ISO/IEC 27018 [143].

Efficient implementation of this standard assures Cloud users that their CSP is well placed to keep data private and secure and helps foster transparency in Cloud provider's privacy practices, while advancing stronger protections for customer data in the Cloud. In summary, ISO/IEC 27018:

- Enables transparency so customers can choose well-governed Cloud services;
- Helps CSPs that process PII to address applicable legal obligations as well as customer expectations;
- Facilitates the creation of contracts for Cloud services;
- Provides Cloud customers with a mechanism to ensure Cloud providers' compliance with legal and other obligation.

Other approaches could also be helpful, such as the Cloud Security Alliance Privacy Level Agreement[24], which provides assurances for a given Cloud service in relation to PII processing.

4.3.1.5 Trust

ISO/IEC PDTR 23186 *Information technology – Cloud computing – Framework of trust for processing of multi-sourced data* [144] is a project that is currently under development. This document describes a framework of trust for the processing of multi-sourced data that includes data use obligations and controls, data provenance, chain of custody, security and immutable proof of compliance as elements of the framework.

4.3.2 Internet of Things and technical standardization

Given that IoT is a combination of several technologies, standardization efforts in this domain could also be viewed as a culmination of wide-ranging and diverse initiatives [7]. For instance, in 2009, the working group ISO/IEC JTC 1/WG 7 on **Sensor Networks** was established to address the areas related to **M2M** (Machine-to-Machine) and IoT, besides sensor network standardization. ETSI also established a technical committee called Smart M2M to focus on M2M standardization as well as to address standardization gaps identified in European Commission's Large Scale Pilot projects [145].

In this context, ETSI TC SmartM2M [146] is developing standards to enable M2M services and applications and certain aspects of IoT. The committee's focus is on an application-independent 'horizontal' service platform with architecture capable of supporting a very wide range of services including smart metering, smart grids, eHealth, city automation, consumer applications and car automation [145]. Examples of some published and standards under development (by ETSI) related to IoT and the topics of this whitepaper are:

- ETSI TR 103 290 (04/2015): "Machine-to-Machine communications; Impact of smart city activity on IoT environment".
- ETSI TR 103 375 (10/2016): "SmartM2M IoT Standards landscape and future evolutions".
- ETSI TR 103 376 (10/2016): "SmartM2M IoT use cases and standards gaps".
- ETSI TR 118 538: "oneM2M Developer guide: Implementing security example" (under development) [147].

Besides sensor networks and M2M, the group of technologies that facilitate **Automatic Identification and Data Capture (AIDC**) are inherent to the IoT paradigm. AIDC technologies such as bar coding and Radio-Frequency Identification (RFID) provide quick, accurate and cost-effective ways to identify, track, acquire and manage data and information about items, personnel, transactions and resources. **ISO/IEC JTC 1/SC 31** [148] and **CEN/TC 225** [149] develop standards related to **AIDC technologies** (see Section 4.3.2.1).

[24] https://cloudsecurityalliance.org/download/privacy-level-agreement-version-2/

Finally, in order to synchronize standardization efforts (among ISO/IEC JTC 1 committees and with external organizations such as ITU-T) as well as to identify market requirements, a Special Working Group (SWG) on IoT was formed towards the end of 2012. Based on the recommendations of this SWG, in 2014, a new working group ISO/IEC JTC 1/WG 10 that was responsible for the development of foundational standards for IoT (e.g. terms and definitions, reference architecture, etc.) was created. Finally, ISO/IEC JTC 1 decided to create **ISO/IEC JTC 1/SC 41 Internet of Things and related technologies** [150] in November 2016 and recommended it to overtake the work of WGs 7 and 10 (see Section 4.3.2.2).

The rest of this section summarizes the standardization activities in abovementioned two parts.

4.3.2.1 AIDC technologies

AIDC is an industry term that describes the identification and/or direct collection of data into a microprocessor-controlled device, such as a computer system or a programmable logic controller (PLC), without the use of a keyboard. AIDC technologies provide a reliable means not only to identify but also to track items. It is possible to encode a wide range of information, beginning with a basic item or the identification of a person, to comprehensive details about the item or person, e.g. item description, size, weight, color, etc.

The **ISO/IEC JTC 1/SC 31 Automatic identification and data capture techniques** develops standardization of data formats, data syntax, data structures, data encoding, and technologies for the process of automatic identification and data capture and of associated devices utilized in inter-industry applications and international business interchanges and for mobile applications. This sub-committee has following working groups (WGs):

- JTC 1/SC 31/WG 1 Data carrier;
- JTC 1/SC 31/WG 2 Data structure;
- JTC 1/SC 31/WG 4 Radio communications;
- JTC 1/SC 31/WG 8 Application of AIDC standards.

ISO/IEC JTC 1/SC 31 has published more than 100 standards. These standards provide answers to a range of questions including: how a bar code is created and read, how an RFID tag is read, how a device such as a phone is used to read and access data as well as how the data associated with the technology are stored and read.

The current work program of ISO/IEC JTC 1/SC 31 includes projects such as:

- The revision of the multipart standard ISO/IEC 15961 regarding “Information technology – Radio frequency identification (RFID) for item management: Data protocol”;
- The development of the multipart standard ISO/IEC 19823 entitled “Information technology – Conformance test methods for security service crypto suites”;
- The development of the multipart standard ISO/IEC 29167 concerning security services in the area of “Information technology – Automatic identification and data capture techniques”.

CEN/TC 225 – AIDC Technologies performs standardization of data carriers for automatic identification and data capture, of the data element architecture therefore, of the necessary test specifications and of technical features for the harmonization of cross-sector applications. Establishment of an appropriate system of registration authorities, and of means to ensure the necessary maintenance of standards. This technical committee has following working groups (WGs):

- CEN/TC 225/WG 1 Optical Readable Media;
- CEN/TC 225/WG 3 Security and data structure;
- CEN/TC 225/WG 4 Automatic ID applications;
- CEN/TC 225/WG 5 RFID, RTLS and on board sensors;

- CEN/TC 225/WG 6 Internet of Things - Identification, Data Capture and Edge Technologies.

CEN/TC 225 takes into account the technical work in ISO/IEC JTC 1/SC 31 and ISO/IEC JTC 1/SC 27, and acts as the focal point for IoT issues within CEN.

4.3.2.2 IoT related standardization

ISO/IEC JTC 1/SC 41 currently has 25 P-Members, including Luxembourg, and 9 O-Members. Currently, 12 delegates from Luxembourg are participating in the activities of SC 41. This committee has the following working groups (WGs) and Table 11 summarized the projects untaken by these WGs:

- JTC 1/SC 41/WG 3 IoT Architecture;
- JTC 1/SC 41/WG 4 IoT Interoperability;
- JTC 1/SC 41/WG 5 IoT Applications.

In addition to these working groups, ISO/IEC JTC 1/SC 41 has also established different study groups depending upon the specific requirements of the current market:

- JTC 1/SC 41/**SG 7 Wearables** intends to study market requirements of smart wearable devices, analyze current standardization and research activities in this field, and identify standardization gaps.
- JTC 1/SC 41/**SG 16 Study Group on Reference Architecture and Vocabulary Harmonization**. This study groups works for the harmonization of various reference architectures and vocabulary within a subcommittee. It provides suggestions for removing overlaps and inconsistencies of reference architecture projects.
- JTC 1/SC 41/**SG 17 Study Group on Societal and human factors in IoT based services**. This study group has been recently created to identify categories of IoT based services to humans, with specific respect to their relevant technical properties, and explore how they relate to SC41's current and potential future work.
- JTC 1/SC 41/**SG 18 Study Group on Integration of IoT and Blockchain**. The objective of this study group is defined for providing an analysis of the requirements of the market and the status of current standardization activities in JTC 1 and other SDOs.
- JTC 1/SC 41/**SG 19 Study Group on Realizing Context Specific Solution / System Architecture based on IoT RA.** This study group has been created to investigate Realizing Context Specific Solution / System Architecture based on IoT RA.

The ITU-T's Study Group (SG 20) – IoT and its applications including smart cities and communities (SC&C) has defined the following seven study questions (roughly equivalent of WGs in ISO/IEC committees) to realize its program of work:

- Q1/20 - End to end connectivity, networks, interoperability, infrastructures and **Big data** aspects related to IoT and SC&C.
- Q2/20 - Requirements, capabilities and use cases across verticals.
- Q3/20 - Architectures, management, protocols and Quality of Service.
- Q4/20 - e/Smart services, applications and supporting platforms.
- Q5/20 - Research and emerging technologies, terminology and definitions.
- Q6/20 - Security, Privacy, Trust, and Identification for IoT and SC&C.
- Q7/20 - Evaluation and assessment of Smart Sustainable Cities and Communities.

Table 11 outlines projects related to IoT standardization.

IoT aspect	Standardization Committee	Project		
		Identifier	Title	Current status
Foundations (vocabulary, architecture and frameworks)	ISO/IEC JTC 1/SC 41 WG 3	ISO/IEC 20924	Definitions and vocabulary	Under development
		ISO/IEC 30141	Reference architecture	Published
		PWI TR JTC1-SC41-1	Technical Report (TR) on IoT Edge Computing	Under development
		ISO/IEC 30147	Methodology for trustworthiness of IoT system/service	Under development
Interoperability, connectivity, conformance and testing	ISO/IEC JTC 1/SC 41 WG 4	ISO/IEC 21823-1	Interoperability for IoT Systems – Part 1: Framework	Under development
		ISO/IEC 21823-2	Interoperability for IoT Systems – Part 2: Transport interoperability	Under development
		ISO/IEC 21823-3	Interoperability for IoT Systems – Part 3: Semantic interoperability	Under development
Applications, platforms, use cases, middleware, tools and implementation guidance	ISO/IEC JTC 1/SC 41 WG 5	ISO/IEC TR 22417:2017	IoT use cases	Published
IoT Security	ITU-T SG 17	X.1361 (ex X.iotsec-2)	Security framework for IoT based on the gateway model	Under development
		X.secup-iot	Secure software update procedure for IoT devices	Under development
		X.nb-iot	Security requirements and frameworks for Narrow Band IoT	Under development
		X.ibc-iot	Security framework for use of identity-based cryptography in support of IoT services over Telecom networks	Under development
		X.ssp-iot	Security requirement and framework for IoT service platform	Under development

IoT aspect	Standardization Committee	Project		
		Identifier	Title	Current status
PII protection in IoT environments	ITU-T SG 17	X.iotsec-3	Technical framework of PII handling system in IoT environment	Under development
	ETSI CYBER	ETSI TS 103 458 v1.1.1 (06/2018)	Application of Attribute Based Encryption for PII and personal data protection on IoT devices, WLAN, Cloud and mobile services – High-level requirements	Published

Table 11: IoT related technical standardization

IoT Foundations

- **ISO/IEC 20924:** This draft contains definitions and provides a list of terms and concepts that are often used in the IoT context.
- **ISO/IEC 30141:** This draft provides a common reference architecture for IoT implementation. As mentioned above a study group, SG 19, has been created to investigate "Realizing Context Specific Solution/System Architecture" based on IoT reference architecture.
- **PWI TR JTC 1-SC 41-1:** This technical report is in progress as a recommendation of SG 10 "IoT Edge computing". It will examine the requirements relating to IoT Edge Computing from use cases, and provide an analysis of activity on this topic amongst SDOs to understand if gaps exist in the standards landscape. Another contribution of this report would be to propose a reference architecture for IoT Edge Computing along with the major technologies involved in IoT Edge Computing.
- **ISO/IEC 30147:** The purpose of this draft is to make a methodology to implement and maintain trustworthiness in an IoT system/service because the existing methodologies are targeted to each application area and do not cover all the challenges in IoT systems/services.

Interoperability, connectivity, conformance and testing

- **ISO/IEC 21823:** These three drafts are dedicated to facilitate the interoperability issues within and outside of an IoT system. The goal of the first part (ISO/IEC 21823-1) "Framework" is to ensure that all parties involved in building and using IoT systems have a common understanding of interoperability as it applies to IoT systems and the various entities within them. Similarly, ISO/IEC 21823-2 and ISO/IEC 21823-3 are for transport- and semantic-interoperability within and outside of IoT system respectively.

Applications, platforms, use cases, middleware, tools and implementation guidance

- **ISO/IEC TR 22417:2017:** This technical report was published in 2017 [150]. The purpose of this report is to identify IoT scenarios and use cases based on real world applications and requirements. This document comprises 25 use cases for IoT. The use cases are a well-known tool for expressing requirements at a high level and demonstrating their real-life relevance. They provide a practical context for considerations on interoperability and standards based on user experience. Furthermore, they clarify where existing standards can be applied and highlight where standardization work is needed.

IoT Security

In addition to SG 20, the ITU-T SG 17 (Security) has initiated the following projects related to security, privacy and/or data protection of IoT.

- **ITU-T X.iotsec-2:** This Recommendation provides a security framework for IoT analyzing security threats and challenges inherent to the IoT environment, and describes security capabilities that could mitigate identified threats and challenges. A methodology for determining which security capabilities will require specification is also provided.
- **X.secup-iot:** This Recommendation aims to provide a basic model for secure update of IoT software/firmware and a common secure update procedure for IoT software/firmware implemented in IoT devices and systems.
- **X.nb-iot:** This Recommendation analyzes the potential deployment scheme and typical application scenarios of Narrow Band (NB) IoT, specifies the security threats and requirements specific to the NB IoT deployments and thus establishes the security framework for the operator to safeguard concerning technology applications.
- **X.ibc-iot:** This Recommendation provides a security framework for the use of identity-based cryptography in support of IoT services over telecom networks.
- **X.ssp-iot:** This Recommendation intends to define security requirements and framework for IoT service platforms, and to describe security measures that could mitigate the security threats and challenges.

PII protection in IoT environments

- **X.iotsec-3:** This Recommendation aims to provide a technical framework to handle PII in IoT environments. Since IoT devices collect different kinds of data, including PII, which can be shared with multiple service providers, this project will define a framework to flexibly reflect user's intention on data usage.
- **ETSI TS 103 458 v1.1.1 (06/2018):** The current draft under development tries to specify high-level requirements for the application of Attribute Based Encryption (ABE) to protect PII and personal data on IoT devices/services, Cloud services, Wireless Local Area Networks and mobile services, where access to data has to be given to multiple parties and under different conditions. With a focus on the confidentiality of data, including personal data and PII, this draft may help in supporting the GDPR.

ISO/IEC JTC 1/SC 41 currently has proposals to develop standards related to IoT trustworthiness framework and requirements of IoT data exchange platform for various IoT services.

4.3.2.3 Other relevant initiatives

ISO/IEC JTC 1/SC 25 Interconnection of information technology equipment [151] is for the standardization of microprocessor systems; and of interfaces, protocols, architectures and associated interconnecting media for information technology equipment and networks, generally for commercial and residential environments. It is dedicated for the development of network interfaces, in liaison with committees for external utility networks, to support smart grid applications at the customer premises. It is also intended to support embedded and distributed computing environments, storage systems, other input/output components, home and building electronic systems including customer premises smart grid applications for electricity, gas, water and heat.

WG 1: Home electronic systems of SC 25 is responsible for the **Home Electronic System (HES)** series of standards. It develops standards for the interconnection of electrical and electronic equipment and products for **homes** and **small buildings**. The primary markets for WG 1 standards are developers, manufacturers, and installers of these products and related services. Homes are made intelligent with interconnected **sensors, actuators, user interfaces**, and **controllers**, which may be embedded in smart consumer appliances. Such networks use a variety of media: IT cabling, wireless and power line communication. Home networks using structured cabling specified by subcommittee 25 are now routinely offered for many new and renovated homes. Wireless and power line carrier technologies are facilitating the introduction of networks into existing homes.

4.3.3 Big data and technical standardization

The standardization activities for Big data started in 2014 with the establishment of **ISO/IEC JTC 1/WG 9 Big data**. This group initiated several projects that focused on developing foundational standards for Big data. However, in 2018, JTC 1/WG 9 was disbanded following the creation of **ISO/IEC JTC 1/SC 42 Artificial Intelligence** [152]. The newly established committee JTC 1/SC 42 has been given the responsibility of covering various aspects related to artificial intelligence, including among other topics the Big data processing. Subsequently, the work-program of JTC 1/SC 42 has included the projects that were initiated by JTC 1/WG 9. In this sense, JTC 1/SC 42 will ensure the review of two international standards that were published by JTC 1/WG 9 and will finalize the two standards that were under development of JTC 1/WG 9. Additionally, JTC 1/SC 42 has started the development of two standards that are specific to artificial intelligence.

ISO/IEC JTC 1/SC 42 has established three study-groups (SGs) and one working-group (WG):

- SG 1: Computational approaches and characteristics of artificial intelligence systems;
- SG 2: Trustworthiness;
- SG 3: Use cases and applications;
- WG 1: Foundational standards.

The committee has 20 P-members including Luxembourg and 7 O-members. The National Mirror Committee of SC 42 in Luxembourg consists of 14 delegates most of them were actively involved in the work of WG 9 Big data and the development of Big data standards.

The topic of artificial intelligence being vast, the work of other technical committees also contributes to the development of the standards in the domain. Table 12 provides an overview of most relevant standardization projects related to Big data led by different technical committees and the rest of this section discusses in detail these projects, along the Big data aspects defined in the table.

Big data aspect	Standardization Committee	Project		
		Identifier	Title	Current status
Vocabulary and definitions	ISO/IEC JTC 1/SC 42	ISO/IEC DIS 20546	Definition and vocabulary	Under development
		ISO/IEC AWI 22989	Artificial intelligence (AI) concepts and terminology	Under development
	ISO/TC 69/WG 12	ISO/NP 3534-5	Terms used in Big data (predictive analysis)	Under development
Reference architecture (ISO/IEC 20547 series)	ISO/IEC JTC 1/SC 42	ISO/IEC AWI TR 20547-1	Framework and application process	Under development
		ISO/IEC TR 20547-2	Use cases and derived requirements	Published
		ISO/IEC DIS 20547-3	Reference architecture	Under development
	ISO/IEC JTC 1/SC 27	ISO/IEC AWI 20547-4	Security and privacy	Under development
	ISO/IEC JTC 1/SC 42	ISO/IEC TR 20547-5	Standards roadmap	Published
Processing, including artificial intelligence	ISO/TC 69/WG 12	ISO 23347	Big data analytics – data science life cycle	Under development
		ISO/NP TR 23348	Big data analytics – Model validation	Under development
	ISO/IEC JTC 1/SC 42	ISO/IEC AWI 23053	Framework for AI systems using machine learning	Under development
Data quality and metadata	ISO/IEC JTC 1/SC 7	ISO/IEC 25012:2008	Software product Quality Requirements and Evaluation (SQuaRE) – Data quality model	Published
Security and privacy	ISO/IEC JTC 1/SC 27	ISO/IEC AWI 20547-4 (repeated here for the sake of completeness)	Reference architecture –security and privacy	Under development

Table 12: Big data technical standardization projects

4.3.3.1 Vocabulary, definitions and reference architecture

As described above, the projects initiated by ISO/IEC JTC 1/WG 9 are as follows:

- ISO/IEC DIS 20546 - Information Technology – Big data – Definition and Vocabulary;
- ISO/IEC 20547- Information Technology – Big data – Reference Architecture:
  - ISO/IEC AWI TR 20547-1, Part 1: Framework and Application Process,
  - ISO/IEC TR 20547-2, Part 2: Use Cases and Derived Requirements,
  - ISO/IEC DIS 20547-3, Part 3: Reference Architecture,
  - ISO/IEC AWI 20547-4, Part 4: Security and Privacy,
  - ISO/IEC TR 20547-5, Part 5: Standards Roadmap.

ISO/IEC DIS 20546 containing definitions, provides a list of terms and concepts that are often used in the context of Big data. Along the concise definition of the term "Big data" itself, the standard introduces various characteristics of data usually referred to as the Vs of Big data (see Section 1.1.3). Moreover, a concept of distributed processing is introduced with the explanation of existing paradigms (vertical and horizontal scaling) and possibilities (distributed file system, cluster computing, etc.). In addition, the concepts of **Cloud computing** and **IoT** are explained as the enabling technologies for Big data analytics.

ISO/IEC AWI TR 20547-1: introduces the key concepts of Big data references architecture that are further detailed in other parts. It sets a background for Big data standardization and suggests a way to apply reference architecture when building a Big data system.

ISO/IEC TR 20547-2: this standard [153] is a technical report containing use cases served to identify the major challenges related to Big data and overview the possible solutions. The standard was published this year but contains use cases provided 4 years ago, making some of presented challenges obsolete. In this context, the group started the work on the update of the document by collecting new use cases with up-to-date challenges.

ISO/IEC DIS 20547-3: introduces a reference architecture for a Big data ecosystem. On the one hand, the standard describes the roles and activities of various stakeholders that may be involved in a Big data project. On the other hand, it offers a functional view of Big data ecosystem. It provides an overview of the components of Big data ecosystem and of their functionalities, such as data collection, storage and processing, for example. The most frequent ways of communication between these components, referred to as communication interfaces, are also listed in the standard. Finally, the standard makes reference to the transversal activities that impact all functional components, such as data governance, data management and security and privacy protection. The latter should be further detailed in the **ISO/IEC DIS 20547-4** that is under responsibility of ISO/IEC JTC 1/SC 27 IT security techniques.

ISO/IEC TR 20547-5: this technical report [154] presents the existing standards that are in one or another way could be helpful when designing, developing and implementing Big data product or service. The goal of the document is to provide the repository of relevant standards and to identify the gaps for future standards development. As Big data attracts more and more interest from other standardization committees the list of related standards continue to grow. Thus, JTC 1/WG 9 initiated the work on the update of this document ISO/IEC TR 20547-5 that is being continued by ISO/IEC JTC 1/SC 42.

The rise of interest in Big data analytics and artificial intelligence, along with the increase in a number of tools for these technologies, led to the need of developing standard definitions that could be used by different involved parties. Thus, **ISO/TC 69 Applications of statistical methods** has initiated the work on the definition of Big data analytics terms:

- ISO/NP 3534-5 - Statistics – Vocabulary and symbols – Part 5: Terms used in Big data (predictive analytics).

In its turn, ISO/IEC JTC 1/SC 42 approved the project that will introduce the concepts related to artificial intelligence:

- ISO/IEC 22989 - Artificial Intelligence Concepts and Terminology.

4.3.3.2 Big data processing

Big data could be regarded as an asset that has its own value. However, the outcomes of Big data processing could potentially be assets that are even more valuable. In order to support the processing of Big data and increase its trustworthiness, different technical standardization committees have recently started working on this topic.

One of the first aspects of data processing that was addressed by standardization community is the extended support of relational databases to new types of data, such as XML, JSON, multi-dimensional arrays, etc. The work is carried out by WG 3 Database language of **ISO/IEC JTC 1/SC 32 Data management and Interchange** [155]. The group has been developing SQL related standards for about 30 years. Since 2015, they have produced a number of standards and technical reports that would help addressing Big data variety challenge within the frame of traditional databases.

More recent developments tackle the analytical framework for Big data. Thus, **ISO/TC 69 Applications of statistical method**s [156] has recently established a WG 12 Big data Analytics. To start with and aside from definitions, the group is going to work on the following newly approved projects:

- ISO 23347 Statistics - Big data Analytics - Data Science Life Cycle;
- ISO 23348 Statistics - Big data Analytics - Model Validation.

The goal of these projects is to provide a standardized framework for statistical data analysis and address some specific challenges related to the analysis of Big data. ISO 23347 Data Science Lifecycle will describe the end-to-end data science life cycle in the context of Big data, and the statistical methods for describing the distribution of data values in huge datasets. ISO 23348 Model Validation will provide guidelines on the techniques of checking and validation of models and results of Big data analysis, including model parametrization, resampling techniques, etc. The standard will also suggest how to compare the performance of different models and how to measure the quality of the outputs.

Another committee that is going to address the challenges related to Big data analytics is ISO/IEC JTC 1/SC 42. In addition to working on the foundational standards of Big data, initiated by ISO/IEC JTC 1/WG 9, and the definition of terms related to artificial intelligence, SC 42 started the work on a general framework for machine learning:

- ISO/IEC 23053 - Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML).

Among other topics that are on the agenda of SC 42 are:

- Computational approaches and characteristics of AI systems:
  - Technologies used by AI systems;
  - Specialized AI systems and underlying computational approaches, architectures, characteristics, etc.

- **Trustworthiness:**
  - Approaches to establish trust in AI systems through transparency, verifiability, explainability, controllability, etc.;
  - Approaches to achieve AI systems' robustness, resiliency, reliability, accuracy, safety, security, privacy, etc.;
  - Types of sources of bias in AI systems with a goal of minimization, including but not limited to statistical bias in AI systems and AI aided decision-making, etc.
- Use Cases and Applications:
  - Identify different AI application domains (e.g., social networks and embedded systems) and the different context of their use (e.g., fintech, health care, smart home, and autonomous cars);
  - Collect representative use cases.

4.3.3.3 Standardization activities with a focus on Big data quality and metadata

The quality of data analysis depends heavily on the quality of data. The committee **ISO/IEC JTC 1/SC 7 Software and systems engineering** [157] has been working on a family of standards referred as ISO/IEC 25000 System and Software Quality Requirements and Evaluation (SQuaRE). These standards aim at providing comprehensive guidelines to achieve system and software quality, and one of the pillars they highlight is Data Quality Model, detailed in ISO/IEC 25012. ISO/IEC 25012 [158] identifies 15 characteristics (see Table 13) describing data quality that fall into two major categories: inherent (data values respecting domain rules and restrictions, metadata, etc.) and system dependent data quality (when data is used within a computer system):

Inherent	System dependent	Both inherent and system dependent
• Accuracy • Completeness • Consistency • Credibility • Currentness	• Availability • Portability • Recoverability	• Accessibility • Compliance • Confidentiality • Efficiency • Precision • Traceability • Understandability

Table 13: Data quality characteristics defined in ISO/IEC 25012

In order to go farther and develop standards that are specific to Big data quality, the WG 2 Metadata of **ISO/IEC JTC 1/SC 32** has established an ad-hoc group **Metadata for Big data Quality**. The group analyzed the ISO/IEC 25012 but also other standards related to data quality, namely ISO 8000 family (quality of industrial data) and ISO 19113, ISO 19115, ISO 19157 (geospatial data). As a result, they provided preliminary report where they extend the list of data quality characteristics identified in ISO/IEC 25012 (including, for example, scalability and reliability) and suggested a classification of metadata for Big data quality:

- Common;
- Service Oriented;
- Software oriented;
- Others.

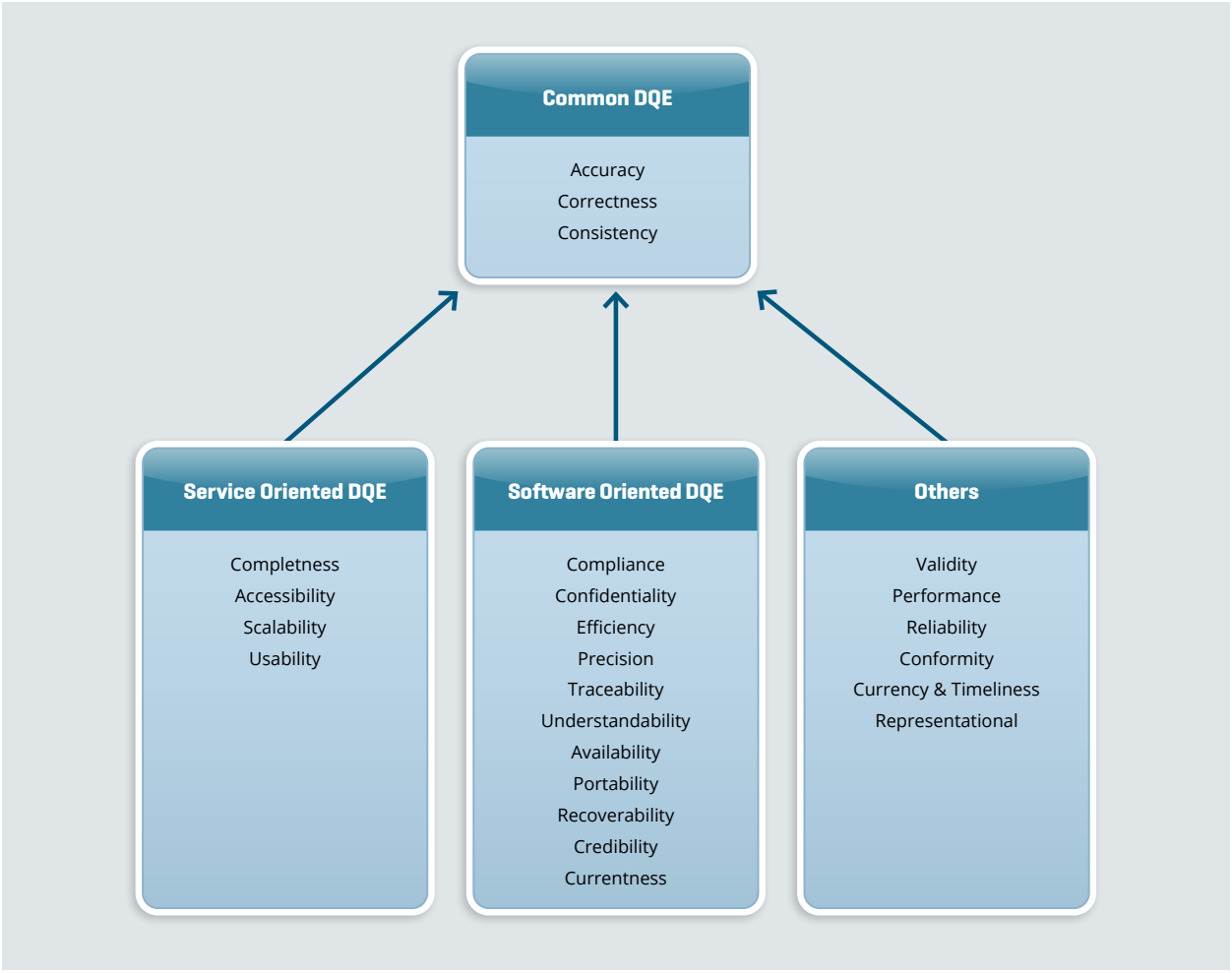


Figure 7: Classes and attributes for Data Quality Management [159]

The goal of the ad-hoc group is to submit a new working item proposal on the usage of metadata for Big data quality management.

4.3.3.4 Standardization activities for Big data security and privacy

As shown in Table 12, ISO/IEC JTC 1/SC 27 [125] is currently responsible for developing **ISO/IEC 20547-4 Big data Reference Architecture – Security and Privacy**. In line with ISO/IEC 20547-3, this standard will outline the activities associated with Big data security and privacy, describe functional components supporting it, and introduce the steps to achieve Big data privacy- and security-by-design.

Moreover, the WG 4 of SC 27 has been working on the assessment model for Big data security in products and services (see Section 4.2.1.1). As a result of their analysis, they propose to develop a new series of standards on Big data security and privacy processes. The goal of these standards would be to facilitate the development of a trustable Big data industry. The scope for the proposed four parts of this new project covers definition of "process reference, assessment and maturity models for the domain of Big data security and privacy. These models are focused on process architecture and the processes used to achieve Big data security and privacy, most specifically on the maturity of those processes. The processes include a set of indicators of process performance and process capability. The indicators are used as a basis for collecting the objective evidence that enables an assessor to assign ratings. These processes are described in different terms, such as process purpose, outcomes, activities and tasks." [160].

4.3.3.5 Other relevant initiatives

ITU-T initiatives: ITU-T approaches Big data from telecommunications perspective, often addressing it with relation to Cloud, IoT or communication protocols. ITU-T standardization activities related to Big data fall under the work of four study groups (SGs):

- SG 3 Tariff and accounting principles and international telecommunication/ICT economic and policy issues;
- SG 13 Future networks (& Cloud);
- SG 17 Security;
- SG 20 IoT, smart cities & communities.

SG 3 "Tariff and accounting principles and international telecommunication/ICT economic and policy issues" aims at providing better understanding of how the growth of ICT could affect the economy. In this context and with respect to Big data, they are working on 2 deliverables:

- A technical paper on economic and policy aspects of Big data in international telecommunication services and networks;
- ITU-T D policy framework and principles for data protection in the context of Big data relating to international telecommunication services.

SG 13 "Future Networks (& Cloud)" is exploring the opportunities and addressing the challenges raised by Smart ICTs and future networks, such as 5G. Under the responsibility of SG 13 there are topics like:

- Big data functional architecture;
- Big data functional requirements on data provenance, integration, exchange, preservation;
- Big data driven networks including requirements to such networks, functional architecture, data traffic management and planning, etc.;
- **Cloud-based Big data services** including Cloud computing requirements and capabilities, functional architecture for BdaaS **(Big data as a Service)**, requirements for data storage and federation;
- Machine learning (ML) for future networks including 5G exploring explore the possibilities that ML techniques offer for future networks in terms of security, data exchanges, etc.

SG 17 Security is working among other documents on a range of deliverables related to the security of Big data such as:

- Guidelines for Cloud service customer data security;
- Security requirements and framework for Big data analytics in mobile internet services;
- Guidelines on security of Big data as a service;
- Security guidelines for Big data infrastructure and platform;
- Security guidelines of lifecycle management for telecom Big data.

SG 20 "IoT, smart cities & communities" has been working on the **standards for IoT** communication and services in the **smart cities**. As IoT is a major source of Big data, one of the SG's projects is entitled **Specific requirements and capabilities of the Internet of Things for Big data**. To study further the relationship between IoT and Big Data, a focus group on Data Processing and Management (FG DPM) to support IoT and Smart Cities & Communities was established under direct responsibility of SG 20. The FG DPM study the questions of interoperability, data formats, using blockchain for data management, as well as the data quality, privacy and other aspects of enabling trust. Another topic of interest in SG 20 is Open Data, including its types and the usage for smart cities.

IEEE Initiatives: Another initiative worth mentioning in the context of privacy protection is the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. With the goal of prioritizing human well-being, the involved experts are working on two outputs:

- Ethically Aligned Design: A Vision for Prioritizing;
- Standards projects related to ethical considerations in AI and autonomous systems.

Among the standardization projects currently under consideration within this framework, IEEE P7002 “Data Privacy Process” aims to address privacy concerns while collecting and processing personal data in AI.

5

Links between scientific research and technical standardization

5. Links between scientific research and technical standardization

This chapter highlights some links between the results of Chapters 3 and 4 where state-of-the-art scientific developments and technical standardization activities are presented respectively in the context of security, privacy and data protection in Smart ICT.

5.1 Cloud computing

A comprehensive overview of privacy and data protection in Cloud computing was given in Chapters 3 and 4 while fundamentals were clarified in Chapter 1 and its relationship with the Big data and IoT paradigms in Chapter 2. In this section, the links between these two lines of works are presented through the perspective of **framework of trust, interoperability and portability, terminology** as well as **pricing strategy**.

- **Framework of trust:** For CSPs, mechanisms to build trust play a decisive role in attracting more consumers. With loss of control over resources from the perspective of the Cloud user, CSP should guarantee users with transparent control over all the data that is stored in the Cloud. In [45] [46] [47], research works have made efforts to establish trust mechanisms among different levels of Cloud computing services. Similarly, [53] made contributions through an integrated trust mechanism with cryptographic RBAC, which can ensure security during data sharing in Cloud. Technical standards could play an active role in enabling CSPs to put trust mechanisms in place; for instance, where CSP's operations and trust mechanisms are certified against international standards. Relevant technical standardization committees (e.g., ISO/IEC JTC 1/SC 38 and JTC 1/SC 27) are already involved in developing technical standards to build trust between users and service providers. For instance, while ISO/IEC 27017:2015 [21] is already contributing towards building trust relationship between users and service providers by means of security controls, other efforts are under development (e.g., ISO/IEC PDTR 23186 *Information technology -- Cloud computing -- Framework of trust for processing of multi-sourced data*).
- **Interoperability and portability:** While multiple CSPs offer Cloud services to users, users may need to transfer the application to another service provider for many reasons (e.g., the availability of CSP's servers is not high, better prices from another provider, resilience goals etc.). Therefore, interoperability between different Cloud services need to be guaranteed. The prospect for Cloud users to move between different vendors is not possible without common standards concerning Cloud computing interoperability. While standardization efforts for interoperability have begun, involving processes for messages transmission, data transmission and VM transfers, there are still open issues that need to be addressed. Recently, several research works have focused on how to cope with the secure interoperability in Cloud environment with the lack of standardized system for data format, communication interface, etc. Communication protocols [50] [51] to cope with the security of the communication protocols and trust model for the interoperability have also been proposed. Meanwhile, [55] [56] [57] make efforts and contributions on exploring the comprehensive policy management framework in Cloud computing and addressing the secure interoperation. In [51] [162], some standardization efforts have been made to support the shared identity among different Cloud providers. To build a common terminology as well as to facilitate interoperability and portability, an international standard ISO/IEC 19941 *Information technology -- Cloud computing -- Interoperability and portability* has been published in 2017.

- **Terminology:** Many emerging innovative Cloud services are being proposed on a daily basis in a rapidly growing market. However, from the users' point of view, there is a confusion about the terminology since there is a lack of consistent description of Cloud services and components offered by CSPs. At the same time, there are different definitions for similar Cloud-related terms in research efforts. Hence, a standardized terminology is necessary. The international standard ISO/IEC 19086-series *Information technology Standard – Cloud computing – Service Level Agreement (SLA)* has been updated in 2017 which aims at unifying the technical terminologies in Cloud computing. It is based on two significant international standards related to Cloud computing ISO/IEC 17788 (Overview and vocabulary) and ISO/IEC 17789 (Reference Architecture). ISO/IEC CD 22123 (Concepts and terminology) is also currently under development.
- **Pricing strategy:** Cloud service providers usually have their own pricing strategies due to the lack of a standardized metering indicator and billing system. Consequently, for the end users, the complexity to choose a service provider increases. These challenges create a need for a standardized metering indicator and billing system that can offer Cloud users a tool for choosing, using and evaluating Cloud services, bring the foundations for settlement among different CSPs, and provide keystones for Cloud service metering and pricing. Recently, ISO/IEC JTC 1/SC 38 started working on a technical report ISO/IEC NP TR 23613 *Information technology – Cloud service metering and billing element*s to determine the gaps between market requirements and current industry practices.

5.2 Internet of Things

IoT, an essential component of the emerging cyber-world, is a paradigm that involves many information and communication technologies, as explained in the previous chapters. The network of connected devices and objects, capable of capturing and disseminating data, allows the development of new innovative services for the benefit of the society by improving services across numerous sectors such as healthcare, transport, and environmental management. While there is a considerable potential, challenges specifically related to privacy and data protection needs to be addressed in order to fully benefit from this paradigm. In the following, links between research efforts and technical standardization activities are presented along these topics:

- Trustworthiness;
- Terminology;
- Reference architecture;
- Interoperability.

Trustworthiness: Market's perception of trustworthiness depends on the indices of data protection measures, regulatory compliance, security and privacy, among others [163] [164]. However, ensuring privacy within this rapidly expanding network of connected physical and virtual objects remains challenging, emphasizing the need for new protocols in various research areas. One predominant challenge emphasized by the scientific community is that inconsistencies arise across IoT systems as systems with varying privacy policies interact with one another. Therefore, notification schemes, continuous online consistency checking and resolution protocols and strategies are necessary.

Although security is an important factor to achieve trust and is necessary for data protection in IoT, existing solutions require excessive computation and memory [70]. There is a need for new methods of detecting malware, technical standards, and minimum technical requirements modeling attacker's characteristics. One example discussed in [165] considers a use case scenario to autonomous vehicles as they heavily depend on multiple sensors. Another would be the use of smart door locks, which have recently attracted a lot of attention, with a market estimated at over $24 Billion [166], for having similar security flaws. Hence, solutions for IoT security is a major open issue.

Given the diverse nature of IoT devices in terms of communication protocols as well as other technical specifications, research aims to investigate new methods for security. One basic requirement for IoT security solutions is quick response time and scalability, given the real-time nature of many IoT devices. This implies that there is a need for lightweight solutions that detect and repair as part of self-healing architecture. In some cases, countermeasures need to be reprogrammed (e.g., when an unexpected attack occurs) and repair instructions need to be delivered securely to appropriate nodes so that the node's running programs could be amended by the runtime architecture. This solution might require additional hardware support for providing authentication, encryption and tamper-proof keys.

SDOs are developing a broad range of technical standards related to security, privacy and/or data protection of IoT. For instance, ITU-T X.iotsec-2 to provide a security framework for IoT analyzing security threats and challenges inherent to the IoT environment, and to describe security capabilities that could mitigate identified threats and challenges. ITU-T X.secup-iot on the other hand aims to provide a basic model for secure update of IoT software/ firmware and a common secure update procedure for IoT software/firmware implemented in IoT devices and systems.

The following needs to be taken into account to exploit the full potential of the IoT paradigm:

- One of the first topics that appeared in most of the analyzed research papers was on **terminology**. It was evident that various research papers have different definitions for similar IoT-related terms, including the description and definition of IoT itself [37] [164], highlighting the need for standardized structure and unified definitions. The international standard under development ISO/IEC 20924 *Information technology – Internet of Things – Definition and vocabulary* will provide a list of terms and concepts that are often used in the IoT context.
- In [167] the authors emphasize the need for having a **reference architecture** for IoT to support the security and privacy of the network. The need for a reference architecture is additionally supported by, for instance, [168] [169] and [170] proposing various architectures that could be used as a reference model for IoT systems, implying that this is an active direction of research for IoT. The recently published technical standard ISO/IEC 30141 *Information technology – Internet of Things – Reference architecture* provides functional view, system view, user view, information view and communication view of the IoT reference architecture.
- Another important topic constantly emphasized in research is how to tackle **interoperability challenges** that come with the diverse communication protocols and lack of standardized systems, as explained in Chapter 3 and highlighted in [171] [172] [173] [174] and [175]. The committee ISO/IEC JTC 1/SC 41 is working on the ISO/IEC 21823 series of standards concerning (framework, transport and semantic) interoperability in IoT.

5.3 Big data

On the one hand, research efforts as well as the requirements from businesses concerning Big data have been evolving over time and on the other hand, technical standards are providing a common framework to connect research and market needs. For instance:

- **Technical terminology:** unique and consistent terminology/vocabulary simplifies and increases the productivity of research works and businesses. To achieve this, ISO/IEC JTC 1/SC 42 has initiated a project (currently under development) ISO/IEC DIS 20546 that is providing standardized definitions and vocabulary, hence a common language.
- **Reference architecture:** Several Big data architecture models have been proposed in the scientific literature (e.g., [176] [177] [178] [179]). In order to build unified and consistent architecture, technical standards in this area could benefit research and business sectors. The ISO/IEC 20547 series has already published two standards related to use cases and derived requirements as well as standards roadmap. As part of this series, other standards such as ISO/IEC AWI TR 20547-1 *Framework and application process* and ISO/IEC DIS 20547-3 *Reference architecture* are currently under development.
- **Trustworthiness:** Data is the core component of Big data analysis and AI algorithms. A user might be reluctant to share personal information if she does not trust the analysis procedure or the outcomes. Some projects by ISO/IEC JTC 1/SC 42 have been initiated for addressing this issue such as the ISO/IEC AWI 20547-4 where a reference architecture is defined with relevant *security and privacy processes*.
- **Big data processing:** Recent years have witnessed research efforts on privacy by design and data protection by design. Similarly, techniques such as deep learning [180] are being used as a tool for training and learning in Big data processing and AI. The security, privacy and data protection in this context is an important dimension and relevant technical standards could be highly beneficial.

6

Conclusions

6. Conclusions

Security, privacy and data protection are becoming essential elements for building trust in ICT. These properties gain paramount importance specifically in the context of Smart ICT where an integration of Cloud computing, IoT and Big data is not only bringing technologies to different aspects of everyday lives but also providing tools for collecting, inferring and analyzing massive amounts of data from multiple (seemingly diverse) sources in order to present wide-ranging and deep insights.

In the past few years, the identification of potential risks and development of innovative solutions to protect users' data and privacy in Smart ICT domains has attracted an unparalleled attention of the scientific community, across the world. For instance, in the context of Cloud computing, research has focused on the challenges ranging from users' loss of control over data and applications, security flaws in the virtualization technology that could lead to information leakage, unavailability of data due to failures (e.g., server crashes, power outages), and security policy constraints. Similarly, solutions have been proposed to strengthen security and privacy at each architectural layer of IoT and Big data. For example, lightweight mechanisms to improve the security of IoT devices at the hardware layer and adaption of process isolation, access and information flow control, as well as methods for installing software updates at the IoT storage and processing layer have been proposed. On the other hand, privacy-enhancing techniques for data publication (e.g., K-anonymity, L-diversity and T-closeness) have been adopted for Big data and encryption techniques such as Homomorphic Encryption have been used for secure data storage and privacy-preserving data processing. These research efforts are reducing the hindrances in the widespread adoption of Smart ICT.

On the other hand, given that Smart ICT is becoming an important component of today's global economy as well as the society and life, a careful analysis and development of relevant technical standards has become necessary. For instance, a standard providing a privacy framework helps in establishing a common terminology, defining fundamentals of PII processing, and laying out privacy design principles. Similarly, a technical standard such as the ISO/IEC 27018:2014 on "Code of practice for protection of PII in public Clouds acting as PII processors" could provide good insights for privacy-preserving implementation of a Smart ICT application.

The European Commission, as of May 2018, has brought into force the General Data Protection Regulation (GDPR) that defines a set of data protection rules for all organizations operating in the EU. This regulation would mean that people will have more control over their personal data and businesses will benefit from a level playing field. The EU is also creating mechanisms for setting up a cybersecurity certification framework (the Cybersecurity Act[25]) for ICT products, services and processes to enhance cyber resilience. Within this framework, an issued certificate will be valid in all EU countries, making it easier for users to gain confidence in the security of these technologies and for organizations to carry out their businesses across borders. The CEN/CLC JTC 13 "Cybersecurity and data protection" is developing standards that could support relevant EU regulation, among others. The European agency for cybersecurity – ENISA – has also established liaison with ISO/IEC JTC 1/SC 27 in order to engage in and contribute to the development of the future standards.

Luxembourg is also creating an ecosystem that helps businesses of all sizes and focus areas in addressing the challenges concerning security, privacy and data protection, and is rapidly positioning itself to be at the forefront of the secure digital revolution. The national cybersecurity strategy aims at strengthening public confidence in digital environments, infrastructure protection, and promotion of the economy. University of Luxembourg and the SnT are performing cutting-edge research to improve the security, privacy and data protection capabilities of several emerging paradigms including Cloud computing, Big data, IoT, FinTech and communication technologies, to name a few. ILNAS – with the support of ANEC G.I.E. – is actively following standardization developments

[25] http://www.consilium.europa.eu/en/press/press-releases/2018/06/08/eu-to-create-a-common-cybersecurity-certification-framework-and-beef-up-its-agency-council-agrees-its-position/

related to security, privacy and data protection, particularly in the Smart ICT domains, within the framework of the national strategy and policy for ICT technical standardization[26]. In this context, among other projects, ILNAS fosters and strengthens the national ICT sector's involvement in standardization work through three leading projects: a) developing market interest and involvement, b) promoting and reinforcing market participation, and c) supporting and strengthening the education about standardization and related research activities.

Firstly, ILNAS is a member of recognized international and European standardization organizations and ensures that Luxembourg is well positioned in various international and EU developments. For instance, through ILNAS, Luxembourg is already a participating member of technical committees related to Smart Secure ICT (e.g., ISO/IEC JTC 1/SC 27, JTC 1/SC 38, JTC 1/SC 41, JTC 1/SC 42, CEN/CLC JTC 13 and ETSI TC CYBER), and all interested national stakeholders have the possibility to participate in the standards development process. Relevant information to become a delegate (e.g., of ISO/IEC JTC 1/SC 27) is available here[27]. One of the primary focus areas for ILNAS remains the Joint Technical Committee ISO/IEC JTC 1 since it develops most recognized ICT International Standards.

Secondly, ILNAS and University of Luxembourg (through SnT) have established a partnership in order to facilitate standards-related education and research. A university certificate program called "Smart ICT for Business Innovation"[28] is currently ongoing and it provides professional training at national level, while addressing Smart ICT across technical standardization as well as business innovation spectrum. Based on the experiences of the two editions of university certificate program (2016-2017, 2018-2019), ILNAS and University of Luxembourg aim to launch a full-fledged Master degree "Smart Secure ICT for Business Innovation" where security, privacy, data protection and technical standardization will be at the heart of the program and be taught transversely to various Smart ICT topics.

To ensure cutting-edge Master curriculum and to support the national market, ILNAS and University of Luxembourg have established a four-year research program dedicated to "Digital Trust for Smart ICT"[29]. This research program aims to build a solid base of knowledge and expertise in Smart ICT, taking into account the aspects related to digital trust and standardization, at all steps. In this context, ILNAS, ANEC G.I.E. and University of Luxembourg – in collaboration with the Ministry of the Economy – is publishing this white paper with the goal of providing a comprehensive view of data protection and privacy aspects (as enablers of digital trust) in Smart ICT from research as well as technical standardization perspectives. In this white paper:

- A Smart ICT data model defining the interactions between different Smart ICT domains is presented. This model analyzes how data serves as the common thread to all three Smart ICT domains and enables an understanding of the integrative Smart ICT components.
- The literature (research and scientific developments) concerning security, privacy and data protection in Smart ICT has been summarized.
- An overview of various developments in the areas of technical standardization has been provided. This includes details about technical committees and projects that focus on security, privacy and data protection and, relevant information about standardization activities in Cloud computing, Big data and IoT.
- Based on the above two points, some links between research developments and technical standardization projects have been highlighted.

As part of this research program, ILNAS and University of Luxembourg will strive to keep the contents of this white paper updated with latest developments, perform research activities to address state-of-the-art challenges, contribute to the technical standardization domain, and create a synchrony between scientific research and technical standards, specifically within the Smart Secure ICT framework.

[26] https://portail-qualite.public.lu/dam-assets/fr/publications/normes-normalisation/orientations-strategiques/politique-luxembourgeoise-pour-la-normalisation-technique-des-tic-2015-2020/policy-ict-technical-standardization-2015-2020.pdf

[27] https://portail-qualite.public.lu/fr/normes-normalisation/participer-normalisation/experts-normalisation.html

[28] https://portail-qualite.public.lu/fr/normes-normalisation/education-recherche/projets-phares-dans-l_education-a-la-normalisation.html

[29] https://portail-qualite.public.lu/fr/normes-normalisation/education-recherche/programme-recherche.html

References

[1] ILNAS, White Paper - Digital Trust for Smart ICT, Luxembourg, 2016.

[2] "Data protection and privacy," Commission Nationale pour la Protection des Données (CNPD), Luxembourg.

[3] S. D. C. d. Vimercati, S. Foresti and P. Samarati, "Managing and accessing data in the cloud: Privacy risks and approaches," in *7th International Conference on Risk and Security of Internet and Systems (CRiSIS)*, 2012.

[4] "ISO/IEC 17788:2014 Information technology -- Cloud computing -- Overview and vocabulary," [Online]. Available: https://www.iso.org/standard/60544.html. [Accessed August 2018].

[5] A. Botta, W. d. Donato, V. Persico and A. Pescapé, "Integration of Cloud computing and Internet of Things: A survey," *Future Generation Computer Systems*, vol. 56, pp. 684-700, 2016.

[6] P. Mell and T. Grance, "The NIST Definition of Cloud Computing," National Institute of Standards and Technology, 2011.

[7] ILNAS, White paper - Internet of Things: Technology, Economic view and Technical Standardization, Luxembourg, 2018.

[8] "Gartner Report," https://www.gartner.com/doc/3841268/forecast-analysis-internet-things, January 2017.

[9] J. Greenough, "The Internet of Things is Rising: How the IoT Market will Grow across sectors," Business Insider Intelligence, October 2014.

[10] A. Thierer and A. Castillo, "Projecting growth and Economic Impact of Internet of Things," The Mercatus Center at George Mason University, June, 2015.

[11] E. Fernandes, A. Rahmati, K. Eykholt and A. Prakash, "Internet of Things Security Research: A Rehash of Old Ideas or New Intellectual Challenges?," *IEEE Security & Privacy*, vol. 15, no. 4, 5 2017.

[12] M. Cox and D. Ellsworth, "Managing big data for scientific visualization," 1997.

[13] D. Laney, "3D Data Management: Controlling Data Volume, Velocity and Variety," META Group, 2001.

[14] P. Zikopoulos, D. Deroos, K. Parasuraman, T. Duetsch, D. Corrigan and J. Giles, "Harness the power of Big Data - The IBM Big Data Platform," McGraw Hill.

[15] E. Curry, "The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches," in *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe*, Springer, 2016, pp. 29-37.

[16] "A European strategy on the data value chain" [Online]. Available: http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=3488. [Accessed August 2018].

[17] J. M. Cavanillas, E. Curry and W. Wahlster, New Horizons for a Data-Driven Economy - A roadmap for usage and exploitation of Big data in Europe, Springer.

[18] CSA, "The Treacherous 12 - Cloud Computing Top Threats in 2016," Cloud Security Alliance, 2016.

[19] D. Mendez, I. Papapanagiotou and B. Yang, "Internet of Things: Survey on Security and Privacy," *Information Security Journal: A Global Perspective (CoRR)*, 2017.

[20] J. Soria-Comas and J. Domingo-Ferrer, "Big Data Privacy: Challenges to Privacy Principles and Models," *Data Science and Engineering*, vol. 1, no. 1, pp. 21-28, 2016.

[21] "ISO/IEC 27017:2015 Information technology -- Security techniques -- Code of practice for information security controls based on ISO/IEC 27002 for cloud services," [Online]. Available: https://www.iso.org/standard/43757.html. [Accessed August 2018].

[22] "ISO/IEC 19086-1:2016 Information technology -- Cloud computing -- Service level agreement (SLA) framework -- Part 1: Overview and concepts," [Online]. Available: https://www.iso.org/standard/67545.html. [Accessed August 2018].

[23] "ISO/IEC 19086-3:2017 Information technology -- Cloud computing -- Service level agreement (SLA) framework -- Part 3: Core conformance requirements," [Online]. Available: https://www.iso.org/standard/67547.html. [Accessed August 2018].

[24] "ISO/IEC 27001:2013 Information technology -- Security techniques -- Information security management systems -- Requirements," [Online]. Available: https://www.iso.org/standard/54534.html. [Accessed July 2018].

[25] "ISO/IEC 27002:2013 Information technology -- Security techniques -- Code of practice for information security controls," [Online]. Available: https://www.iso.org/standard/54533.html. [Accessed August 2018].

[26] "ISO/IEC 29100:2011 Information technology -- Security techniques -- Privacy framework," [Online]. Available: https://www.iso.org/standard/45123.html. [Accessed August 2018].

[27] M. Strohbach, H. Ziekow, V. Gazis and N. Akiva, "Towards a Big Data Analytics Framework for IoT and Smart City Applications," in *Modeling and Processing for Next-Generation Big Data Technologies*, Springer, 2015, pp. 257-282.

[28] E. Cavalcante, J. Pereira, M. Pitanga-Alves, P. Maia, R. Moura, T. Batista, F. Delicato and P. Pires, "On the interplay of Internet of Things and Cloud Computing: A systematic mapping study," *Computer Communications*, Vols. 89-90, pp. 17-33, 2016.

[29] S. K. Dash, S. Mohapatra and P. K. Pattnaik, "A survey on application of wireless sensor network using Cloud computing," *International Journal on Computer Science Engineering and Technology*, vol. 1, no. 4, pp. 50-55, 2010.

[30] G. C. Fox, S. Kamburugamuve and R. D. Hartman, "Architecture and measured characteristics of a Cloud based Internet of Things," in *IEEE International Conference on Collaboration Technologies and Systems*, 2012.

[31] D. Yao, C. Yu, H. Jin and J. Zhou, "Energy efficient task scheduling in mobile Cloud computing," *Network and Parallel Computing - Springer*, pp. 344-355, 2013.

[32] K. Jeffery, "Keynote: CLOUDS: A large virtualisation of small things," in *The 2nd International Conference on Future Internet of Things and Cloud*, 2014.

[33] ILNAS, White Paper Big Data, Luxembourg, April 2016.

[34] M. D. Assuncao, R. N. Calheiros, S. Bianchi, M. A. Netto and R. Buyya, "Big Data computing and clouds: Trends and future directions," *Journal of Parallel and Distributed Computing*, pp. 3-15, 2015.

[35] J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," Google Inc..

[36] K. H. Lee, Y. J. Lee, H. Choi, Y. D. Chung and B. Moon, "Parallel Data Processing with MapReduce: A Survey," *SIGMOD*, vol. 40, no. 4, pp. 11-20, 2011.

[37] R. Minerva, A. Biru and D. Rotondi, "Towards a definition of the Internet of Things (IoT)," Issue 1 - IEEE, May 2015.

[38] Y. Qin, Q. Z. Sheng, N. J. Falkner, S. Dustdar, H. Wang and A. V. Vasilakos, "When Things Matter: A Survey on Data-Centric Internet of Things," *Journal of Network and Computer Applications*, 2016.

[39] "Microsoft Azure IoT Reference Architecture," Microsoft, 2018.

[40] "Overview of Internet of Things," Google Cloud, [Online]. Available: https://cloud.google.com/solutions/iot-overview. [Accessed July 2018].

[41] D. Gil, A. Ferràndez, H. Mora-Mora and J. Peral, "Internet of Things: A Review of Surveys Based on Context Aware Intelligent Services," Senors - MDPI, 2016.

[42] N. Mathur and R. Purohit, "Issues and Challenges in Convergence of Big Data, Cloud and Data Science," *International Journal of Computer Applications*, vol. 160, no. 9, 2017.

[43] Y. Qin, Q. Z. Sheng, N. J. Falkner, S. Dustdar, H. Wang and A. V. Vasilakos, "When Things Matter: A Survey on Data-Centric Internet of Things," *Journal of Network and Computer Applications*, vol. 64, pp. 137-153, 2016.

[44] A. Gholami and E. Laure, "Big data security and privacy issues in the Cloud," *International Journal of Network Security & its Applications (IJNSA)*, vol. 8, no. 1, 2016.

[45] M. Stihler, A. O. Santin, A. L. Marcon and J. Silva Fraga, "Integral Federated Identity Management for Cloud Computing," in 2012 *5th International Conference on New Technologies, Mobility and Security (NTMS)*, 2012.

[46] B. Keltoum and B. Samia, "A dynamic federated identity management approach for cloud-based environments," in *Proceedings of the Second International Conference on Internet of things, Data and Cloud Computing* - ICC '17, 2017.

[47] K. Bendiab, S. Shiaeles and S. Boucherkha, "A New Dynamic Trust Model for ``On Cloud" Federated Identity Management," in 2018 *9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, 2018.

[48] N. M. Gonzalez, M. A. T. Rojas, M. V. M. Silva, F. Redigolo, T. C. M. Brito Carvalho, C. C. Miers, M. Naslund and A. S. Ahmed, "A Framework for Authentication and Authorization Credentials in Cloud Computing," in *12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, 2013.

[49] R. Banyal, P. Jain and V. Jain, "Multi-factor authentication framework for Cloud Computing," in *CIMSim*.

[50] R. Lomotey and R. Deters, "Saas authentication middleware for mobile consumers of Iaas Cloud," in *SERVICES IEEE*, 2013.

[51] J. Sendor, Y. Lehmann, G. Serme and A. S. Oliveira, "Platform level support for authorization in Cloud services with oauth 2," in *IC2E IEEE*, 2014.

[52] C. Jincui and J. Liqun, "Role-Based Access Control Model of Cloud Computing," *Energy Procedia*, vol. 13, pp. 1056-1061, 2011.

[53] L. Zhou, V. Varadharajan and M. Hitchens, "Integrating trust with cryptographic role-based access control for secure Cloud data storage," in *IEEE TrustCom*.

[54] B. Tang, R. Sandhu and Q. Li, "Multi-tenancy authorization models for collaborative Cloud services," in *CTS*.

[55] A. Ben Fadhel, D. Bianculli and L. Briand, "A comprehensive modeling framework for role-based access control policies," *J. Syst. Softw*., vol. 107, pp. 110-126, 2015.

[56] F. Jaidi, F. L. Ayachi and A. Bouhoula, "A Comprehensive Formal Solution for Access Control Policies Management: Defect Detection, Analysis and Risk Assessment".

[57] I. Saenko and I. Kotenko, "Using Genetic Algorithms for Design and Reconfiguration of RBAC Schemes," in *Proceedings of the 1st International Workshop on AI for Privacy and Security - PrAISe '16*, 2016.

[58] A. Brinkmann, C. Fiehe, A. Litvina, I. Lück, L. Nagel, K. Narayanan, F. Ostermair and W. Thronicke, "Scalable monitoring system for Clouds," in *UCC*.

[59] H. Kuijs, C. Reich, M. Knahl and N. Clarke, "A Scalable Architecture for Distributed OSGi in the Cloud," in *Proceedings of the 6th International Conference on Cloud Computing and Services Science*, 2016.

[60] S. Fischer-Hübner, J. Angulo and T. Pulls, "How can Cloud Users be Supported in Deciding on, Tracking and Controlling How their Data are Used?," in *IFIP Advances in Information and Communication Technology*, 2014, pp. 77-92.

[61] J. R. Raphael, *The worst Cloud outages of 2013 (so far)*, 2013.

[62] R. A. Popa, J. R. Lorch, D. Molnar, H. J. Wang and L. Zhuang, "Enabling security in Cloud storage slas with Cloudproof," in *USENIX ATC*, 2011.

[63] S. Zhu and G. Gong, "Fuzzy authorization for Cloud storage," in *IEEE Transactions on Cloud Computing (TCC)*, 2014.

[64] P. Voigt and A. Bussche, "Scope of Application of the GDPR," in *The EU General Data Protection Regulation (GDPR)*, 2017, pp. 9-30.

[65] D. Choi, S.-H. Jin and H. Yoon, "Trust Management for User-Centric Identity Management on the Internet," in *2007 IEEE International Symposium on Consumer Electronics*, 2007.

[66] M. Berdufi, Trust management in a multicloud computing environment, Università degli Studi di Camerino, 2016.

[67] R. K. Kalluri and C. Rao, "Addressing the security, privacy and trust challenges of Cloud computing," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 5, pp. 6094-6097, 2014.

[68] Y. Zhang and J. Joshi, "Access Control and Trust Management for Emerging Multi-domain Environments," *Annals of Emerging Research in Information Assurance, Security and Privacy Services*, pp. 421-452, 2009.

[69] B. Russell, *Realizing Linux Containers (LXC)*, 2015.

[70] Y. Yang, L. Wu, G. Yin, L. Li and H. Zhao, "A Survey on Security and Privacy Issues in Internet-of-Things," *IEEE Internet of Things Journal*, vol. 4, pp. 1250-1258, 10 2017.

[71] G. Hunt, G. Letey and E. Nightingale, "The seven properties of highly secure devices," *tech. report MSR-TR-2017-16*, 2017.

[72] K. Yang, M. Hicks, Q. Dong, T. Austin and D. Sylvester, "A2: Analog Malicious Hardware," in 2016 *IEEE Symposium on Security and Privacy (SP)*, 2016.

[73] A. Antonopoulos, C. Kapatsori and Y. Makris, "Hardware Trojans in Analog, Mixed-Signal, and RF ICs," in *The Hardware Trojan War: Attacks, Myths, and Defenses*, S. Bhunia and M. M. Tehranipoor, Eds., Cham, Springer International Publishing, 2018, pp. 101-123.

[74] S. Sicari, A. Rizzardi, L. A. Grieco and A. Coen-Porisini, "Security, privacy and trust in Internet of Things: The road ahead," *Computer Networks*, vol. 76, pp. 146-164, 1 2015.

[75] A. Rahmati, M. Salajegheh, D. Holcomb, J. Sorber, W. P. Burleson and K. Fu, "TARDIS: time and remanence decay in SRAM to implement secure protocols on embedded devices without clocks," in *Proceedings of the 21st USENIX conference on Security symposium*, Berkeley, 2012.

[76] E. Fernandes, J. Paupore, A. Rahmati, D. Simionato, M. Conti and A. Prakash, "FlowFence: Practical Data Protection for Emerging IoT Application Frameworks," in *USENIX Security Symposium*, 2016.

[77] M. Giannikos, K. Kokoli, N. Fotiou, G. F. Marias and G. C. Polyzos, "Towards secure and context-aware information lookup for the Internet of Things," in 2013 *International Conference on Computing, Networking and Communications (ICNC)*, 2013.

[78] B. Tepekule, U. Yavuz and A. E. Pusane, "On the use of modern coding techniques in QR applications," in *2013 21st Signal Processing and Communications Applications Conference (SIU)*, 2013.

[79] L. Zhou, Q. Wen and H. Zhang, "Preserving Sensor Location Privacy in Internet of Things," in *2012 Fourth International Conference on Computational and Information Sciences*, 2012.

[80] T. Yu, V. Sekar, S. Seshan, Y. Agarwal and C. Xu, "Handling a Trillion (Unfixable) Flaws on a Billion Devices: Rethinking Network Security for the Internet-of-Things," in *Proceedings of the 14th ACM Workshop on Hot Topics in Networks*, New York, NY, USA, 2015.

[81] A. K. Simpson, F. Roesner and T. Kohno, "Securing vulnerable home IoT devices with an in-hub security manager," in *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, 2017.

[82] D. Formby, P. Srinivasan, A. Leonard, J. Rogers and R. A. Beyah, "Who's in Control of Your Control System? Device Fingerprinting for Cyber-Physical Systems," in *NDSS*, 2016.

[83] J. S. Kumar and D. R. Patel, "A survey on internet of things: Security and privacy issues," *Int. J. Comput. Appl. Technol.*, vol. 90, 2014.

[84] E. Borgia, "The Internet of Things vision: Key features, applications and open issues," *Comput. Commun.*, vol. 54, pp. 1-31, 12 2014.

[85] A. Levy, M. P. Andersen, B. Campbell, D. Culler, P. Dutta, B. Ghena, P. Levis and P. Pannuto, "Ownership is Theft: Experiences Building an Embedded OS in Rust," in *Proceedings of the 8th Workshop on Programming Languages and Operating Systems*, New York, NY, USA, 2015.

[86] E. Bertino, S. Calo, H. Chen, N. Li, T. Li, J. Lobo, I. Molloy and Q. Wang, "Some usability considerations in access control systems," in *Symposium on Usable Privacy and Security (SOUPS) 2008*, 2008.

[87] A. P. Felt, E. Ha, S. Egelman, A. Haney, E. Chin and D. Wagner, "Android Permissions: User Attention, Comprehension, and Behavior," in *Proceedings of the Eighth Symposium on Usable Privacy and Security*, New York, NY, USA, 2012.

[88] N. Papernot, P. McDaniel, A. Sinha and M. Wellman, "Towards the Science of Security and Privacy in Machine Learning," 11 2016.

[89] L. Xu, C. Jiang, J. Wang, J. Yuan and Y. Ren, "Information Security in Big Data: Privacy and Data Mining," *IEEE Access*, vol. 2, pp. 1149-1176, 2014.

[90] A. Evfimievski, J. Gehrke and R. Srikant, "Limiting privacy breaches in privacy preserving data mining," in *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems - PODS '03*, New York, New York, USA, 2003.

[91] A. Juels and T. Ristenpart, "Honey Encryption: Encryption beyond the Brute-Force Barrier," *IEEE Security Privacy*, vol. 12, pp. 59-62, 7 2014.

[92] Z. Huang, E. Ayday, J. Fellay, J.-P. Hubaux and A. Juels, "Genoguard: Protecting genomic data against brute-force attacks," 2015.

[93] R. Chatterjee, J. Bonneau, A. Juels and T. Ristenpart, "Cracking-Resistant Password Vaults Using Natural Language Encoders," in *2015 IEEE Symposium on Security and Privacy*, 2015.

[94] C. Benjamin, M. Fung, K. Wang, R. Chen and S. Philip, "Yu Privacy-preserving data publishing: A survey of recent developments," *ACM Comput. Surv.*, pp. 1-53, 2010.

[95] R. C.-W. Wong and A. W.-C. Fu, "Privacy-Preserving Data Publishing: An Overview," *Synthesis Lectures on Data Management*, vol. 2, pp. 1-138, 1 2010.

[96] Z. He, "Privacy Preserving Data Publishing," 2018.

[97] X. Liu, Q. Xie and L. Wang, "Personalized extended (α , k)-anonymity model for privacy-preserving data publishing : PERSONALIZED EXTENDED (α , k)-ANONYMITY MODEL," *Concurr. Comput.*, vol. 29, p. e3886, 3 2017.

[98] K. Ito, J. Kogure, T. Shimoyama and H. Tsuda, "De-identification and Encryption Technologies to Protect Personal Information," *Fujitsu Sci. Tech. J.*, vol. 52, pp. 28-36, 2016.

[99] A. Machanavajjhala, J. Gehrke, D. Kifer and M. Venkitasubramaniam, "ell-Diversity: Privacy Beyondkappa-Anonymity," in *null*, 2006.

[100] K. El Emam, L. Arbuckle, G. Koru, B. Eze, L. Gaudette, E. Neri, S. Rose, J. Howard and J. Gluck, "De-identification methods for open health data: the case of the Heritage Health Prize claims dataset," *J. Med. Internet Res.*, vol. 14, p. e33, 2 2012.

[101] K. El Emam, "Methods for the de-identification of electronic health records for genomic research," *Genome Med.*, vol. 3, p. 25, 4 2011.

[102] N. Li, T. Li and S. Venkatasubramanian, "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity," in *2007 IEEE 23rd International Conference on Data Engineering*, 2007.

[103] V. Khanaa, Dean-Information., B. University, R. Udayakumar, A. Prof. and I. T. B. University, "Protecting Privacy When Disclosing Information: K Anonymity and its Enforcement Through Suppression," *International Journal of Business Intelligents*, vol. 001, pp. 28-31, 2012.

[104] P. Jain, M. Gyanchandani and N. Khare, "Big data privacy: a technological perspective and review," *Journal of Big Data*, vol. 3, no. 25, 2016.

[105] K. Rajendran, M. Jayabalan and M. E. Rana, "A Study on k-anonymity, l-diversity, and t-closeness Techniques," *IJCSNS*, vol. 17, p. 172, 2017.

[106] M. M. Groat, W. Hey and S. Forrest, "KIPDA: k-indistinguishable privacy-preserving data aggregation in wireless sensor networks," in *2011 Proceedings IEEE INFOCOM*, 2011.

[107] T. S. Gal, Z. Chen and A. Gangopadhyay, "A Privacy Protection Model for Patient Data with Multiple Sensitive Attributes," *IJISP*, vol. 2, pp. 28-44, 7 2008.

[108] X.-M. He, X. S. Wang, D. Li and Y.-N. Hao, "Semi-Homogenous Generalization: Improving Homogenous Generalization for Privacy Preservation in Cloud Computing," *J. Comput. Sci. Technol.*, vol. 31, pp. 1124-1135, 11 2016.

[109] A. R. Beresford and F. Stajano, "Location privacy in pervasive computing," *IEEE Pervasive Comput.*, pp. 46-55, 2003.

[110] F. Liu, K. A. Hua and Y. Cai, "Query l-diversity in Location-Based Services," in *2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware*, 2009.

[111] D. Riboni, L. Pareschi, C. Bettini and S. Jajodia, "Preserving Anonymity of Recurrent Location-Based Queries," in *2009 16th International Symposium on Temporal Representation and Reasoning*, 2009.

[112] N. S. Kumar, G. V. R. Lakshmi and B. Balamurugan, "Enhanced Attribute Based Encryption for Cloud Computing," *Procedia Comput. Sci.*, vol. 46, pp. 689-696, 1 2015.

[113] A. Acar, H. Aksu, A. S. Uluagac and M. Conti, "A Survey on Homomorphic Encryption Schemes: Theory and Implementation," *ACM Comput. Surv.*, vol. 51, pp. 79:1--79:35, 7 2018.

[114] D. Lubicz and T. Sirvent, "Attribute-Based Broadcast Encryption Scheme Made Efficient," in *Progress in Cryptology -- AFRICACRYPT 2008*, 2008.

[115] Homomorphic encryption and Bitcoin [Online]. Available: https://www.lesswrong.com/posts/XCuwfWFuiGxCWXFtW/homomorphic-encryption-and-bitcoin. [Accessed August 2018].

[116] E. Ayday, J. L. Raisaro, M. Laren, P. Jack, J. Fellay and J.-P. Hubaux, "Privacy-preserving computation of disease risk by using genomic, clinical, and environmental data," in *Proceedings of USENIX Security Workshop on Health Information Technologies (HealthTech'13)*, 2013.

[117] M. Hussain and M. Hussain, *Advance Applications of Identity Based Encryption*.

[118] H. Vaghashia and A. Ganatra, "A Survey: Privacy Preservation Techniques in Data Mining," *Int. J. Comput. Appl. Technol.*, vol. 119, pp. 20-26, 2015.

[119] Y. Lindell and B. Pinkas, "Privacy preserving data mining," in *Annual International Cryptology Conference*, 2000.

[120] R. Agrawal and R. Srikant, "Privacy-preserving data mining," *ACM*, 2000.

[121] Y. Lindell and B. Pinkas, "Secure Two-Party Computation via Cut-and-Choose Oblivious Transfer," *J. Cryptology*, vol. 25, pp. 680-722, 10 2012.

[122] V. S. Verykios, E. Bertino, I. N. Fovino, L. P. Provenza, Y. Saygin and Y. Theodoridis, "State-of-the-art in privacy preserving data mining," *SIGMOD* Rec., vol. 33, p. 50, 3 2004.

[123] ILNAS, "Standards Analysis Smart ICT - Luxembourg," 2018. [Online]. Available: https://portail-qualite.public.lu/content/dam/qualite/publications/normalisation/2018/standards-analysis-smart-ict-2-0.pdf.

[124] "Cooperation between ISO, IEC, CEN and CENELEC," [Online]. Available: https://www.cencenelec.eu/intcoop/StandardizationOrg/Pages/default.aspx. [Accessed July 2018].

[125] "ISO/IEC JTC 1/SC 27 IT Security Techniques," [Online]. Available: https://www.iso.org/committee/45306.html. [Accessed August 2018].

[126] "ISO/IEC 29101:2013 Information technology -- Security techniques -- Privacy architecture framework," [Online]. Available: https://www.iso.org/standard/45124.html. [Accessed August 2018].

[127] "ISO/IEC 29134:2017 Information technology -- Security techniques -- Guidelines for privacy impact assessment," [Online]. Available: https://www.iso.org/standard/62289.html. [Accessed August 2018].

[128] "ISO/IEC 29151:2017 Information technology -- Security techniques -- Code of practice for personally identifiable information protection," [Online]. Available: https://www.iso.org/standard/62726.html. [Accessed August 2018].

[129] "ISO/IEC 29190:2015 Information technology -- Security techniques -- Privacy capability assessment model," [Online]. Available: https://www.iso.org/standard/45269.html. [Accessed August 2018].

[130] "ISO/IEC 29146:2016 Information technology -- Security techniques -- A framework for access management," [Online]. Available: https://www.iso.org/standard/45169.html. [Accessed August 2018].

[131] "ISO/IEC 29191:2012 Information technology -- Security techniques -- Requirements for partially anonymous, partially unlinkable authentication," [Online]. Available: https://www.iso.org/standard/45270.html. [Accessed August 2018].

[132] "ISO/IEC 24760-1:2011 Information technology -- Security techniques -- A framework for identity management -- Part 1: Terminology and concepts," [Online]. Available: https://www.iso.org/standard/57914.html. [Accessed July 2018].

[133] "ISO/IEC 24760-2:2015 Information technology -- Security techniques -- A framework for identity management -- Part 2: Reference architecture and requirements," [Online]. Available: https://www.iso.org/standard/57915.html. [Accessed July 2018].

[134] "ISO/IEC 24760-3:2016 Information technology -- Security techniques -- A framework for identity management -- Part 3: Practice," [Online]. Available: https://www.iso.org/standard/57916.html. [Accessed July 2018].

[135] "CEN/CLC JTC 13 Cybersecurity and data protection," [Online]. Available: https://www.cenelec.eu/dyn/www/f?p=104:7:1375868458618501::::FSP_LANG_ID,FSP_ORG_ID:25,2307986#1. [Accessed August 2018].

[136] "CEN/CLC/TC 8 Pricacy management in porducts and services," [Online]. Available: https://www.cencenelec.eu/standards/Sectors/DefenceSecurityPrivacy/Privacy/Pages/default.aspx. [Accessed August 2018].

[137] "ETSI/TC CYBER," [Online]. Available: https://www.etsi.org/technologies-clusters/technologies/cyber-security. [Accessed August 2018].

[138] "ITU-T SG 17 Security," [Online]. Available: https://www.itu.int/en/ITU-T/about/groups/Pages/sg17.aspx. [Accessed August 2018].

[139] "ISO/IEC 17826:2016 Information technology -- Cloud Data Management Interface (CDMI)," [Online]. Available: https://www.iso.org/standard/70226.html. [Accessed August 2018].

[140] "ISO/IEC 19944:2017 Information technology -- Cloud computing -- Cloud services and devices: Data flow, data categories and data use," [Online]. Available: https://www.iso.org/standard/66674.html. [Accessed August 2018].

[141] "ISO/IEC 19941:2017 Information technology -- Cloud computing -- Interoperability and portability," [Online]. Available: https://www.iso.org/standard/66639.html. [Accessed August 2018].

[142] "ISO/IEC 17789:2014 Information technology -- Cloud computing -- Reference architecture," [Online]. Available: https://www.iso.org/standard/60545.html. [Accessed August 2018].

[143] "ISO/IEC 27018:2014 Information technology -- Security techniques -- Code of practice for protection of personally identifiable information (PII) in public clouds acting as PII processors," [Online]. Available: https://www.iso.org/standard/61498.html. [Accessed August 2018].

[144] "ISO/IEC PDTR 23186 Information technology -- Cloud computing -- Framework of trust for processing of multi-sourced data," [Online]. Available: https://www.iso.org/standard/74844.html. [Accessed September 2018].

[145] "Terms of Reference for Technical Committee Smart Machine-to-Machine communications," ETSI, [Online]. Available: https://portal.etsi.org/tbsitemap/smartm2m/smartm2mtor.aspx. [Accessed August 2018].

[146] "ETSI/TC Smart M2M," [Online]. Available: https://portal.etsi.org/tbsitemap/smartm2m/smartm2mtor.aspx. [Accessed August 2018].

[147] "oneM2M; Developer guide: Implementing security example (oneM2M TR-0038 v2.0.0 release 2A)," ETSI, 2017.

[148] "ISO/IEC JTC 1/SC 31 Automatic identification and data capture techniques," [Online]. Available: https://www.iso.org/committee/45332.html. [Accessed August 2018].

[149] "CEN/TC 225 - AIDC technologies," [Online]. Available: https://standards.cen.eu/dyn/www/f?p=204:7:0::::FSP_ORG_ID:6206&cs=1E12277AECC001196A7556B8DBCDF0A1C. [Accessed August 2018].

[150] "ISO/IEC TR 22417:2017 Information technology -- Internet of things (IoT) use cases," [Online]. Available: https://www.iso.org/standard/73148.html. [Accessed July 2018].

[151] "ISO/IEC JTC 1/SC 25 Interconnection of information technology equipment," [Online]. Available: https://www.iso.org/committee/45270.html. [Accessed August 2018].

[152] "ISO/IEC JTC 1/SC 42 Artificial Intelligence," [Online]. Available: https://www.iso.org/committee/6794475.html. [Accessed August 2018].

[153] "ISO/IEC TR 20547-2:2018 Information technology -- Big data reference architecture -- Part 2: Use cases and derived requirements," [Online]. Available: https://www.iso.org/standard/71276.html. [Accessed July 2018].

[154] "ISO/IEC TR 20547-5:2018 Information technology -- Big data reference architecture -- Part 5: Standards roadmap," [Online]. Available: https://www.iso.org/standard/72826.html. [Accessed July 2018].

[155] "ISO/IEC JTC1/SC 32 Data management and Interchange," [Online]. Available: https://www.iso.org/committee/45342.html. [Accessed August 2018].

[156] "ISO/TC 69 Applications of statistical methods," [Online]. Available: https://www.iso.org/committee/49742.html. [Accessed August 2018].

[157] "ISO/IEC JTC 1/SC 7 Software and systems engineering," [Online]. Available: https://www.iso.org/committee/45086.html. [Accessed August 2018].

[158] "ISO/IEC 25012:2008 Software engineering -- Software product Quality Requirements and Evaluation (SQuaRE) -- Data quality model," [Online]. Available: https://www.iso.org/standard/35736.html. [Accessed July 2018].

[159] J. Metadata, "N2393 Progress report on Metadata for Big Data Quality".

[160] "ISO/IEC JTC 1/SC 27 N18548 Proposal for new work item on Big data security and privacy - Processes (ISO/IEC NP 27045)".

[161] ILNAS, "Standards Analysis ICT Sector (version 8.0)," November 2017. [Online]. Available: https://portail-qualite.public.lu/dam-assets/publications/normalisation/2017/standards-analysis-ict-8-0.pdf. [Accessed August 2018].

[162] AWS Documentation Team, "AWS Identity and Access Management User Guide: AWS IAM User Guide," 2018.

[163] M. Nitti, R. Girau and L. Atzori, "Trustworthiness Management in the Social Internet of Things," IEEE Transactions on Knowledge and Data Engineering, vol. 26, no. 5, pp. 1253-1266, 2014.

[164] J. Voas, "Demystifying the Internet of Things," in Demystifying the Internet of Things, 2016, pp. 80-83.

[165] A. Lima, F. Rocha, M. Völp and P. Esteves-Veríssimo, "Towards Safe and Secure Autonomous and Cooperative Vehicle Ecosystems," in 2Nd ACM Workshop on Cyber-Physical Systems Security and Privacy, Vienna, Austria, 2016.

[166] "Smart Lock Market Size, Share, Analysis | Industry Report, 2018-2024," [Online]. Available: https://www.grandviewresearch.com/industry-analysis/smart-lock-market. [Accessed 25 07 2018].

[167] I. Addo, S. Ahamed, S. Yau and A. Buduru, "A Reference Architecture for Improving Security and Privacy in Internet of Things Applications," IEEE International Conference on Mobile Services, pp. 108-115, 2014.

[168] S. Krčo, B. Pokrić and F. Carrez, "Designing IoT architecture(s): A European perspective," IEEE World Forum on Internet of Things (WF-IoT), pp. 79-84, 2014.

[169] M. Weyrich and C. Ebert, "Reference Architectures for the Internet of Things," IEEE Software, vol. 33, no. 1, pp. 112-116, 2016.

[170] A. Barba and F. de C. Giorno, "A Reference Architecture for the IoT Services' Adaptability - Using Agents to Make IoT Services Dynamically Reconfigurable," in 3rd International Conference on Internet of Things, Big Data and Security, Funchal, Madeira, Portugal, 2018.

[171] M. Blackstock and R. Lea, “IoT interoperability: A hub-based approach,” in International Conference on the Internet of Things (IoT), 2014.

[172] P. Desai, A. Sheth and P. Anantharam, “Semantic Gateway as a Service Architecture for IoT Interoperability,” IEEE International Conference on Mobile Services, pp. 313-319, 2015.

[173] G. Aloi and al., “A Mobile Multi-Technology Gateway to Enable IoT Interoperability,” IEEE First International Conference on Internet-of-Things Design and Implementation (IoTDI), pp. 259-264, 2016.

[174] C. Pereira, A. Pinto, A. Aguiar, P. Rocha, F. Santiago and J. Sousa, “IoT interoperability for actuating applications through standardised M2M communications,” IEEE 17th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), pp. 1-6, 2016.

[175] O. Vermesan, “Advancing IoT Platforms Interoperability, vol.1,” River Publisher, pp. 1-92.

[176] “Big data architecture and patterns, Part 1 Introduction to big data classification and architecture,” IBM, [Online]. Available: https://www.ibm.com/developerworks/library/bd-archpatterns1/index.html. [Accessed August 2018].

[177] “Big data architectures,” Microsoft Azure, 2017. [Online]. Available: https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/.

[178] “Modern Data Architecture with Apache Hadoop,” Hortonworks and Attunity, [Online]. Available: http://hortonworks.com/wp-content/uploads/2012/06/Hortonworks-Attunity-whitepaper.pdf. [Accessed August 2018].

[179] X. Zheng, M. Fu and M. Chugh, “Big data storage and management in SaaS applications,” Journal of Communications and Information Networks, vol. 2, no. 3, 2017.

[180] J. Schmidhuber, “Deep learning in neural networks: An overview,” Neural networks, vol. 61, no. 2015, pp. 85-117.

October 2018 · Data protection and Privacy in Smart ICT · Version 1.0 · © ILNAS/UNI.LU/ANEC/SnT

ILNAS
Institut Luxembourgeois de la
Normalisation, de l'Accréditation, de la
Sécurité et qualité des produits et services

uni.lu
UNIVERSITÉ DU
LUXEMBOURG

Southlane Tower I · 1, avenue du Swing · L-4367 Belvaux · Tel. : (+352) 24 77 43 -70 · Fax : (+352) 24 79 43 -70 · E-mail : info@ilnas.etat.lu

www.portail-qualite.lu