CLOUD
ACCOUNTABILITY
PROJECT

D:C-5.2 Validation of the accountability metrics

Deliverable Number	D35.2
Work Package	WP 35
Version	Final
Deliverable Lead Organisation	UMA
Dissemination Level	PU
Contractual Date of Delivery (release)	30/09/2014
Date of Delivery	06/10/2014

Editor

David Nuñez (UMA), Carmen Fernandez-Gago (UMA)

Contributors

David Nuñez (UMA), Carmen Fernandez-Gago (UMA), Isaac Agudo (UMA), Jesús Luna (CSA), Alain Pannetrat (CSA)

Reviewers

Vasilis Tountopoulos (ATC), Massimo Felici (HP)

A4Cloud
www.a4cloud.eu
Accountability For Cloud and Other Future Internet Services
FP7-ICT-2011-8-317550-A4CLOUD

SEVENTH FRAMEWORK PROGRAMME

Executive Summary

One of the objectives of A4Cloud is to develop measurement techniques for supporting the concept of Accountability. Essentially, what is needed for ensuring accountability is to be able to demonstrate that the accounts provided by an organisation (to regulators, auditors, data subjects or other service providers) are adequate and appropriate for the context by meeting certain internal and external criteria, and to have in place mechanisms for dealing with the situation (including sanctions and other measures possibly leading to the remediation of failures), if this is not the case. From an organisational point of view the focus is on measuring whether the fundamental types of activities that an accountable organisation should undertake are in place and effective.

The definition and utilization of meaningful metrics is among the mechanisms that support the ability of demonstrating accountability, which is central to its definition. Metrics are widely used as an instrument for verifying and showing the compliance of non-functional requirements, such as those related to security and privacy, and are also of total relevance for accountability. The goal of this work package is to develop metrics for supporting accountability, through the demonstration that proper mechanisms for privacy, security and information governance are in place.

The previous deliverable of this work package focused on setting up the foundations towards the elicitation of metrics, in particular by analyzing the principal properties to be measured (i.e., the attributes of accountability) from the metrics perspective, and proposing a top-down approach for metrics elicitation through the use of a Metrics Metamodel.

The top-down approach is complemented by a bottom- up approach based on the analysis of relevant control frameworks, giving as result a Catalogue of metrics for Accountability. This deliverable presents the main results of the Metrics for Accountability work package (WP C-5), epitomized by the Accountability Metrics Catalogue. The catalogue is composed of 39 metrics, organized in three categories: Verifiability and Compliance, devoted to demonstrating compliance to good practices and regulations; Transparency, Responsibility and Attributability, which is related to measuring the characteristics about the internal processes that provide Accountability; and, Remediability and Incident Response, which encompasses metrics related to remediation, redress, and incident response. This catalogue constitutes the final outcome of the metrics elicitation process. Finally, the process of validation of the accountability metrics is described, together with the results of the validation.
In summary, the main contributions of the deliverable are:

- A bottom-up approach to elicit metrics for accountability that complements the top-down approach presented in the previous deliverable.
- By using the complementary two approaches a catalogue of metrics for accountability is derived.
- A validation process is presented for the metrics in the catalogue that is based on questionnaires.

Table of Contents

1 Introduction

1.1 Purpose

Accountability is a complex concept, whose definition varies depending on the discipline where it has to be applied. Thus, for the A4Cloud context, the consortium has agreed on using the following definitions of Accountability [1]:

> *"**Conceptual Definition of Accountability**: Accountability consists of defining governance to comply in a responsible manner with internal and external criteria, ensuring implementation of appropriate actions, explaining and justifying those actions and remedying any failure to act properly.*

One of the important aspects behind the accountability concept is the ability of an organization to demonstrate their conformity with required obligations [8]. The concept of Accountability goes beyond behaving in a responsible manner, and deals also with showing compliance and providing transparency to the internal process of accountability provision. One of the goals of the A4Cloud project is the demonstration of this through the measurement of the degree of such conformity and the provision of meaningful evidence. Thus, measurement becomes an important tool for assessing the accountability of an organization by external authorities (and organizations themselves, in the case of self-assessment).

Essentially, what is needed for ensuring accountability is to be able to demonstrate that the accounts provided by an organisation (to regulators, auditors, data subjects or other service providers) are adequate and appropriate for the context by meeting certain internal and external criteria, and to have in place mechanisms for dealing with the situation if this is not the case. From an organisational point of view the focus is on measuring whether the fundamental types of activities that an accountable organisation should undertake are in place and effective.

Conceptually, the notion of accountability can be decomposed into several properties. Such properties, referred as *attributes of accountability*, include transparency, verifiability, observability, liability, responsibility and attributability. Thus, it would be logical to think that if we are interested in assessing how accountable an organisation is we should be able to assess or provide techniques for measuring the attributes that influence on accountability. How much or to what extent they should be measured is a key issue. One of the goals of A4Cloud is, therefore, to develop a collection of metrics for performing meaningful measures on the attributes that influence accountability.

From a technical viewpoint, metrics are widely used as an instrument for verifying the compliance of non-functional requirements, such as those related to security, privacy, or accountability. Metrics are also a tool that facilitates the decision making process, since they can be seen as an input of the management review process of an organization [20]. For example, they are an important aspect of maturity models, since they are used to support management decisions, improve quality assessment, monitoring of performance, etc. Hence, metrics for accountabillity can be considered as a means for showing that proper mechanisms for privacy, security and information governance are in place and indeed support accountability.

This deliverable presents the main results of the Metrics for Accountability work package (WP C-5), epitomized by the Accountability Metrics Catalogue. While the previous deliverable [2] was focused on the analysis of the principal properties to be measured (i.e., the attributes of accountability) from the metrics perspective, and proposing a top-down approach for metrics elicitation through the use of a Metrics Metamodel, this deliverable exposes the final outcome of the metrics elicitation process, a catalogue of 39 metrics, together with the results of its validation. Moreover, in this deliverable we propose techniques for expressing confidence in the measurement results, by considering quality factors of the evaluation as an additional dimension of the metric.

The structure of this deliverable is as follows: Section 2 describes the role of metrics in relation to the concept of Accountability. Fundamental concepts necessary to define metrics for accountability are presented in this section, such as basic notions of metrology, and the conceptual relationship between Metrics, Accountability and Evidence. The role of metrics with respect to the notion of the account and the Accountability Maturity Model is also discussed. Section 3 describes the approaches followed for eliciting accountability metrics. In particular, two complementary approaches are presented: a top-down approach, based on a Metrics Metamodel (defined in a previous deliverable), and a bottom-up approach, based on the analysis of relevant control frameworks. Section 4 presents the Accountability Metrics Catalogue, which contains 39 metrics organized in three categories. Section 5 proposes possible approaches for further extending the metrics presented here. In particular, a method for expressing confidence in the measure results is proposed. We also present some methods for defining derived metrics. Section 6 deeps in the process of validation of the proposed metrics. Different validation strategies are discussed, and the design of a validation methodology, together with its results, is presented. Finally, Section 7 concludes the deliverable and outlines the future work.

1.2 Glossary of Acronyms / Abbreviations

AMM	Accountability Maturity Model
CCM	Cloud Control Matrix
CMMI	Capability Maturity Model Integration
CSP	Cloud Service Provider
GAPP	Generally Accepted Privacy Principles
NIST	National Institute of Standards and Technology
OCF	Open Certification Framework
PII	Personally Identifiable Information
STAR	Security, Trust and Assurance Registry

2 The Role of Metrics in Accountability

Metrics have a central role in cloud computing, as reflected by the NIST definition of the cloud [28], in which five essential characteristics are identified, namely, on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service. The latter characteristic is defined by the capacity of cloud systems for measuring aspects related to the utilization of services, in order to provide automatic control and optimization of the usage of cloud resources, and ultimately, to support transparency and enhance trust of cloud consumers with regard to cloud providers. Metrics in cloud computing environments are also of paramount importance for other reasons. For instance, metrics can also be derived on the consumer side, enabling cloud consumers to monitor the quality of service of the cloud provider and to verify the compliance of agreed terms. Metrics are also a tool that facilitate the decision making process of cloud consumer organizations, as they can be used for making informed decisions with regard to the election and evaluation of cloud providers.

As for cloud service governance, metrics are very useful means for assessing performance of operational processes and for demonstrating the implementation of appropriate practices through the provision of quantifiable evidence of the application of such practices. Metrics also support accountability governance. The accountability process should assess accountability of an organization in a systematic way, and the definition and use of specialized metrics are means for achieving this. Metrics can be used as an instrument for identifying strengths and weaknesses in the security and privacy mechanisms in place.

From the perspective of the accountability framework, metrics are a means for demonstrating accountability, through the provision of quantifiable evidence of the application of proper practices and the performance of operational processes. This way, progress in the implementation of accountability practices can be justified in a quantitative way. This is discussed in more detail in Section 2.2, where the relation between metrics and the notion of the account is explored. With respect to compliance, the establishment of a proper metrics programme could help to demonstrate the commitment to a proactive approach for accountability [30], by measuring pertinent aspects present in regulations and policies.

Additionally, metrics are a key aspect in the definition of maturity models. The adoption and systematic use of metrics is an indispensable practice for organizations that strive to achieve a repeatable and optimizing behaviour. Measuring the behaviour of organizations is central to define an Accountability Maturity Model, as discussed in Section 2.3. In theory, a mature organization (from the Accountability perspective) should present a quantitative, and hence, measurable behaviour. Mature organizations are therefore characterized by an ingrained use of metrics within their internal processes.

2.1 Concepts

In this section we introduce some fundamental concepts that shape the work on metrics in accountability. Some of these concepts were already introduced in previous work of this WP (see [5] and [2] for more details).

2.1.1 Fundamental Concepts of Metrology

In this section, we provide some basic definitions of concepts that are fundamental to measurement in general. This section is just a summary of the findings presented in a previous deliverable, D:C-5.1 [2]. These concepts are:

- **Attribute**: property or characteristic of an object that can be distinguished quantitatively or qualitatively by human or automated means [18].
- **Metric or measurement result**: A set of indicators, together with an associated interpretation, that is designed to facilitate decision-making and improve performance and accountability through collection, analysis, and reporting of relevant data (adapted from [20], [17]).
- **Measure**: variable whose value is assigned as a result of measurement [18].
- **Measurement method**: logical sequence of operations, described generically, used in quantifying an attribute with respect to a specified scale [18].

- **Indicator:** measure that provides an estimate or evaluation of specified attributes derived from an analytical model with respect to defined information needs [20].
- **Evidence**: data collected to support a metric, including the data inputs necessary to calculate and validate the metric (adapted from [27]).

Another important concept that we should present again, is the notion of scales of measurement. In the classical theory of measurement [39], the *scales of measurement* (or *levels of measurement*) are a set of categories for classifying measurement methods regarding their characteristics. Identifying the scale for each particular metric is essential for interpreting and analysing its results. Moreover, since each scale has a set of permitted operations, knowing its scale allows us to assess the validity of a metric, or at least, to discard senseless metrics.

- Nominal scales: This type of scale is applicable for mapping entities to names or categories. It is also known as categorical scale. Values in a nominal scale do not have any kind of relation to each other. For this reason, only the equality operation (=) is permitted for nominal values. From a statistical viewpoint, only modes can be computed.
- Ordinal scales: This scale permits to assign an order relation to its values, which is used to put measured entities in order. For this reason, ordinal scales are said to have magnitude. However, there is no information for measuring the differences between values. A simple example of this scale is the set of values “Low – Medium – High”. There is an order relation that permits to state that High is greater than Medium, which in turn is greater than Low, but it makes no sense to measure the difference between Low and Medium. Ordinal scales are also nominal. Ordinal scales therefore permit to use equality (=) and inequality (≤) operations, as well as medians and percentiles. Certain non-parametric statistical tests that only require ordinal data, known as ranking tests [37], can also be performed.
- Interval scales: This type of scale permits to measure differences between values. Additionally, interval scales are also ordinal scales. Thus, their values can be compared and ordered. Interval scales permit additions and substractions of their values. Therefore, means and standard deviations can also be computed. Although multiplications and divisions, and hence any other operations that depend of those, such as ratios, cannot be performed, multiplications and divisions of differences between values can indeed be computed.
- Ratio scales: This type of scale improves interval scales by adding a meaningful zero value. Ratio scales are also interval scales. All the operations that are valid for interval scales apply here. In addition, multiplication and division are also meaningful.

Nominal and ordinal metrics are often grouped as **qualitative** metrics, whereas interval and ratio metrics are **quantitative**. This differentiation is very important when facing the processing the results of metrics, which will happen when aggregating and compositing metrics or when producing interpretation of the results of a metrics. Qualitative metrics may need to be converted to quantitative, in order to make possible complex processing, such as aggregated metrics. Note that this process often consists on defining a transformation from a qualitative domain (which at most possess a partial ordering) to a numeric one, which implies making assumptions on the validity of such transformation. On the contrary, quantitative metrics may need to be converted to qualitative when facing the reporting of final assessments, in order to be easily interpreted by people; for example, a numeric metric could be transformed to a simple Green/Yellow/Red label.

2.1.2 Accountability Attributes

Accountability is conceptually decomposed into attributes, as presented in the work from the Conceptual Framework [4]. **Accountability attributes** capture concepts that are strongly related to and support the principle of accountability. These include: key properties of accountability (e.g. transparency); conceptual elements (e.g. remediation); consequences (e.g. sanctions); related objects (e.g., obligations). There exist emerging relationships (e.g. implication and inclusion) among attributes dependent on different viewpoints of analysis (which are related to societal, legal and ethical aspects of accountability). For instance, from a legal perspective, responsibilities imply obligations, which consequently may involve sanctions. From a social perspective, transparency implies both observability and verifiability (and vice versa, transparency is obtained by combining observability and verifiability). This section defines and focuses on accountability attributes: *observability*, *verifiability*, *attributability*, *transparency*, *responsibility*, *liability* and *remediability*.

- **Observability** is a property of an object, process or system which describes how well the internal actions of the system can be described by observing the external outputs of the system.
- **Verifiability** is a property of an object, process or system that its behaviour can be verified against a requirement or set of requirements.
- **Attributability** is a property of an observation that discloses or can be assigned to actions of a particular actor (or system element).
- **Transparency** is a property of an accountable system that it is capable of 'giving account' of, or providing visibility of, how it conforms to its governing rules and commitments.
- **Responsibility** is the state of being assigned to take action to ensure conformity to a particular set of policies or rules.
- **Liability** is the state of being legally obligated or responsible.
- **Remediability** is the state (of a system) of being able to correct faults or deficiencies in the implementation of a particular set of policies and rules and/or providing a remedy to a party, if any, harmed by the deficiency.

Note that this set of Accountability Attributes was the one in place during most of the lifetime of the workpackage. However, latest updates on the Conceptual Framework have added two additional attributes, as exposed in D:C-2.1 [1], that we are not conisdering from the metrics perspective in this deliverable as the timeframe sinde their inclusion and the time this document had to be ready did not allow it:

- **Appropriateness** is a property required to the mechanisms implemented by an organisation to put into effect the principles and obligations of accountability and demonstrate this on request.
- **Effectiveness** is a property assessed over the mechanisms implemented by an organisation to put into effect the principles and obligations of accountability and demonstrate this on request.

Since this addition is posterior in time to most of the work here presented, we will not consider these two last attributes.

2.1.3 Accountability Evidence

The concept of "Evidence" from the point of view of Accountability is defined in [3] as:

> ***"Accountability evidence** can be defined as a collection of data, metadata, and routine information and formal operations performed on data and metadata, which provide attributable and verifiable account of the fulfilment of relevant obligations with respect to the service and that can be used to convince a third party of the veracious (or not) functioning of an observable system."*

This definition is broad enough to permit the consideration of different types of evidence sources, ranging from observations of technical systems (e.g., network logs) to organizational documentation (e.g., internal policies of an organization). This is reflected in the consideration as evidence, not only of data (and metadata), which is usually associated to technical characteristics and observations of a system or process, but also of "routine information", which comprises information regarding the internal processes of organizations. As noted by [3] "data, metadata, formal operations and routine information can be called the ***supporting evidence elements*** for accountability". From now on, and within the context of metrics for accountability, we refer to these elements as "Evidence".

2.1.4 A Conceptual Model of Accountability Metrics

Taking into consideration the definition of the Accountability Attributes and Evidence given in the previous section, we can informally define what is the relationship between these concepts and the notion of metrics. Figure 1 shows this relationship, which links Accountability Attributes to Evidence through the use of Metrics.

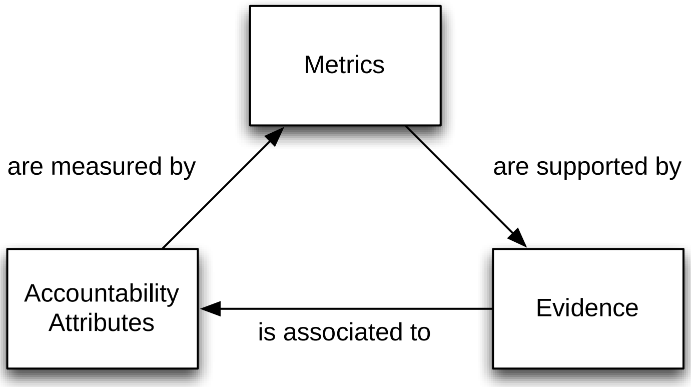

Figure 1: Conceptual Relationship between Accountability Attributes, Metrics and Evidence

The concept of Evidence conveys all the information supporting the evaluation performed by a Metric. As discussed briefly in the next section and in section 3.1.2, metrics do not directly measure or evaluate a property (i.e. an accountability attribute), but the evidence associated to it. As shown in Figure 1, evidence plays a major role in the relationship between metrics and accountability attributes.

2.2 Metrics and the notion of the Account

A central concept related to Accountability is the notion of the Account. It is defined in MS:C-2.3 [4], as "*a report or description of an event through the use of measurable forms of evidence*". That is, the notions of Account, Evidence and Event share the following relation shown in Figure 2:

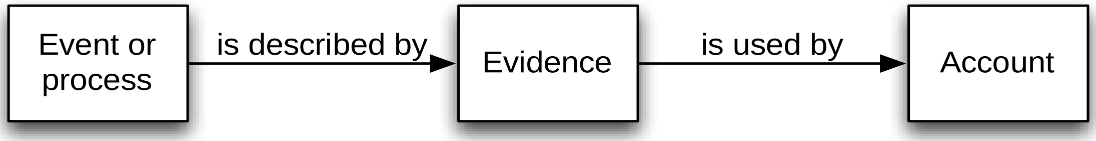

Figure 2: Relation between the Notions of Account, Evidence and Event

We distinguish two types of accounts:

- Declarative account: The account consists of a description, such as a report, of a process or event. The account has primarily a retrospective function. It is important to note that not all forms of account as perceived from a legal perspective are actually measurable.
- Evaluative account: In this case, the account takes the form of an assessment of a process or an event, through the use of evidence. Same as above, "evidence" from a legal point of view does not necessarily relate to something measurable. Thus, evaluative accounts are restricted to measurable forms of evidence.

As mentioned earlier, the concept of Evidence is central for the process of eliciting metrics. From the metrics point of view, not all forms of account, as perceived from a legal perspective, are actually measurable, any assessment or evaluation of a property or attribute can only be made using as input some tangible information. The term "evidence" was used in this context to refer to the information used to support the assessment within a metric. Examples of "evidence" are an observation of a system, a system log, a certification asserted by a trusted party, a textual description of a procedure, etc. Hence, a metric does not directly measure a property of a process, a behaviour, or a system, but the evidence associated to them. It can be seen that the notion of evidence in this context is very broad and it is not limited to computerised data (such as a system log), but can be applied to more general information (e.g. the description of a process within an organisation, a certification, etc.). Note that it is unfeasible to measure all evidence highlighted from a legal point of view. Our analysis here is concerned with "measurable forms of evidence".

Taking into consideration this distinction and the concepts developed in deliverable D:C-5.1 [2], we note two main findings:

- There is a parallelism between the notion of "evidence" and declarative accounts. In some sense, we could argue that the notion of evidence is analogous to a declarative account (when related to measurable aspects).
- Metrics are of relevance for evaluative accounts during the assessment or evaluation procedure.

In MS:C-2.3 [4], an account is said to be a means for demonstrating accountability. Hence, metrics for accountability, supported by Declarative Accounts, could be considered as instruments to this end, by measuring how well accountability attributes are achieved in a particular context. The result of the application of metrics could be used by Evaluative Accounts to demonstrate compliance and fulfilment of accountability obligations. The relation among these concepts is depicted in the following Figure.

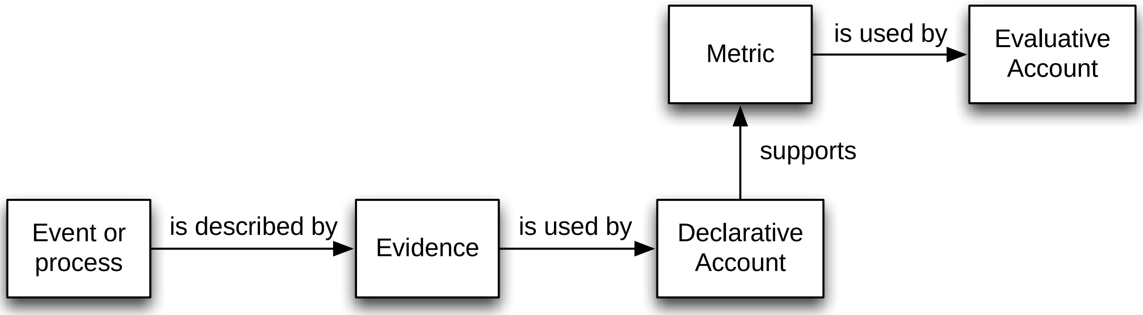

Figure 3: Conceptual Relationship between Metrics and the Account

2.3 Metrics and the Accountability Maturity Model

The metrics discussed in this document play a central role for organizations aiming to assess their level of accountability. However, a problem at the state of the art refers to selecting the actual set of “elements” that assessors have to consider for the quantification of accountability. Given the real-world experience with control frameworks for assessing both security (e.g., ISO/IEC 27002 [19]) and privacy (e.g., Generally Accepted Privacy Principles [6]), we have proposed an accountability maturity model (AMM) for the cloud [1] in A4Cloud. The proposed AMM can capture both, the maturity of individual organisations in terms of accountability practices, as well as a measurement of the appropriateness of the measures used across the whole cloud supply chains.

Quantitatively speaking, evaluating the controls of the proposed AMM is central for organizational maturity models because of their role in quality assessment, monitoring of processes performance and support of management decisions. Despite its inherent subjectivity, human stakeholders (e.g., cloud auditors, decision makers) are nowadays quite familiarized with the use of high-level quantifiers known as *maturity levels* while assessing a control framework. The AMM proposes two different approaches for quantifying/assigning maturity levels. The first approach assesses, at a high-level of abstraction, the controls of the framework in order to assign any of five possible maturity levels (*Ad Hoc, Repeatable, Defined, Managed, or Optimised*). This approach is common to many state of the art maturity models, and has been broadly discussed in MS:C-2.3 [4]. In theory, a mature organisation (from the perspective of maturity models) should present a quantitative, and hence, measurable behaviour. However, despite the basic quantitative reasoning enabled by the use of maturity levels, the outcome is usually too high-level and it cannot be directly applied to automate the organization's accountability management (e.g., adaptation of data protection mechanisms in case of cyber-incidents).

The second quantification approach proposed by the AMM takes a more fine-grained perspective, and uses low-level quantifiers to assess each individual accountability control. The use of granular metrics is relevant to cloud consumers, enabling them to monitor the quality of service of the cloud provider and to verify the compliance of agreed terms. Metrics are also a tool that facilitate the decision making process of cloud consumer organizations, as they can be used for making informed decisions with regards to the election and evaluation of cloud providers. The adoption and systematic use of low-level metrics into the AMM is an indispensable practice for organizations that strive to achieve a repeatable and optimizing behaviour, therefore enabling the implementation of realistic levels of automation. Mature organisations are therefore characterized by an ingrained use of metrics within their internal processes.

The traditional usage of "maturity levels" provides only an aggregated view of the accountability level achieved by an organization, but further refinements are difficult to achieve and decision makers might fail to visualize the weakness on their implemented controls (and therefore to improve their accountability practices). In a similar manner, a CSP (Cloud Service Provider) providing low-level accountability metrics to their (prospective) customers might facilitate both the decision making process (i.e., choosing the adequate provider), and the continuous monitoring of the accountability levels. This is discussed in much more detail in the Conceptual Framework deliverable [1].

3 Eliciting Metrics for Accountability

In this section we describe the process we have followed for eliciting metrics for accountability. In order to measure the accountability attributes we need to have a clear target of the aspects of the attributes that are to be measured. The definitions of the attributes are in some cases vague, subjective or ambiguous, thus it is difficult to measure specific aspects. We need a suitable model that allows us to identify measurable factors from the definitions of the attributes. Once these specific factors are identified we need to derive metrics for them based on the analysis of existing control frameworks. Thus, the process of eliciting accountability consists of two complementary approaches:

- A *top-down* approach: our initial approach is based on the definition of a Metamodel for Accountability Metrics, design to aid during the initial phases of the elicitation of metrics. It is described in Section 3.1.
- A *bottom-up* approach: it is used for complementing the previous one, based on the analysis of relevant control frameworks. It is explained in more detail in Section 3.2.

3.1 Metrics Metamodel

In this section we present a metamodel for describing metrics for accountability attributes, which helps during the process of elicitation of metrics for accountability. This metamodel is intended for the modelization of complex properties, as the accountability attributes, and metrics for measuring them. One of the main goals of this metamodel is permitting a top-down and recursive decomposition of properties. This aspect is detailed in Section 4.1. The Metrics Metamodel was already presented previously in [2] and [31], but we include it here for the sake of completeness.

The accountability attributes belong to the family of non-functional properties, which include all properties that are not directly related to functionality, but to a quality or behavioural attribute of a system [38]. Non-functional properties, such as the ones related to security and privacy, are of key importance with regard to the analysis and evaluation of the different aspects of a system, a service or an organization, such as quality and trustworthiness. However, their evaluation is traditionally complicated because of several reasons. Firstly, because of their subjective and ambiguous nature; secondly, non-functional properties usually present multi-dimensionality, possessing several facets; and finally, in some cases, the optimization of a non-functional property may be inconsistent with others. The consequence of all these factors is that it is very difficult to assess if this kind of properties have been met since there is no clear-cut criteria for deciding. This problem is very similar to the one of non-functional requirements in the area of requirements engineering [27].

3.1.1 Top-Down Decomposition Approach for Eliciting Accountability Metrics

It is clear then that the non-functional nature of accountability attributes is an important hindrance for defining meaningful metrics. As stated before, most of the problems we face are related to the level of abstraction of the attributes of accountability. Some of such attributes are defined in a very high-level of abstraction, which is prone to vagueness and ambiguity, and are then not useful from a metrics perspective. Furthermore, there is a disparity in the level of abstraction between different attributes. Thus, a tentative solution is to consider a stratified view of the attributes, where high-level attributes represent more vague and wide concepts and low-level attributes represent more tangible and empirical notions. This would allow also a fine-grained decomposition of attributes, if needed. Hence, we propose a top-down decomposition approach that works on two different levels:

- The **conceptual level**, where all the high-level concepts related to Accountability (e.g., the core attributes for accountability) are defined as well as the relations among them. These high-level concepts can be further refined into more concrete ones. This level will include the attributes being identified in WP:C-2 as the core attributes, and will also comprise sub-attributes that still are high-level enough for not being useful for metrics, but needed in order to define correctly the concepts related to accountability. Thus, the rationale for their definition is mainly conceptual.
- The **measurable level**, where we deal with "tangible" and empirical concepts. In certain cases, these attributes could be decomposed even more. Metrics will be initially defined for these peripheral concepts.

The idea we propose is to first go downwards in order to "break down accountability" into simpler and more low-level concepts, constructing a tree-like model (possibly, a directed graph) until we reach measurable things. This is a common approach in security metrics. Therefore, in this model, measurable concepts are in the peripheral nodes. Next, from this model, and using inference techniques over its relations, we could go upwards and construct metrics for high-level concepts. This aspect is currently under development and will be provided in next versions of the deliverable.

3.1.2 The Metamodel for Metrics for Accountability Attributes

In this section, we propose a model-driven approach that includes the definition of a metamodel for describing metrics and accountability properties. The goal of this metamodel is to serve as a language for describing: (i) accountability properties in terms of entities, evidence and actions, and (ii) metrics for measuring them. Note that this metamodel could be extended for its application to non-functional properties in general, however, this is out of the scope of this work since we currently focus on those related to the accountability concept.

One of the main aspects of this metamodel is that metrics are defined to take two main kinds of inputs: **Evidence** and **Criteria**. From our point of view, any assessment or evaluation (i.e, a metric) can only be made using as input some tangible and empirical evidence, such as an observation, a system log, a certification asserted by a trusted party, a textual description of a procedure, etc. That is, a metric does not directly measure a property of a process, behaviour, or a system, but uses the evidence associated with them in order to derive a meaningful measure. That is the idea that we are trying to capture in this metamodel: Evidence is the fundamental support of any evaluation method and is what gives an objective dimension to assessments. On the other hand, criteria are all the elements that convey contextual input that may constrain what should be measured, such as stakeholder's preferences, regulations and policies. It is clear then that each metric will have different nature depending on the criteria. Therefore, in this metamodel, both Evidence and Criteria are central to the definition of metrics.

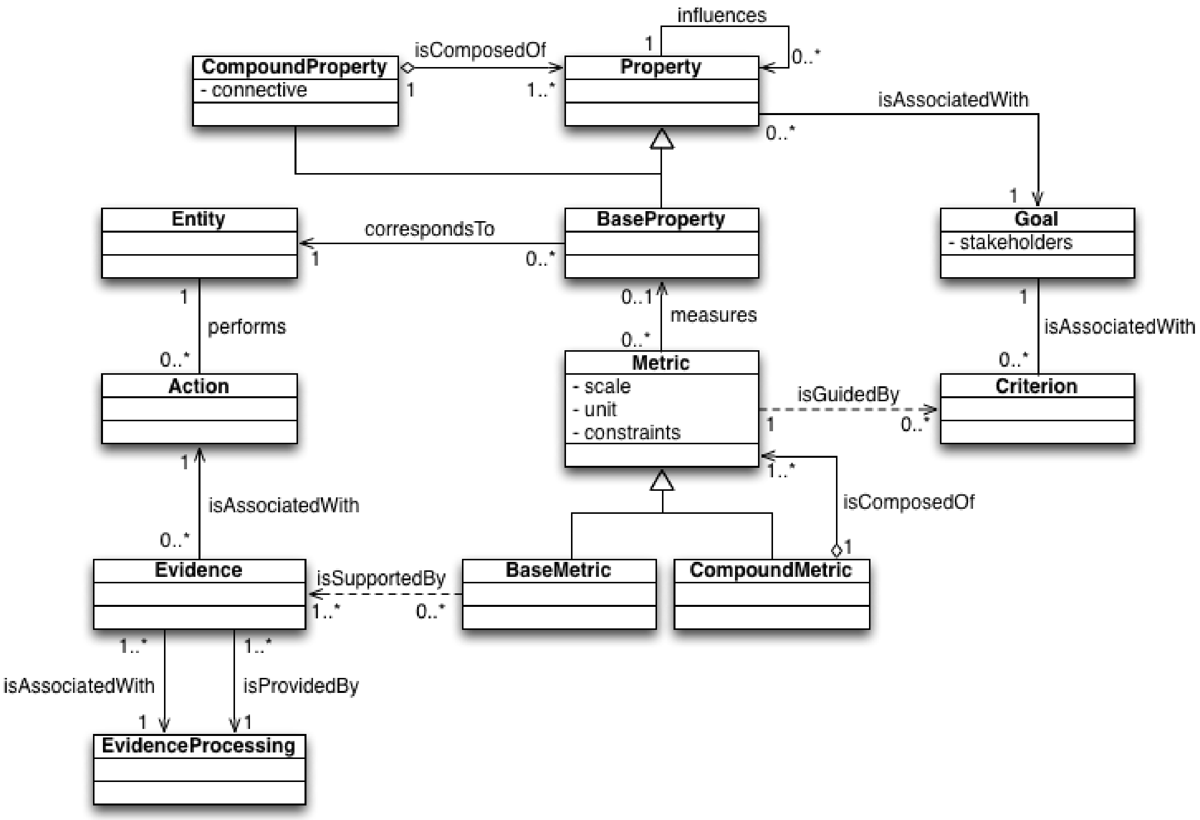

Figure 4: Metamodel for Metrics for Accountability

In this section, we will present our metamodel (see Figure 4), and provide a detailed description of each of its elements and the relations among them:

- **Goal:** High-level description of the property (or family of properties) that is modelled. These elements also contain a reference to the stakeholder (or stakeholders) for which the goal is oriented.
- **Property:** As mentioned earlier, non-functional properties are qualities or behavioural characteristics of an entity. Ideally, properties can be distinguished quantitatively or qualitatively by some evaluation method; however, properties may be defined as very high-level concepts. Thus, we consider that properties can be further decomposed into more basic ones in some cases. In these cases, **BaseProperty** elements can be defined in terms of entities and the actions between them, whereas **CompoundProperty** elements are defined in terms of other properties, making possible a top-down decomposition of properties, from a high-level and abstract way to a tangible and more accessible one. **CompoundProperty** elements then have a connective attribute, which is used for describing the logical connective used for combining properties. In addition, properties may also influence other properties, not necessarily taking part of a composition relation; the model then permits to express these influence relations between properties.
- **Entity:** This element is used to describe the entity that meets the modelled property. An entity is a physical or conceptual object that performs actions and that meets properties. For example, an organization, a process or a system can be considered as entities.
- **Action:** We define this as a process that occurs over a period of time and is performed by or has an effect on entities. Even though, actions have an effect in the environment, we cannot deal directly with these consequences, but with the evidence associated to them.
- **Evidence:** We define evidence as a collection of information with tangible representation about the effect of actions. Evidence is used to support a metric. That is, evidence is not an abstract concept about the consequence of activities, but actual data that can even be processed by a machine. Note, however, that evidence may come from sources with different levels of certainty and validity, depending on the method of collection or generation of such evidence.
- **EvidenceProcessing**: In our model, we assume that evidence, although it is associated to the effect of actions, does not directly stem from them. Instead, evidence is originated or collected by means of an **EvidenceProcessing** element. In this way, we model the fact that there may not exist a perfect correlation between the effects or consequences of actions and the evidence associated with them. The **EvidenceProcessing** element makes this difference explicit. With the inclusion of this element in our metamodel, we emphasise that the method of collection and processing of evidence is as important as the evidence itself. For this reason, there should also be evidence associated with each **EvidenceProcessing** element, describing how it works. Such evidence may be used by a metric during the evaluation process.
- **Metric:** We define this as an evaluation method for assessing the level of satisfaction of a non-functional property in a quantitative or qualitative way, on the basis of evidence and contextual criteria. Metrics can be of two types: **BaseMetric** for metrics that use evidence as inputs for their calculations, and **CompoundMetric** for aggregated metrics that are defined as a function of other metrics. Aggregated metrics may rely on auxiliary metrics that are not associated with any property and that are defined solely for facilitating the definition of the parent metric. In both cases, metrics may use **Criterion** elements for guiding the evaluation with respect to the context of the metric. This element has the following fields:
  - **Scale**: This field describes the type of measurement scale used in this metric. The scale can be nominal, ordinal, interval or ratio.
  - **Unit**: This field represents the measurement unit adopted as standard for measuring the property. The definition of a measurement unit is only necessary in the case of quantitative metrics.
  - **Constraints**: This field conveys the contextual constraints that may affect the application and validity of the metric.
- **Criterion**: This element captures all the contextual input that may constrain what should be measured by the metric, such as regulation, best practices, organisational policies and contracts, and stakeholders’ preferences. It could be the case that one could define different metrics for the same property. The assessment methodology for each metric will depend on the contextual input given for the metrics evaluation. The **Criterion** element will be the responsible of conveying such contextual information.

The intention behind this metamodel is to be used as part of the process of elicitation and evaluation of accountability properties in a cloud context. Hence, the stakeholder who is interested in assessing such properties would be the one that takes the role of owner of the model described using this metamodel. Each particular model defined using this language reflects the viewpoint of the model owner with regard to the context of application. Customization of models to specific situations is then done in different ways:

- Decomposition and interlinking of properties: the modeler can freely identify the goals and their associated properties, which can be further decomposed into other subproperties or interlinked through influence relations.
- Modelling of entities and their actions: Entities and actions can be modelled with the level of abstraction desired by the model owner, as the metamodel simply dictates that entities perform actions.
- Identification of meaningful evidence sources: the **EvidenceProcessing** element is used to model the sources of evidence that stem from the effect of actions.
- Definition of different metrics in terms of evidence and criteria: the possibility of defining different metrics for the same property is another characteristic that supports the customisation of models. Thus, the context and preferences of the model owner with regard to evaluation of properties can be reflected. Each metric would have different sources of evidence and criteria.

3.1.3 Shortcomings of this Approach

Although a top-down approach may seem like a natural strategy for reasoning about high-level concepts such as Accountability, it does not guarantee to reach measurable concepts. In fact, when facing the elicitation of metrics using the metrics metamodel, we found to be difficult to derive metrics starting from the model of the accountability attributes. Actually, the value of the proposed metrics metamodel lays principally in aiding to correctly identify and specify the subconcepts that are relevant or influence the Accountability Attributes, rather than being a method for extracting relevant metrics. For this reason, we needed a complementary strategy for our top-down approach.

The goal of a complementary strategy would be to use information located closer to the sources of Evidence in order to facilitate its linking to the Accountability Attributes through the use of metrics. In other words, we needed a bottom-up approach, from Evidence to Metrics. Taking into consideration that the main goal of the metrics for accountability is to demonstrate that proper mechanisms for privacy, security and information governance are in place, we devised a bottom-up approach based on analysing control frameworks. This approach is explained in the next section.

3.2 Bottom-up Approach

Control frameworks that are relevant for accountability, such as the Cloud Control Matrix [9], the Generally Accepted Privacy Principles [6], and NIST SP 500-83 [29], are specifically designed for covering the categories of mechanisms that implement security, privacy and information governance. For this reason, it is fair to assume that they can be used as sources of evidence from where metrics can be derived. Moreover, control frameworks are widely used in organizations during audits and certifications, so evidence of their application can be reasonably extracted from audit records or similar data (see Figure 5).

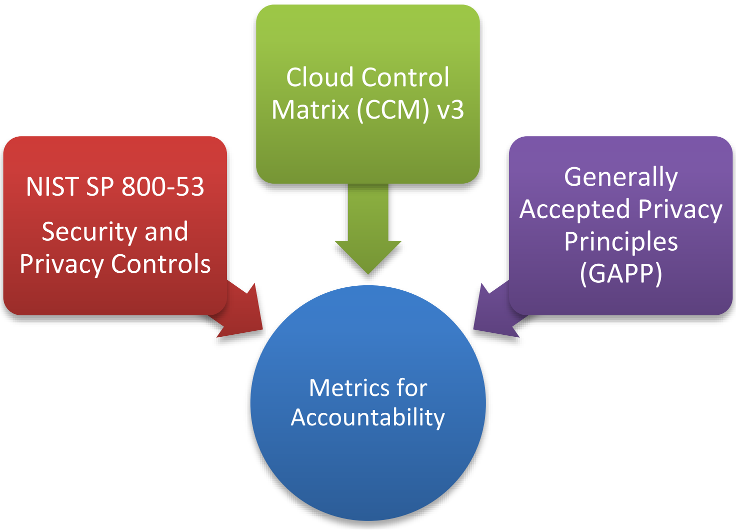
Cloud Control Matrix (CCM) v3
NIST SP 800-53 Security and Privacy Controls
Generally Accepted Privacy Principles (GAPP)
Metrics for Accountability

Figure 5: Bottom-up approach for metrics elicitation

The goal is that metrics derived this way could evaluate the existence and quality of implemented mechanisms that are object of the controls, and that are designed for ensuring accountability. These mechanisms are not only of technical nature, but also organisational. The main advantage of this approach is that the derived metrics will be automatically aligned with the principles of Accountability, since a quantitative improvement in the measured results will have beneficial effects on the fulfillment of the controls. This approach could also be extended for considering other sources apart from control frameworks, such as regulations.

The steps of the bottom-up approach are as follows:

1. To analyse relevant control frameworks in the light of Accountability Attributes. The goal of this step is to select those controls that influence Accountability to some extent.
2. To study the nature of the control, in order to identify whether there is any quantifiable element in the description of the control that is susceptible to being measured. Qualitative elements may be identified too, if they have at least an ordinal nature.
3. To define a metric that measures the identified elements, using the qualitative or quantitative elements identified in the previous step.
4. To check that the metric supports the concept of Accountability and, in particular, the Accountability Attributes for which it is related to.

Control ID	IP-4: Complaint Management
Control	The organization implements a process for receiving and responding to complaints, concerns, or questions from individuals about the organizational privacy practices
Supplemental Guidance	Complaints, concerns, and questions from individuals can serve as a valuable source of external input that ultimately improves operational models, uses of technology, data collection practices, and privacy and security safeguards. Organizations provide complaint mechanisms that are readily accessible by the public, include all information necessary for successfully filing complaints (including contact information for the Senior Agency Official for Privacy (SAOP)/Chief Privacy Officer (CPO) or other official designated to receive complaints), and are easy to use. Organizational complaint management processes include tracking mechanisms to ensure that all complaints received are reviewed and appropriately addressed in a timely manner.

Table 1: Example of Control related to Accountability

As an illustrative example of this approach, let us consider the control IP-4 from NIST 800-53 Rev. 4 [29], shown in Table 1. This control is entitled “Complaint Management” and dictates that “*the organization implements a process for receiving and responding to complaints, concerns, or questions from individuals about the organizational privacy practices*”. This control is relevant to Accountability since it is directly supporting one of the Accountability Attributes, namely Remediability, which is devoted to the establishment of policies and procedures for providing remedy to a party after a failure; other attributes, such as Transparency, are also influenced. Once we have identified this control as relevant, we may proceed to the second step where we study the description of the control in order to identify quantifiable elements. In the extended description of the control it is stated that "*the organization responds to complaints, concerns, or questions from individuals within an organization-defined time period*". Hence, we conclude from this description that timely response of complaints is important for supporting Remediability. To this end, measuring the actual time of complaint responses could provide a meaningful and quantitative measure of this sub-aspect of Remediability. Next, we define the metric "*Mean time to respond to complaints*" as the average time that it takes for the organization to respond to complaints from affected stakeholders. Finally, we argue that an organization that truthfully strives to minimize the result of this metric would indeed enhance its Remediability state, and therefore, its "Accountability level". Using metrics like this, the organization can optimize its policies and procedures following a quantitative approach towards accountability enhancement.

3.2.1 Analysis of Relevant Control Frameworks

We introduce in this section the control frameworks that are more relevant to Accountability in the Cloud, and which ones were used as a basis for the bottom-up presented before.

3.2.1.1 CSA Cloud Control Matrix (CCM)

The Cloud Security Alliance's Cloud Control Matrix (CCM) [9] is a control framework specifically designed for the purposes of cloud security. CCM provides fundamental security principles to guide cloud vendors and to assist prospective cloud customers in assessing the overall security risk of a cloud provider. Furthermore, this security control framework strengthens existing information security control environments by delineating control guidance by service provider and consumer, and by differentiating according to the cloud model type and environment. The CCM is one of the pillars in CSA's Open Certification Framework [11], and its ultimate goal is to normalise security expectations, cloud taxonomy and terminology, and security measures implemented in the cloud. In its latest version, CCM v3.0 provides a controls framework divided in 16 domains that are cross-walked to other industry-accepted security standards, regulations, and other relevant controls frameworks to reduce audit complexity. Overall, CCM is comprised of 136 controls.

3.2.1.2 AICPA/CICA Generally Accepted Privacy Principles

The AICPA/CICA Generally Accepted Privacy Principles (GAPP) [6] is a set of privacy-related principles for guiding the definition and management of privacy programs. Each of these principles has associated a set of criteria of accomplishment, summing up to a total of 73 GAPP criteria. The GAPP control framework is categorised in 10 thematic areas or "principles" (i.e., which we considered as domains), each of them grouping a number of criteria, which we will consider as controls. It is worth noticing that from a very broad perspective, we can find some similarities among the GAPP principles and A4Cloud's accountability attributes, as described in [1].

3.2.1.3 NIST 800-53 Rev. 4 - Privacy Control Framework

The NIST Special Publication 800-53 Revision 4 [29] consists of a wide catalogue of security and privacy controls intended for US federal information systems and organizations. The last revision of this publication included a comprehensive set of 26 controls specialized in privacy and data protection.

3.3 Combining both Approaches

Both approaches can be combined in order to derive meaningful accountability metrics. The top-down approach helps us understanding the high-level concepts that stem from the Accountability notion, whereas the bottom-up approach helps us to extract metrics related to these concepts from relevant control frameworks, such as the ones presented in the previous section. Figure 6 shows an example of the combination of both approaches, where the high-level concept of the Transparency attribute is recursively decomposed in lower-level sub-concepts. At the same time, particular controls from different control frameworks are analysed in order to extract metrics relevant to the concepts identified by the top-down approach. In the example shown in the Figure 6, controls that deal with the procedures and quality of the notification processes, such as GAPP 1.2.7 and CCM STA-05, are linked through a metric to low-level sub-concepts of Transparency, such as the accessibility of the notifications. In this example, a naïve metric is depicted, which describes the existence of such notification processes.

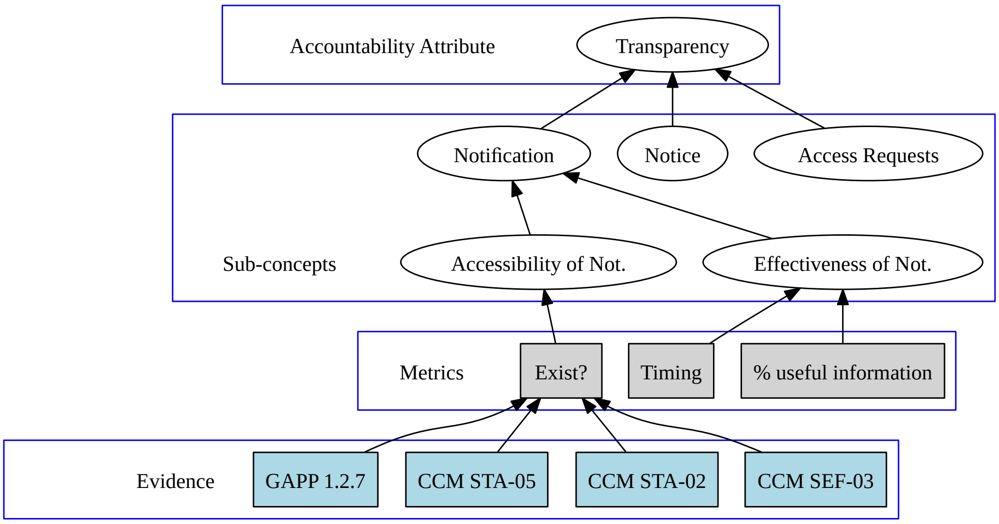

Figure 6: Combining the top-down and bottom-up approaches

4 Accountability Metrics Catalogue

The result of applying the methodologies presented in Section 3 is a catalogue of 39 metrics that we present by using the following generic template that captures the main features of each metric:

- **Metric ID**: A numeric identifier of the metric.
- **Name of the Metric**: A distinctive name that summarizes the goal of the metric.
- **Description**: A brief explanation of the purpose of the metric.
- **Accountability Attributes**: An enumeration of the Accountability Attributes that are influence by the results of the metric.
- **Associated Evidence**: A description of which evidence sources could be used for extracting the information necessary to compute the metric.
- **Input**: The specification of the input parameters that are used for computing the metric.
- **Formulation and output**: A description of the method for computing the metric, as well as the identification of what is the output. In most cases, the formulation would be an arithmetic formula or a description of levels.
- **References**: Identification of relevant references, in particular, to those controls that are associated to the metric. In this case, the reference to the control includes the name of the control framework and the identifier of the control.

Note that the template is independent of the attributes and could be applied to any metric derived for any of the new accountability attributes that we are not considering for this deliverable for time reasons. According to the metrics that we have identified we realised that they can be classified into three main different blocks, depending on the accountability attributes that they cover. Thus, the catalogue is classified into the following thematic categories:

- Verifiability and Compliance
- Transparency, Responsibility and Attributability
- Remediability and Incident Response

In the following sections we will present the metrics according to the above classification.

4.1 Verifiability and Compliance

This set of metrics is devoted to aspects related to demonstrating compliance to good practices and regulations. In the Conceptual Framework of A4Cloud, these aspects are subsumed under the “Verifiability” attribute.

4.1.1 Metric 1. Authorized collection of PII

Description: This metric describes the coverage of authorizations for collecting personally identifiable information (PII).
Accountability Attributes: Verifiability
Associated Evidence: Authorization records for collecting PII. These records should exist if the organization actually asked for permission.
Input: This metric is computed using the following parameters:
A: Number of collected PII records for which authorization from the data subject exists
T: Total number of collected PII records
Formulation and output: Output = (A/T)*100
References: NIST SP 800-53 R4 (AP-1, IP-1)

4.1.2 Metric 2. Privacy Program Budget

Description: This metric describes the percentage of the organization’s IT budget that is allocated for establishing and maintaining a privacy program
Accountability Attributes: Verifiability
Associated Evidence: Privacy Program

Input: This metric is computed using the following parameters:

B	Allocated budget to the privacy program
T	Total budget

Formulation and output: Output = (B/T)*100
References: NIST SP 800-53 R4 (AR-1). See also NIST SP 800-55, Appendix A-2.

4.1.3 Metric 3. Privacy Program Updates

Description: This metric describes the frequency of updates to the privacy program, policies and procedures by a competent role (e.g. Data Protection Officer (DPO)).
Accountability Attributes: Verifiability, Responsibility
Associated Evidence: Privacy Program records
Input: This metric is computed by using the following parameters:

N:	Number of scheduled updates to the privacy program, per year

Formulation and output: Output = N
References: NIST SP 800-53 R4 (AR-1). GAPP (1.1.2, 1.2.1)

4.1.4 Metric 4. Periodicity of Privacy Impact Assessments for Information Systems

Description: This metric describes the periodicity of Privacy Impact Assessments for Information Systems
Accountability Attributes: Verifiability
Associated Evidence: Data privacy policies describing the plans and procedures for performing Privacy Impact Assessments.
Input: This metric is computed using the following parameters:

N	Number of scheduled Privacy Impact Assessments, per year

Formulation and output: Output = N
References: NIST SP 800-53 R4 (AR-2)

4.1.5 Metric 5. Number of privacy audits received

Description: This metric describes the number of independent reviews and assessments performed to the privacy program, policies and procedures in place.
Accountability Attributes: Transparency, Verifiability
Associated Evidence: Privacy audits records.
Input: This metric is computed using the following parameters:

A	Number of audits received

Formulation and output: Output = A
References: CCM v3 (AAC-02). GAPP (8.2.7)

4.1.6 Metric 6. Successful audits received

Description: This metric describes the percentage of independent reviews and assessments performed to the policies and procedures in place for complying with applicable contractual and regulatory obligations.
Accountability Attributes: Observability, Transparency, Verifiability
Associated Evidence: Privacy audits records.
Input: This metric is computed using the following parameters:

S	Number of successful audits received
A	Total number of audits received

Formulation and output: Output = (S/A)*100
References: CCM v3 (AAC-02). GAPP (8.2.7)

4.1.7 Metric 7. Record of Data Collection, Creation, and Update

Description: This metric describes a percentage of the extent to which date is recorded when collecting, creating and updating private records. Date of data collection, creation and update is relevant for complying with data retention schedules.
Accountability Attributes: Verifiability
Associated Evidence: Records associated to private data.
Input: This metric is computed using the following parameters:

N	Number of collected PII records for which the date is indicated
T	Total number of collected PII records

Formulation and output: Output = (N/T)*100
References: NIST SP 800-53 R4 (AP-1, IP-1, DM-2). GAPP (5.2.2)

4.1.8 Metric 8. Data classification

Description: This metric describes a percentage of the extent to which private data is identified and classified according to sensitivity and risk.
Accountability Attributes: Verifiability
Associated Evidence: Records associated to private data.
Input: This metric is computed using the following parameters:

N	Number of collected PII records that have been classified
T	Total number of collected PII records

Formulation and output: Output = (N/T)*100
References: NIST SP 800-53 R4 (SE-1). GAPP (1.2.3). CCM v3 (DSI-01)

4.1.9 Metric 9. Coverage of Privacy and Security Training

Description: This metric describes the percentage of relevant employees who have received training on the privacy program and policies in place. The definition of relevant employee could vary (e.g., those that handle private data)
Accountability Attributes: Verifiability
Associated Evidence: Training records
Input: This metric is computed using the following parameters:

N	Number of relevant employees
T	Number of employees who have received training

Formulation and output: Output = (T/N)*100
References: CCM v3 (BCR-11, CCC-02, HRS-10). GAPP (1.2.7, 1.2.9, 1.2.10). NIST SP 800-53 R4 (AR-5).
Note: This metric can be decomposed in two, if it is necessary to distinguish privacy and security training.

4.1.10 Metric 10. Account of Privacy and Security Training

Description: This metric describes the quality of the accounts given with respect to the privacy training and awareness programs in place.
Accountability Attributes: Verifiability
Associated Evidence: Training records
Formulation and output:

- Level 0 – No records of training are maintained.
- Level 1 – Records of training sessions are maintained, but there is no evidence of individual attendance.
- Level 2 – Individual records of attendance are maintained.
- Level 3 – Individual evaluation of the training contents is performed and recorded.
- Level 4 – The training program includes automated procedures for recording attendance as well as for evaluating personnel individually.

References: CCM v3 (BCR-11, CCC-02, HRS-10). GAPP (1.2.7, 1.2.9, 1.2.10). NIST SP 800-53 R4 (AR-5).

Note: This metric can be decomposed in two, if it is necessary to distinguish privacy and security training.

4.1.11 Metric 11. Level of confidentiality

Description: This metric indicates the level of confidentiality achieved by a system regarding client data independently of the means used to achieve this objective.
Accountability Attributes: Verifiability
Associated Evidence: Encryption policies
Formulation and output:

- Level 0 – Data confidentiality does not satisfy any of the next levels.
- Level 1 – Data may be accessible by the cloud provider personnel for regular operational purposes, under the control of an authentication, authorization and accounting (AAA) mechanism.
- Level 2 – Technical and organizational measures are in place so that data may only be accessible to privileged CSP personnel (administrators) for debugging or maintenance purposes, under the control of an AAA mechanism.
- Level 3 – Technical and organizational measures are in place so that data is only accessible to privileged CSP personnel to respond to law enforcement or extraordinary requests made by the client, under the control of an AAA mechanism.
- Level 4 – Data is encrypted by the client with cryptographic keys that cannot be ascertained by the provider.

References: CCM v3 (EKM-01, EKM-04).

4.1.12 Metric 12. Key Exposure Level

Description: This metric indicator of key exposure to reflect the level of confidentiality afforded to cryptographic secrets, from a cloud client point of view.
Accountability Attributes: Verifiability
Associated Evidence: Encryption policies
Formulation and output:

- Level 0 – Access to decrypted data or cryptographic secrets by the CSP is necessary to provide some functionalities of the service.
- Level 1 – Access to decrypted data or cryptographic secrets is available to specific personnel of the CSP, for administrative or debugging purposes only.
- Level 2 – Access to decrypted data or cryptographic secrets is available to specific personnel of the CSP, for administrative or debugging purposes only. It is governed by the principle of dual control and split knowledge.
- Level 3 – Access to decrypted data or cryptographic secrets is available to specific personnel of the CSP in exceptional circumstances only. It is governed by the principle of dual control and split knowledge, under the supervision of a hardware security module.
- Level 4 – Cryptographic secrets needed to decrypt the data are known to the cloud client only.

References: CCM v3 (EKM-01, EKM-04).

4.1.13 Metric 13. Data Isolation Testing Level

Description: This metric describes the level of testing that has been done by the cloud provider to assess how well data isolation is implemented.
Accountability Attributes: Verifiability
Associated Evidence: Isolation tests records
Formulation and output:

- Level 0 – No data isolation testing has been performed.
- Level 1 – Read/write isolation has been tested.
- Level 2 – Secure deletion has been tested, in addition to read/write isolation.
- Level 3 – Absence of known side channel attacks has been tested, in addition to read/write and secure deletion.

References: CCM v3 (IVS-09)

Note: In order to use such a metric, the resources in the scope of the measurement need to be well defined (storage, CPU, network, memory, database, etc.). Additionally, a standard set of tools or procedures need to be defined to establish the tests that should be conducted to assess each level.

4.2 Transparency, Responsibility and Attributability

This set of metrics is intended for measuring the characteristics about the internal processes that provide Accountability. In particular, we focus here on the ones that influence the "Transparency", "Responsibility" and "Attributability" attributes.

4.2.1 Metric 14. Type of Consent

Description: This metric describes the type of consent obtained for collecting, using and sharing private data. The type of consent can be ranked in levels according to its preference.
Accountability Attributes: Transparency.
Associated Evidence: Privacy Policies.
Formulation and output:

- Level 0 – No Consent: Consent is not obtained at or before collection of private data.
- Level 1 – Implied Consent: The consent is infered from the behaviour of the data subject, or even from failing to explicitly object. No opt-out or opt-in mechanisms are offered.
- Level 2 – Opt-out Consent: Data subjects can take measures for prevent the collection of private data, but no opt-in mechanisms are offered.
- Level 3 – Opt-in Consent: Data subjects explicitly grant permission for collecting or using private data.

References: NIST SP 800-53 R4 (IP-1). GAPP (3.2.1)

4.2.2 Metric 15. Type of notice

Description: This metric describes the type of privacy notice provided by the collecting organization, depending on how the privacy notice is offered to the data subjects. Ideally, multi-layer notice should be provided so data subjects have the information necessary to make decisions at any point in time.
Accountability Attributes: Transparency.
Associated Evidence: Privacy Notice
Formulation and output:

- Level 1 – Single notice: The organization provides only a single document describing the privacy notice.
- Level 2 – Multi-layer notice: The organization provides different layers of notices. Each layer can present different degrees of information, as long as the union of all the layers is compliant with applicable privacy regulations.

References: NIST SP 800-53 R4 (TR-1). Opinion 10/2004 on More Harmonised Information Provisions [7].

4.2.3 Metric 16. Procedures for Data Subject Access Requests

Description: This metric describes the quality of the procedures in place for guaranteeing data subjects' access to their personal information.
Accountability Attributes: Transparency.
Associated Evidence: Privacy Program.
Formulation and output:

- Level 0 - No procedures are established for permitting data subject access to their personal information.
- Level 1 - Procedures for data subject access exist but are not documented or consistent.
- Level 2 - Documented and consistent processes for data subject access are established. Employees responsible of such procedures are identified and trained on how to respond to requests. There also exist procedures for handling with denial of acess.

- Level 3 - Automated and self-service procedures for data subject access are in place, including the case of denied access.

References: NIST SP 800-53 R4 (IP-2). GAPP (6.2.1, 6.2.4)

4.2.4 Metric 17. Number of Data Subject Access Requests

Description: This metric describes the number of data subject acces requests received during a given period of time.
Accountability Attributes: Transparency.
Associated Evidence: Access requests records.
Input: This metric is computed using the following parameters:

N	Number of data subject access requests received during a given period of time

Formulation and output: Output = N
References: NIST SP 800-53 R4 (IP-2). GAPP (6.2.1, 6.2.4)

4.2.5 Metric 18. Responded data subject access requests

Description: This metric describes the percentage of data subject access requests that have been responded and for which a record of the request and the response exists.
Accountability Attributes: Observability, Transparency
Associated Evidence: Access request records.
Input: This metric is computed using the following parameters:

R	Number of responses to data subject access requests
N	Number of data subject access requests received during a given period of time

Formulation and output: Output = (R/N)*100
References: NIST SP 800-53 R4 (IP-2). GAPP (6.2.1, 6.2.4)

4.2.6 Metric 19. Mean time for responding Data Subject Access Requests

Description: This metric indicates the mean time for responding to data subject access requests
Accountability Attributes: Transparency.
Associated Evidence: Records of data subject access requests.
Input: This metric is computed using the following parameters:

T_i	Response time for access request i (expressed in a given time unit, such as hours)
N	Total number of data subject access requests, for a given period of time

Formulation and output: Output = $\frac{1}{N}\sum_i^N T\_i$
References: NIST SP 800-53 R4 (IP-2). GAPP (6.2.3)

4.2.7 Metric 20. Readibility (Flesch Reading Ease Test)

Description: This metric describes quantitatively the level of readibility of a given text, computed from the number of sentences, words and syllables. This is of interest for assessing readibility of privacy notices and notifications, which should be written in a clear and concise way. This metric is known as the Flesch Reading Ease Test, and is widely utilized for evaluating readibility (e.g., in [13], for readibility of insurance policies).
Accountability Attributes: Transparency.
Associated Evidence: Privacy notices, Notifications.
Input: This metric is computed using the following parameters:

S	Total number of sentences in the text
W	Total number of words in the text
Y	Total number of syllables in the text

Formulation and output: Output $= 206.835 - 1.015 \cdot \frac{W}{S} - 84.6 \cdot \frac{Y}{W}$
Interpretation: The goal is to maximize the value of the metric. A minimum of 45 could be fixed in order to consider the text reasonably readable. This minimum score is used in other contexts for assessing readibility of policies, such as life insurance policies.
References: NIST SP 800-53 R4 (IP-1). GAPP (2.2.3, 3.1.1)

4.2.8 Metric 21. Rank of Responsibility for Privacy

Description: This metric describes numerically at what level within the organization hierarchy the person responsible for privacy is located.
Accountability Attributes: Responsibility, Liability.
Associated Evidence: Organization's Privacy Program.
Input: This metric is computed using the following parameters:
N Number of superiors of the person responsible for privacy in the organization's hierarchy
Formulation and output: Output = N
Note: The goal is to minimize this metric, so the person responsible for privacy is located near the top of the organization (e.g., CEO).
References: NIST SP 800-53 R4 (AR-1). GAPP (1.1.2).

4.2.9 Metric 22. Certification of acceptance of responsibility

Description: This metric describes the percentage of employees who have certified their acceptance of responsibilities for activities that involve handling of private data.
Accountability Attributes: Responsibility, Liability, Verifiability
Associated Evidence: Certificates of acceptance of responsibility from relevant employees. These certificates could be either electronic or in paper, but tangible evidence of the acknowledgement of responsibility should exist.
Input: This metric is computed using the following parameters:
N Number of employees that are responsible of handling private data
A Number of employees who have effectively certified their responsibility
Formulation and output: Output = (A/N)*100
References: NIST SP 800-53 R4 (AR-5). CCM v3 (BCR-11, HRS-11, SEF-03). GAPP (1.1.1).

4.2.10 Metric 23. Frequency of certifications

Description: This metric describes how often employees certify their acceptance of responsibilities for activities that involve handling of private data.
Accountability Attributes: Verifiability, Responsibility, Liability.
Associated Evidence: Organization's Privacy Program.
Input: This metric is computed by using the following parameters:
N Number of scheduled certifications of acceptance, per year
Formulation and output: Output = N
References: NIST SP 800-53 R4 (AR-5). CCM v3 (BCR-11, HRS-11, SEF-03). GAPP (1.1.1).

4.2.11 Metric 24. Log Unalterability

Description: This metric describes the level of protection of the log management systems against tampering.
Accountability Attributes: Attributability, Verifiability
Associated Evidence: Description of the Log Management System
Formulation and output:

- Level 0 – No integrity mechanisms are in place
- Level 1 – Log integrity is protected only by access control measures.
- Level 2 – Cryptographic mechanisms are in place for guaranteeing log unalterability or WORM (Write Once Read Many) devices are used.

References: CCM v3 (IAM-01).

4.2.12 Metric 25. Identity Assurance

Description: This metric describes the quality of the authentication mechanisms in place.
Accountability Attributes: Verifiability, Attributability
Associated Evidence: …

Formulation and output:

- Level 0 – No authentication mechanisms are in place.
- Level 1 – Simple challenge response mechanisms are allowed and no identity proofing is required.
- Level 2 – Single factor remote network authentication is required; in this case, authentication is successful if the claimant proves control of the authentication token through a secure authentication protocol.
- Level 3 – Multifactor authentication mechanisms are in place. Proofs of control of the authentication token are done through a cryptographic protocol.
- Level 4 - Multifactor authentication with a hardware cryptographic token is required. Strong cryptographic mechanisms are required along physical tokens with a FIPS 140-2 level greater than 2, and identity proofing is done in person.

References: CCM v3 (IAM-01, IAM-02, IAM-12). GAPP (6.2.2, 8.2.2)
Note: This metric is extracted from the NIST standard 800-63-1.

4.2.13 Metric 26. Mean time to revoke users

Description: This attribute describes quantitatively how fast an organization revokes users' access to data and organizationally-owned or managed (physical and virtual) applications, infrastructure systems, and network components, based on user's change in status (e.g., termination of employment or other business relationship, job change or transfer).
Accountability Attributes: Attributability, Responsibility.
Associated Evidence: Records of user revocation.
Input: This metric is computed by using the following parameters:

T_i	Revocation time for user i (expressed in a given time unit, such as hours)
N	Total number of user revocation, for a given period of time

Formulation and output: Output = $\frac{1}{N}\sum_{i}^{N} T\_i$
References: CCM v3 (IAM-11, IAM-02).

4.3 Remediability and Incident Response

This set of metrics is devoted to measuring aspects related to remediation, redress, and incident response. The concept of Accountability goes beyond showing compliance and providing transparency, and must include also response to threats, incidents and failures to comply. These aspects are subsumed under the "Remediability" attribute.

4.3.1 Metric 27. Mean time to respond to complaints

Description: This metric indicates the average time that the organization takes for responding to complaints from stakeholders.
Accountability Attributes: Remediability, Transparency.
Associated Evidence: Records of complaints and resolutions.
Input: This metric is computed using the following parameters:

T_i	Response time for complaint i
N	Total number of complaints

Formulation and output: Output = $\frac{1}{N}\sum_{i}^{N} T\_i$
References: NIST SP 800-53 R4 (IP-4). GAPP (10.2.1, 10.2.2)

4.3.2 Metric 28. Number of complaints

Description: This metric indicates the number of complaints received during a given period of time.
Accountability Attributes: Transparency, Remediability.
Associated Evidence: Records of complaints.
Input: This metric is computed using the following parameters:

N	Number of complaints received during a given period of time

Formulation and output: Output = N

References: NIST SP 800-53 R4 (IP-4). GAPP (10.2.1, 10.2.2)

4.3.3 Metric 29. Reviewed complaints

Description: This metric indicates the percentage of complaints that have been reviewed during a given period of time.
Accountability Attributes: Remediability, Transparency.
Associated Evidence: Records of complaints.
Input: This metric is computed using the following parameters:

R	Number of complaints that have been reviewed during a given period of time
T	Total number of complaints received during a given period of time

Formulation and output: Output = (R/T)*100
References: NIST SP 800-53 R4 (IP-4). GAPP (10.2.1, 10.2.2)

4.3.4 Metric 30. Number of privacy incidents

Description: This metric provides the number of privacy incidents and breaches that have occurred in a given period of time.
Accountability Attributes: Transparency, Observability.
Associated Evidence: Incident Management records
Input: This metric is computed using the following parameters:

N	Number of privacy incidents and breaches over a given period

Formulation and output: Output = N
References: GAPP (1.2.7). CCM v3 (SEF-04, STA-05)

4.3.5 Metric 31. Coverage of incident notifications

Description: This metric provides the percentage of privacy incidents and breaches for which affected stakeholders were notified, for a given period of time.
Accountability Attributes: Transparency, Observability, Remediability.
Associated Evidence: Incident Management records
Input: This metric is computed using the following parameters:

N	Number of privacy incidents for which notification exists
T	Total number of privacy incidents over a given period

Formulation and output: Output = (N/T)*100
References: GAPP (1.2.7). CCM v3 (SEF-04, STA-05)

4.3.6 Metric 32. Type of incident notification

Description: This metric describes the quality of the notification procedures after a privacy incident or breach.
Accountability Attributes: Transparency, Remediation.
Associated Evidence: Incident notifications, Privacy Program.
Formulation and output:

- Level 0 – No notification of privacy incidents is done, or it is done inconsistently.
- Level 1 – General notification, usually as a public notice. Affected users may not be aware of the incident
- Level 2 – Individual notification to each affected user.
- Level 3 – Automated and self-service procedures for data subject access are in place, including the case of denied access.

References: GAPP (1.2.7). CCM v3 (SEF-04, STA-05)

4.3.7 Metric 33. Privacy incidents caused by third parties

Description: This metric indicates the number of privacy incidents caused by a third party to whom personal information was transferred (i.e. Data Processors)
Accountability Attributes: Transparency, Remediation, Observability.

Associated Evidence: Incident management records.
Input: This metric is computed using the following parameters:
N Number of privacy incidents and breaches over a given period caused by a third party processor
Formulation and output: Output = N
References: GAPP (7.2.4). CCM v3 (SEF-04, STA-05)

4.3.8 Metric 34. Number of Business Continuity Resilience (BCR) plans tested

Description: This metric indicates the number of business continuity resilience and incident response plans that have been tested in a given interval of time.
Accountability Attributes: Verifiability, Remediability
Associated Evidence: Records of BCR plans
Input: This metric is computed using the following parameters:
N Number of BCR plans tests in a given period of time
Formulation and output: Output = N
References: CCM v3 (BRC-02)

4.3.9 Metric 35. Maximum tolerable period for disruption (MTPD)

Description: This metric indicates the maximum tolerable period for disruption, as defined by the organizations' BCR plans.
Accountability Attributes: Remediability
Associated Evidence: BCR plans
Input: This metric is computed using the following parameters:
MTPD Duration of the maximum tolerable period for disruption, expressed in a given time unit (e.g. minutes)
Formulation and output: Output = MTPD.
References: CCM v3 (BCR-09)

4.3.10 Metric 36. Sanctions

Description: This metric indicates the number and type of sanctions that the organization has received. The EU DPD defines different types of sanctions: (i) a notice addressed to the Data controller (e.g. for compulsory audit), (ii) a fine, (iii) an injunction dictating the end of processing operations, and (iv) a (temporary or permanent) revocation of the authorization allowing the processing of personal data.
Accountability Attributes: Remediability, Transparency, Liability
Associated Evidence: Records of sanctions
Input: This metric is computed using the following parameters:

N1	Number of sanctions received in the form of notices
N2	Number of sanctions received in the form of fines
N3	Number of sanctions received in the form of injuctions
N4	Number of sanctions received in the form of revocations

Formulation and output: Output = (N1, N2, N3, N4)
References: CCM v3 (STA-02)

4.3.11 Metric 37. Incidents with damages

Description: This metric indicates the number of incidents that end up with compensatory or punitive damages.
Accountability Attributes: Remediability, Transparency, Liability
Associated Evidence: Records of incidents
Input: This metric is computed using the following parameters:
N Number of incidents that end up with compensatory or punitive damages, in a given period of time
Formulation and output: Output = N
References: CCM v3 (STA-02)

4.3.12 Metric 38. Total expenses due to compensatory damages

Description: This metric indicates the total expenses incurred due to compensatory damages.
Accountability Attributes: Remediability, Transparency, Liability
Associated Evidence: Records of incidents
Input: This metric is computed using the following parameters:

E_i Expenses due to compensatory damage associated to incident i (expressed in a given currency, such as Euros)

N Total number of incidents that incurred in damages

Formulation and output: Output = $\sum_{1}^{N} E\_i$
References: CCM v3 (STA-02)

4.3.13 Metric 39. Average expenses due to compensatory damages

Description: This metric indicates the average expenses due to compensatory damages per upheld complaint/incident
Accountability Attributes: Remediability, Transparency, Liability
Associated Evidence: Records of incidents
Input: This metric is computed using the following parameters:

E_i Expenses due to compensatory damage associated to incident i (expressed in a given currency, such as Euros)

N Total number of incidents that incurred in damages

Formulation and output: Output = $\frac{1}{N}\sum_{1}^{N} E\_i$
References: CCM v3 (STA-02)

4.4 Summary of the Accountability Metrics Catalogue

Table 2 shows a summary of the whole metrics catalogue and its association to the accountability attributes. For each metric, it is specified which accountability attributes are affected. It can be seen that some attributes, such as Transparency, Verifiability and Remediability are more prominently represented in the catalogue. On the contrary, some attributes such as Observability and Attributability were less present. As me mentioned earlier in this deliverable effectiveness and appropriateness were left out due to a timing question.

Metric	Name	Transparency	Verifiability	Attributability	Observability	Remediability	Responsibility	Liability
	Verifiability and Compliance							

1	Authorized collection of PII		X					
2	Privacy Program Budget		X					
3	Privacy Program Updates		X				X	
4	Periodicity of Privacy Impact Assessments for Information Systems		X					
5	Number of privacy audits received	X	X					
6	Successful audits received	X	X		X			
7	Record of Data Collection, Creation, and Update		X					
8	Data classification		X					
9	Coverage of Privacy and Security Training		X					
10	Account of Privacy and Security Training		X					
11	Level of confidentiality		X					
12	Key Exposure Level		X					
13	Data Isolation Testing Level		X					
	Transparency, Responsibility and Attributability							
14	Type of Consent	X						
15	Type of notice	X						
16	Procedures for Data Subject Access Requests	X						
17	Number of Data Subject Access Requests	X						
18	Responded data subject access requests	X			X			
19	Mean time for responding Data Subject Access Requests	X						
20	Readibility (Flesch Reading Ease Test)	X						
21	Rank of Responsibility for Privacy						X	X
22	Certification of acceptance of responsibility						X	X
23	Frequency of certifications		X				X	X
24	Log Unalterability		X	X				
25	Identity Assurance		X	X				
26	Mean time to revoke users			X			X	
	Remediability and Incident Response							
27	Mean time to respond to complaints	X				X		
28	Number of complaints	X				X		
29	Reviewed complaints	X				X		
30	Number of privacy incidents	X			X			
31	Coverage of incident notifications	X			X	X		
32	Type of incident notification	X				X		
33	Privacy incidents caused by third parties	X			X	X		
34	Number of Business Continuity Resilience (BCR) plans tested		X			X		
35	Maximum tolerable period for disruption (MTPD)					X		
36	Sanctions	X				X		X
37	Incidents with damages	X				X		X
38	Total expenses due to compensatory damages	X				X		X
39	Average expenses due to compensatory damages	X				X		X

Table 2: Summary Table of the Accountability Metrics

5 Extending the Metrics

In this section we describe how the concept of metrics can be extended for conveying more complex information. In particular we present two types of extensions:

- How to express confidence in metrics: The definition of the metrics can be extended for conveying not only the assessment done by the metric, but also a measure of the "confidence" of this assessment. An approach for expressing this aspect is presented in Section 5.1.
- How to define derived metrics: In the previous section we describe a catalogue of accountability metrics, which in fact are base metrics, according to the definition of the Metrics Metamodel. However, in some cases it may be interesting to define new metrics on top of metrics previously defined. That is, to transform and/or aggregate metrics. In Section 5.2 we describe how this can be achieved.

5.1 Expressing Confidence in Metrics

One of the main objectives of this deliverable is to include a method for deriving the level of confidence in the proposed metrics. The term "*confidence*" refers in this context to a measure of the assurance of the reliability of the metrics results. That is, a measure of how reliable is the result of a metric. In this section we describe an approach that can be used for extending the proposed metrics (and other metrics as well) in order to express a measure of the confidence on the assessment done by the metrics.

5.1.1 State of the Art

The idea of taking the confidence in the result of a metric into consideration is something that has been already treated in the literature. For example, in [32] and [33], a taxonomy of the quality metrics is proposed, with the intention of expressing the assurance in the security verification process. In this context, "quality" is referred to an assessment on the confidence of the constituent aspects of the verification process, namely coverage, rigour, depth and independence. A set of levels for each of these aspects is defined, as well as the criteria of assignment. This work is very relevant for us since we will follow a similar approach.

In [26], the authors propose a method for expressing "uncertainty" in a measure, which is the converse of "confidence" in this context. In this work, which is framed in the field of measuring trust, measures of trust can be represented by an interval, so the greater the interval, the higher the uncertainty of the measure. On the contrary, a narrow interval implies greater certainty in the measure. This approach is interesting, but assumes that there is a quantitative measure underneath, i.e., a metric with at least an interval scale. However, this rules out completely qualitative metrics, so in our case this approach is not valid.

In the Metrics Metamodel described in Section 3.1, as well as in [31], the quality of the evidence is considered as a prospective factor that could influence the metric. Although the concept of confidence is not explicitly mentioned, it is stated that the evidence for metrics may come from sources with different levels of certainty and validity, depending on the method of collection or generation of such evidence. That is, the notion of confidence associated to the source material for applying the metrics, i.e., the evidence, is considered implicitly. However, no further proposal is made to this respect.

In the field of computer security evaluation, the Common Criteria standard (ISO/IEC 15408) [18], defines the notion of Evaluation Assurance Level (EAL), an ordinal rating that indicates the depth and rigor of the evaluation processes of a system or a product. One of the key points of the definition of the Evaluation Assurance Levels is that higher levels do not ensure higher security of the system or product, but instead that the evaluation of the security has been more extensive and has counted with stricter verification standards. That is, this concept is analogous to the notion of confidence that we are trying to define. Hence, some aspects of the Common Criteria standard will influence our proposal for specifying the confidence in the metrics results.

5.1.2 A Formal Approach

In this section, we tackle how to formally define the idea behind the informal description of "confidence" presented at the beginning of the section. We formalize the idea of confidence in the metrics by expressing "Confidence" as an orthogonal dimension to the metric results, i.e. the measure itself. Figure 7 illustrates this concept as a bi-dimensional space, where "Measure" and "Confidence" are orthogonal dimensions. In this space, two identical measures (*m*) could have different levels of confidence (*c* and *c'*), as shown in the figure. Thus, it is clear that, in the case presented in this figure, the measure *M'*, which has a higher level of confidence, is preferable to measure *M*.

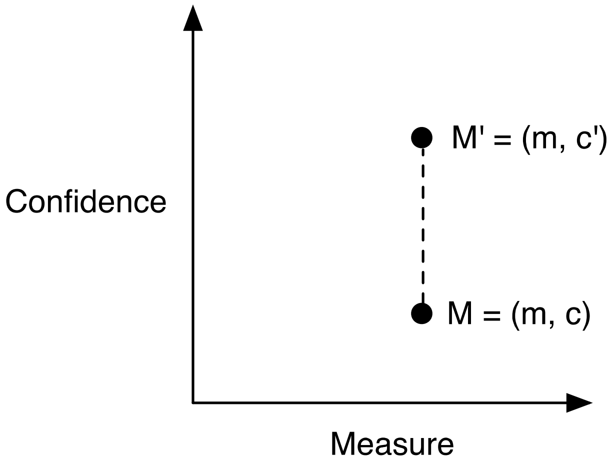

Figure 7: Measures and Confidence in a bi-dimensional space

We could formalize this preference by inducing a partial order over this bi-dimensional space. An example of partial order that captures this idea is the "*product order*". This type of partial order defines that, given two pairs (*m, c*) and (*m', c'*) in the bidimensional space *M* × *C*, then (*m, c*) ≤ (*m', c'*) if and only if *m* ≤ *m'* and *c* ≤ *c'*. That is, consider two different results, M and M', for applying a particular metric, then for the result M' to be greater than or equal to result M, it has to have a greater or equal measure result and a greater or equal confidence.

Once a partial order has been defined, certain pairs of elements can be compared. For example, let *M*, *M'*, and *M''* be elements in the bi-dimensional space, as shown in Figure 8. According to the product order defined before, element *M* can only be compared to elements in the grey area, such as element *M'*. Other elements, such as *M''* cannot be compared to *M* if we establish the product order.

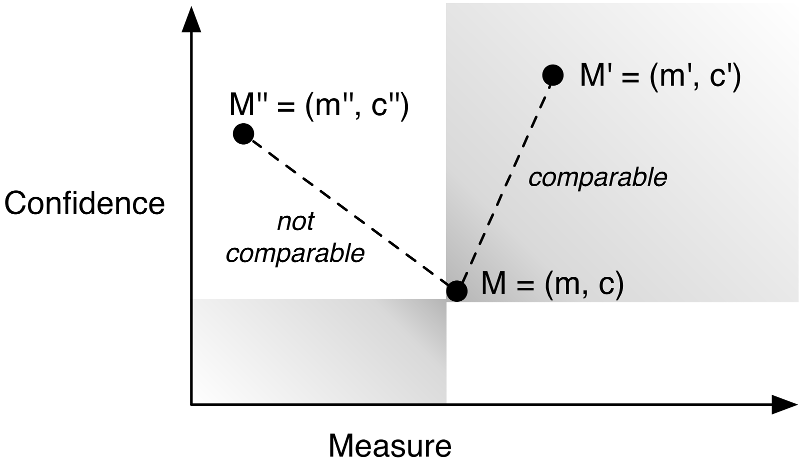

Figure 8: Comparable elements in the bi-dimensional space

The product order is a conservative but safe choice, since that way we ensure that in order to being capable of asserting that a measure, with an associated confidence level, is greater than other, this measure should be greater or equal and the confidence level be great or equal, as well; in other words, if its measure is greater but the confidence level is lower, then one cannot assert anything about the order relationship between the two elements. It is possible, however, to define other kinds of partial orders different to the product order (and even, if necessary, total orders).

5.1.3 Factors of Confidence

Once we have formalized how the confidence dimension fits with respect to the measures done by the metric, we have to define how the confidence level is devised. In order to facilitate the process of measuring the confidence in the metrics, a suitable approach is decomposing this concept in relevant factors. Recall that the notion of confidence we are referring to should indicate the quality, depth and rigour of the evaluation processes associated to the metric.

Based on the work presented in [32] and [33], we follow a similar approach for distinguishing what are the factors that influence the confidence in the metrics results. We identify two factors of confidence:

Source of assessment (C_SA).
This factor is devoted to the identification of the source of the assessment, that is, the actor that performs the evaluation of the metric. Depending on the independence of this actor with respect to the object of evaluation, we can identify several levels of confidence:

- **Level 1 (Self-assessment)**: In this case, the evaluation is performed by the same individuals or organizations that manage the object of evaluation. It is clear that, although one may trust the validity of the assessment and trustfulness of the actors that perform the evaluation, the independence is not formally fulfilled. Setting aside the quality of the assessment, the source of assessment in this case implies a lower level of confidence. An example of this kind of source of assessment is CSA Open Certification Framework (OCF) Level 1, which corresponds to self-assessment questionnaires from actual cloud service providers [10].
- **Level 2 (Third-party assessment)**: This level corresponds to an evaluation that is performed by a specialized and trusted third party, such as an auditor or certification body. In this case, we can assume that the assessment is performed in a fully independent entity, so total independence can be assumed. An example of this kind of source of assessment is CSA OCF Level 2, which corresponds to certification or attestation by authorized auditors [11].
- **Level 3 (User/Publicly Verifiable)**: This level refers to evaluations that can be directly performed by the interested stakeholders. Although in level 2, full independence is achieved from the point of view of the verification process, it is clear that freeing from the need of an intermediary entity is preferable and should be supported when possible, by providing technical and organizational means for the interested stakeholders to perform the evaluations by themselves.

Note that we are only defining an ordinal scale, so trying to reason about the “distance” between level 2 and level 3 makes no sense. That is, level 2 represents already represents a high level of independence. The important fact here is that we consider level 3 to be greater than level 2, given the aforementioned reasons. Level 3 is more aligned with the concept of Accountability since it facilitates the demonstrational facet of Accountability, by directly supporting some of its core attributes such as Observability, Verifiability and Transparency.

Consistency (C_CO)
This factor describes the level of regularity of the evaluation. For an evaluation to be consistent, a systematic and structured procedure must be followed. However, this is not always true in reality, and certain degrees of relaxation exist during assessments. This factor is identified in [32] and [33] as “Rigour” and it is reminiscent to the concept of Evaluation Assurance Level (EAL) in the Common Criteria framework [18]. In Common Criteria, these levels reflect the assurance requirements that must be fulfilled in order to achieve Common Criteria certification; a higher level implies better confidence in the security test procedures. The Consistency factor is similar to that concept. The following levels of Consistency are proposed:

- **Level 1 (Informal procedure)**: In this level, there is no formal procedure specified for performing the evaluation, or if it exists, no proofs of adherence to the procedures are provided.

- **Level 2 (Structured procedure)**: In this level, a formal procedure is defined, but the means for performing the evaluation are manual and possibly open to interpretation and subjectivity. Proofs of adherence to the procedures are available.
- **Level 3 (Automated procedure)**: The highest level of rigour corresponds to the case when a formal evaluation procedure exists and it is performed in a standardized fashion by means of an automatic mechanism. Proofs of adherence to the evaluation procedures are also available. An example of this kind of rigour level is CSA OCF Level 3, which corresponds to certification based on continuous monitoring [11].

Note that an evaluation with level 2 of consistency could be as exact as an automatic assessment. However, as noted in [32], [33], the latency of such evaluation would be much lower than an automatic one. Another important difference is the repeatability of the results, since an automatic method is presumably more accurate than a manual procedure. For this reason, the differentiation of level 2 and level 3 is made, since an automated procedure for evaluation would be preferable in a cloud-based setting like ours. As in the rest of cases, it is meaningless trying to figure out the magnitude of the difference between levels, as this is only an ordinal measure.

Table 3 presents a summary of the levels defined by the factors of Confidence:

Confidence (C)	Source of Assessment (C_SA)	Level 1: Self-assessment (C_SA 1)
		Level 2: Third-party assessment (C_SA 2)
		Level 3: User/Publicly Verifiable (C_SA 3)
	Consistency (C_CO)	Level 1: Informal Procedure (C_CO 1)
		Level 2: Structured Procedure (C_CO 2)
		Level 3: Automated Procedure (C_CO 3)

Table 3: Factors of Confidence

Although the two factors are theoretically independent, there may exist certain correlation between them. For example, a third-party assessment (C_SA Level 2) would probably have a consistency level (C_CO) of 2 or 3, since in most of cases the entity that performs the evaluation is a certified and professional organization, which presumably will follow high-quality procedures for the evaluation. Another example is a publicly verifiable assessment (C_SA Level 3) performed by an interested stakeholder that, however, does not count with the resources for performing a strict and rigorous evaluation, thus achieving an informal level of consistency (C_CO Level 1). These were examples, and all the permutations between these two factors are possible. However, these examples reflect different possibilities that affect the global confidence on the metrics results, and justify the identification of the factors of confidence that were presented in this section.

5.1.4 Establishing the Level of Confidence

Once the factors of Confidence are defined, we can aggregate both factors into a single measure of Confidence. Given that there are just two independent factors, we can set up a “Confidence matrix”, very similar in structure to the “Risk Matrix” typically used in the field of Risk Assessment. The Confidence Matrix is defined in Figure 9.

Consistency Source of Assessment	Informal (Level 1)	Structured (Level 2)	Automated (Level 3)
Self-assessment (Level 1)	**0**	**1**	**1**
Third party assessment (Level 2)	**1**	**2**	**2**
User/Publicly Verifiable (Level 3)	**1**	**2**	**3**

Figure 9: Metric Confidence Matrix

In this matrix, each combination of confidence factors produces a single Confidence Level that ranges from 0 to 3. These levels are defined as follows:

- **Level 0 (Unreliable)**: There is no confidence in the metrics results, since both the independence and the consistency of the assessment are very low.
- **Level 1 (Insufficient)**: In this case, one of the two factors only achieves the lowest level, so the global confidence value will be considered as insufficient. It is clear that confidence in metrics is insufficient when the assessment is self-made or the process is informal.
- **Level 2 (Essential)**: This level is the minimum desired level of confidence. The assessment guarantees an acceptable level of independence and consistency.
- **Level 3 (Maximum)**: This is the preferable level of confidence. However, achieving this level is presumably a costly procedure, since it implies automating the evaluation and making it publicly verifiable.

It is clear that both self-assessed and informally performed evaluations are not sufficient for providing a reliable metric, thus, the maximum attainable level of confidence for these two levels is 1. In particular, when the evaluation is both informal and self-assessed, the confidence is non-existent (level 0). Once both factors reach a level of 2, then an acceptable level of confidence is achieved (level 2). For the particular case when the evaluation is both publicly verifiable and automated, a maximum level of confidence is reached (level 3).

Note that the Confidence level defined above is just a coarse-grained indicator of the aggregation of the two factors of confidence. A finer grained indicator could be possible, but it would have more levels, which complicates its interpretation. Thus, the selection of this scale was done for the sake of simplicity and clarity.

5.2 Defining Derived Metrics

In Section 4 we describe a catalogue of accountability metrics, which are Base Metrics, according to the Metrics Metamodel. That is, these metrics only take as input information extracted from evidence. However, in some cases it may be interesting to define new metrics on top of metrics previously defined. To be more specific, it may be necessary to transform and/or aggregate previously defined metrics. These new metrics will be derived or compound metrics.

We distinguish two main ways of defining derived metrics:

- Aggregation of Metrics: In this case, a new metric is constructed using several metrics as input parameters. This could be mathematically abstracted as a function $F_A(m_1, m_2, ..., m_n) = m$, which takes as input a set of *n* metrics and produces as an aggregated result one metric *m*.

- Transformation of Metrics: In this case, a previously defined metric is simply transformed into another, usually for changing its scale (e.g. from an interval to an ordinal scale) or for normalizing its values. This could be mathematically abstracted as a function $F_T(m) = m'$, which takes as input a metric *m* and produces transformed metric *m'*.

5.2.1 Aggregation of Metrics

Metrics are usually gathered together as a catalogue, as we show them in Section 4, in most cases without any explicit categorization or ranking. However, from the decision-makers point of view, having simply a collection of uncategorized metrics could not be very useful; instead, it is preferable to count with a low number of metrics with aggregated values [36].

When we are dealing with numerical metrics (i.e., at least interval-based scales), then there are plenty of widely known arithmetic operations that could be performed for achieving aggregation [24], [25]. Methods such as averages, medians, min/max, etc. are common choices. However, The Ordered Weighted Averaging (OWA) aggregation operator, introduced by Yager in [40], generalizes most of these operations into a single definition. The OWA operator is defined by a function F, which takes as arguments a set $\vec{a}$ of input values in R^n and a set $\vec{w}$ of n weights in the unit interval. The function F first reorders the elements of the input $\vec{a}$ and generates an ordered set $\vec{b}$. Then, the function F works as follows:

$$F(\vec{a},\vec{w}) = \sum_{i=1}^{n} w_i\, b_i$$

Given this definition, common aggregation operators can be defined by just setting the proper weights. For example, the “Max” operator is achieved with the weights $\vec{w} = (1, 0, \dots, 0)$, the “Min” operator with the weights $\vec{w} = (0, 0, \dots, 0, 1)$, and the “Average” operator with the weights $\vec{w} = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$. The expressive power of OWA enables to define other operators in the space between traditional functions.

Most of the above methods assume that the values to be aggregated belong at least to an interval scale, so additions and substractions are possible. However, in the case of qualitative data, which usually belongs to an ordinal scale, it makes no sense to wonder about the addition or the difference between two ordinal values. Averages and weighted averages cannot be directly performed. To this end, certain techniques have been designed for operating directly with qualitative values (to be more precise, with ordinal data). In [16], and more recently in [21], methods for performing weighted means over ordinal data are presented. These methods are in turn based on concepts from fuzzy logic.

5.2.2 Transformation of Metrics

Derived metrics can also be defined through transformation, i.e., when a single metric is transformed into a new one, usually for changing its type of scale. We distinguish three main kinds of transformation of metrics:

- Normalizations: This kind of transformation is referred to the case when the values of a quantitative metric are operated in some way to adapt them to some reference domain. For example, it is very common to normalize a metric with respect to the unit interval (i.e., [0,1]) or to 0 to 100. Although not technically a normalization, or from increasing to decreasing, etc)
- Qualitative to Quantitative: One of the big problems of qualitative metrics is their difficulty of being part of aggregation and processing. Hence, a typical justification of transforming qualitative to quantitative metrics is for making possible complex processing, such as numerical operations [22] and [23]. It is much more common to aggregate qualitative data using this approach than the techniques described in the previous section for operating directly with qualitative information. Typical ways to achieve this kind of transformation is by simply defining a mapping from qualitative to numerical values (e.g. Low → 0.2). Examples of this kind of transformation are works such as [14] and [15].

- Quantitative to Qualitative: On the contrary, in some cases quantitative metrics may need to be converted to qualitative, in order to be easily interpreted by decision-makers (e.g. Good/Bad, Low/Medium/High). That is, although quantitative analysis is desirable when possible, at the end, for the final assessment it is preferable to count with a qualitative indicator. As in the previous case, this can be acheived by defining a mapping from the numerical to the qualitative values (e.g. 0.2 → Low).

6 Validation of Metrics for Accountability

One of the main objectives of this deliverable is to present the methodologies used for validating the metrics produced in this work package. According to the PMBOK Guide [34], "*Validation is the assurance that a product, service, or system meets the needs of the customer and other identified stakeholders. It often involves acceptance and suitability with external customers*". To this end, there are several approaches that we could follow in order to analyse the suitability of the metrics. We distinguish here several validation strategies that are applicable for validating the proposed metrics.

6.1 Simulation-based Validation

This kind of validation strategy is based on the definition of controlled experiments where simulated data is used for feeding the proposed metrics. The results are then analysed in order to identify inconsistencies and points for improvement. The disadvantages of this kind of validation strategy is that the simulation of the metrics is not an easy task, and in addition, the subsequent analysis does not guarantee to induce a judgement on the validity of the studied object (in this case the metrics). However, it could give a valuable insight on the operation and feasibility of the proposed metrics.

Given the definition of this validation strategy, it is clear then that it is necessary some kind of simulation environment in order to be extract information for the analysis. A natural place where this validation could be performed is in work package WP D-7 (Instantiation for use case), where a prototype of a use case for the project will be implemented. However, at the time of writing of this deliverable it is still not clear whether this environment could be used for simulating evidence suitable for metrics. A more important issue is that the timelines of both work packages (C-5 and D-7) do not facilitate a suitable interaction, as WP C-5 finishes in M24 and instantiation in D-7 will happen during months M20 to M41. Work package D-6, where the assertion of A4Cloud toolkit will be performed, is also another place where some simulations are to be done. However, the same problem arises, as the assertion task will not begin until M31. Thus, it is not realistic to go for this approach for validating the metrics proposed here.

Another option is to simulate the metrics aside from the project instantiation and the assertion toolkit. In some cases, real data that could be used as evidence for some metrics proposed in this deliverable is already available. In particular, the CSA STAR registry [12] is a public repository where assessment reports of cloud providers are gathered. Hence, the STAR registry constitutes a rich source of data. However, from the point of view of the confidence on the metrics results, these assessments are not ideal, as the information comes directly from the cloud provider. Additionally, only part of the metrics could be simulated this way, since not all the metrics are associated to controls from the Cloud Control Matrix, especially to its version 1.1. A simulation with this data would only obtain partial results.

6.2 Opinion-based Validation

In contrast to the other strategies, there are some metrics that, because of its nature, cannot be validated through the use of public or simulated data. For example, most of the metrics that involve purely subjective measures, such as users' satisfaction or users' perception, should be validated with information provided by the users themselves.

In this setting, interview-based methods for gathering users' opinions and experience could be very useful for validating the proposed metrics. Such methods include expert interviews, group discussions, etc. The goal of these methods is to gather direct feedback and suggestions for improvement directly from relevant stakeholders. These approaches have the added value of counting with a sample of the intended stakeholders of the metrics, such as IT professionals or end-users. One of the more prominent examples of this kind of methods is the Delphi methodology, which we will explain in the next section. As we discuss later on, we will adopt this methodology for performing an opinion-based validation of the proposed metrics.

6.2.1 Delphi Methodology

The Delphi methodology is a structured procedure based on surveys of expert opinions, which is usually used in forecasting and decision-making processes. The Delphi methodology requires the participation of a moderator (or a group of moderators), who prepares questionnaires and reviews the responses, and a group of experts, which responds anonymously to the questionnaires. The procedure in the Delphi methodology is iterative; in each round, expert opinion about a certain subject is surveyed by means of the questionnaires. At the end of the round, the moderator reviews the responses, and refines the questions based on the identified consensus and disagreement. The process is repeated several times, until it is reached a reasonable consensus or the moderator believes it is enough. Figure 10 shows the general process of the Delphi methodology.

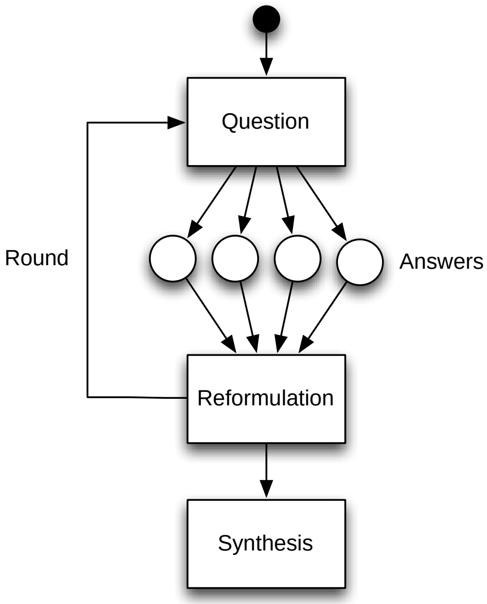

Figure 10: The Delphi methodology

It can be seen that the iterative process of the Delphi methodology supports the refinement of the surveyed questions, but at the same time, requires that experts participate in several rounds, which can be difficult and time consuming. Ideally, this process should be done in person and in one session, but the methodology is flexible, so it could be performed on-line and in different time periods. The only objective is to iteratively refine the research questions based on the opinions of the experts.

In order to elicit the answers to the research question, the use of questionnaires, rather than direct interviews, is preferable in order to avoid the “interviewer effect”, which is any impact (positive or negative) that the interviewer characteristics and behaviour induce on the responses of the interviewees.

6.2.2 Experts Feedback

A secondary means to validating the catalogue of metrics is to present it to a selection of experts, in the case belonging to the community of security and privacy metrics. The difference with the previous method is that, in this case, there is no defined structure at all in the process for asking for feedback. The catalogue is presented to the experts, and they give feedback based on their expertise.

6.3 Selection of the Validation Methodology

In this section, we will identify and justify which validation strategy is used for the proposed catalogue of Accountability Metrics. The first aspect that we must take into consideration is that the proposed metrics are of diverse nature, ranging from technical to organizational aspects. Hence, it would be difficult to find a “perfect” approach for validating all the metrics. Ideally, one could choose one strategy or another depending on the nature of the analysed metric, the associated evidence that is needed for realizing it, and the feasibility of the validation itself. This would imply using a combination of methods, depending on these factors, so the validation would be supposedly easier or more adequate for certain metrics than for others. However, it is impractical to come across with a validation strategy for each proposed metric, and a trade off solution that covers the most of metrics should be found.

Simulation-based validation is in principle an appealing approach, since it would add an experimental spirit to the validation phase. However, as we mentioned in Section 6.1, the feasibility of using this approach here is low, especially timewise. Furthermore, the analysis that has to be performed on the results of the simulation is far from trivial and does not guarantee that the metrics are correctly validated, rather than merely executed. Given these reasons, we conclude that simulation-based validation is not suitable for the validation phase of this work package, although it may be a very interesting way to complement the validation once a suitable simulation environment exists, so resulting data from simulations is adequate, from both the viewpoints of volume and relevance.

On the contrary, the great advantage of opinion-based validation is that, once metrics are defined, it is possible to gather opinions from different stakeholders regarding a wide variety of metrics, regardless its nature and technical complexity, since the opinion will be based on the description of the metric, rather than on an actual implementation. This implies that the a priori feasibility of the implementation of the metric is an aspect to be elicited, together with the rest of validation questions, in order to compensate for the lack of actual implementation.

With regard to the risks of opinion-based validation, one of the main issues is the possible lack of variety of the feedback, due to the appearance of the interviewer effect, which can produce a bias on the result. To this end, it is preferable to resort to methods that offer a structured and anonymous approach, such as the Delphi methodology.

Therefore, in this work package, we opted for a validation based on the Delphi methodology. This decision is based on the following rationale:

- Flexibility: the methodology can be adapted freely to the characteristics of the research question and the characteristics of the participant groups.
- Feasibility: we can ask for opinions for metrics whose simulation would be complicated or impractical.

Additionally, unstructured feedback from experts is also considered, as an auxiliary approach of eliciting feedback. Comments from some experts consulted by members of the work package were examined.

6.4 Validation of Metrics through the Delphi Methodology

As explained in Section 6.2.1, the Delphi methodology is based on surveying experts opinion through an iterative process. Thus, in order to design the validation strategy, we must first define who is participating in the validation in the role of experts and what are the questions we are going to ask them.

Most of the stakeholders of accountability metrics will have a professional background, such as IT technicians, decision-makers, and legal staff. Then, it seems reasonable that the selected sample of experts is representative of this kind of profile. We arranged a series of rounds of validation, as stipulated by the Delphi methodology. The first validation round was in-person, as part of the stakeholders Workshop 4 from WPB-2, which was held in Málaga, Spain, at UMA premises on September 3rd, 2014. It counted with a group of 18 participants from the Ada Byron Research Institute and the Computer Science Department, both from the Universidad de Málaga. This validation session was intended as a starting point of the validation process, and its output was planned to be refined in further rounds of validation. We refer to this session as the “Round 1” of our application of the Delphi methodology. Individual follow-ups for most of the participants (specifically, 14 of them) from the first round, constituted a second round of validation. We refer to this session as “Round 2”.
At the same time, online validation was also possible through the SurveyMonkey system (https://es.surveymonkey.com/s/NB5T8ZK), and publicized through CSA social networks and other channels, including e-mail and in-person promotion at events (such as the ERCIM Working Group Security and Trust Management meeting and the A4Cloud training tutorial, co-located with the IFIP Summer School on Privacy and Identity Management). This online validation round was made available

before the first validation session, so we will consider it as the initial round of validation, although its results could not influence the questions of the first validation session. We refer to this session as "Round 0".

With regards to the content of the validation sessions, we prepared a set of questions regarding the accountability metrics catalogue. Given the size of the catalogue, of approximately 40 metrics, we strived to keep the questions short. In our approach, the experts evaluate the metrics catalogue through some general questions, but at the same time are given the liberty of asking or discussing about any particular metric. This way, the size of the questionnaire is kept short, but there is room for discussing specific aspects if needed. Appendix shows the content of the questionnaire that was used during the validation rounds.

The questionnaire contained three questions in the form of statements about the respondents' opinions with a five-point scale: strongly disagree (1), disagree (2), neither agree nor disagree (3), agree (4), strongly agree (5). These questions were:

- Q1: "*This set of metrics contains meaningful and relevant measures for Accountability in the Cloud*". With this question, we wanted to analyze the level of appropriateness of the catalogue for measuring the concept of Accountability in the Cloud.
- Q2: "*The use and application of this set of metrics would be easy, in general*". The goal of this question is to assess the perceived degree of feasibility of the metrics proposed.
- Q3: "*This set of metrics can be easily understood by a professional audience*". The goal of this question is to evaluate the degree of usability of the catalogue with respect to the facility of being understood by professionals. We focused on professionals since this part of the stakeholders are the ones that most likely will apply and benefit from the metrics for accountability, due the specialization of some of the metrics. The general public (i.e., cloud end-users) needs much more simplified and aggregated information, so we did not considerate for this question.

The motivation behind the election of these questions was twofold. Firstly, past experience from partners of the work package has shown that it is difficult to gather responses to surveys if there is too many questions. Thus, questions should be concise and kept to the minimum. Secondly, we wanted to evaluate the metrics with respect to the most relevant quality criteria for validation. It is clear that there are several aspects that could be assessed for facing the validation. In [35], the author identifies three core quality criteria for security metrics, namely correctness, measurability, and meaningfulness; a fourth criterion, usability, is also found to be very relevant. In relation to the quality criteria proposed by this work, we can find some parallelism with our questions. Our first question is oriented towards eliciting the opinion regarding the correctness and meaningfulness of the metrics, the second question tries to evaluate the metrics with respect to the measurability dimension, and the third question is focused on the usability aspects of the metrics catalogue.

In the original Delphi methodology, the participants are involved through several rounds; however, given the difficulty of engaging a moderately big group of participants during the whole process, we adapted the methodology so the subsequent round after the in-person session was performed individually, in an ad hoc manner. The results of each round were analyzed and changes on the catalogue of metrics were made in order to refine the input for the next round.

6.5 Rounds of Validation

As mentioned before, the validation process took place in rounds, as prescribed by the Delphi methodology. In this section we describe these rounds and its results, which are quantitatively analyzed in order to get insights about the perceived validity of the catalogue.

6.5.1 Round 0 – Online Survey

As mentioned before, an online survey was available from the beginning of the validation through the SurveyMonkey system (https://es.surveymonkey.com/s/NB5T8ZK). This survey was publicized through different channels, such as email lists, Linkedin and CSA's social networks. Although, this online validation round was made available before the first validation session, its results could not influence the questions of the first validation session since we did not close the survey until the end of the validation period. We analysed its responses and used our findings for the validation.

The responses to the questionnaire were very satisfactory. In particular Q3, which assessed the understandability of the catalogue, had a very good score (4.30), as shown in Figure 11. Question Q1, corresponding to the meaningfulness of the catalogue, had also a good rating (4.20). The results for Q2 (3.40) although good, where closer to the neutral value.

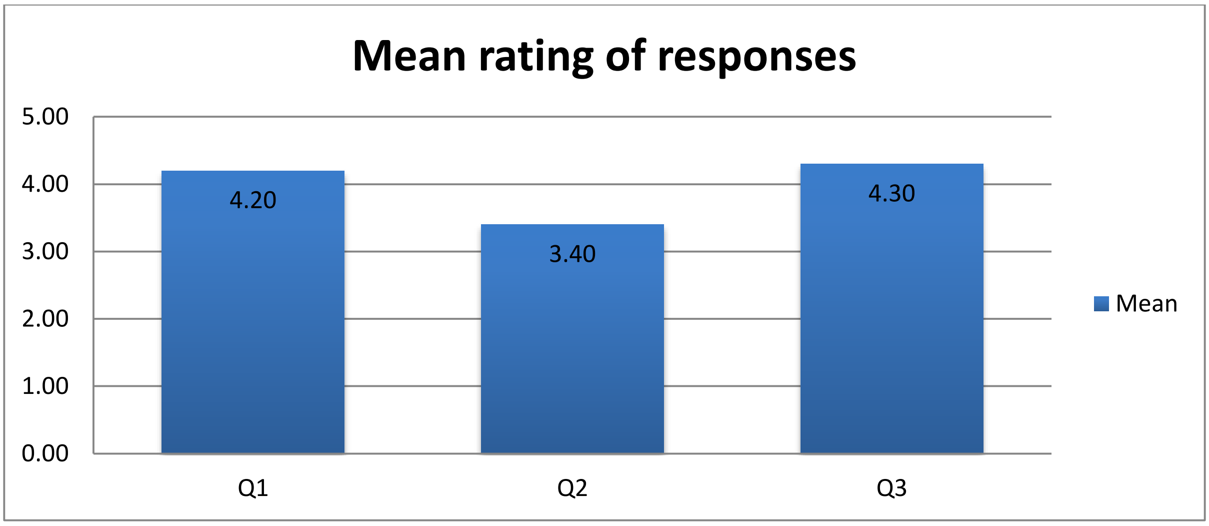

Figure 11: Mean rating of responses (Round 0)

More detail is shown in Figure 12, which shows the distribution of responses with regard to the available options. It can be seen that, except for Q2, there are not “Disagree” or “Strongly disagree” responses. Similar results were obtained on the next round of validation.

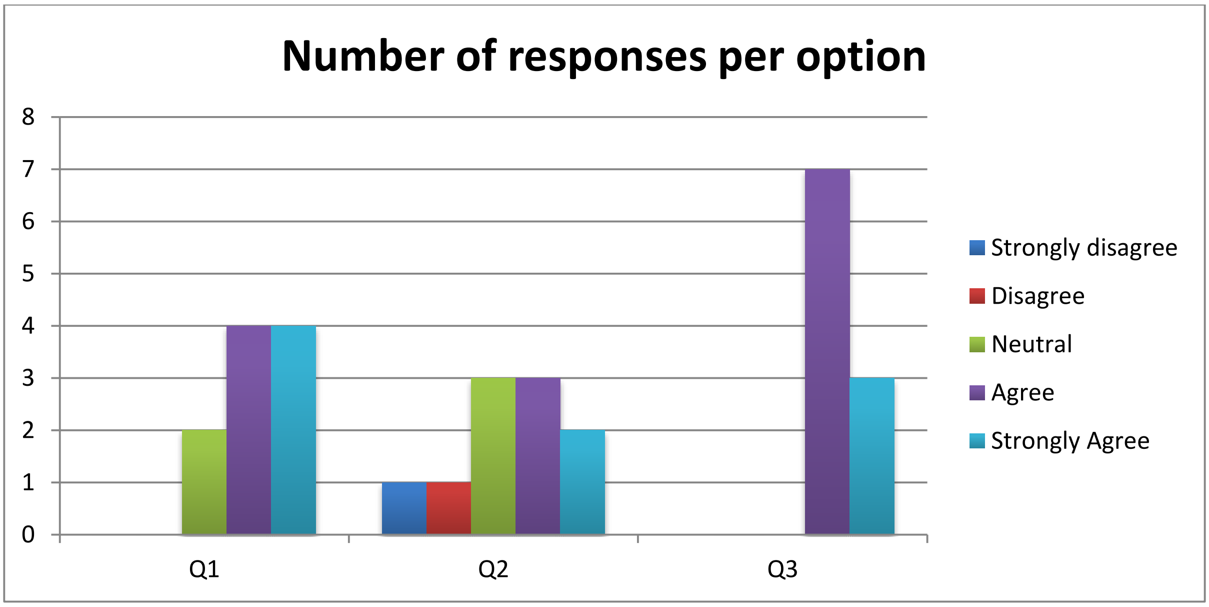

Figure 12: Distribution of responses (Round 0)

The questionnaire also allowed respondants to express their own opinions and comments. The following are the most relevant comments:

- “*I am moderately skeptic about the feasibility of certain metrics. To be more exact, of its acceptance by the cloud service providers, since these metrics can show weaknesses of their internal processes*”. This is true to a certain extent. Indeed, some of these metrics could expose weaknessess of the internal processes of cloud providers, but precisely, this would imply supporting Observability and Transparency, and hence, Accountability.

- "*The metric values should be normalized*". Normalization of the proposed metrics is not always possible. As explained before, normalization implies the transformation of the scale used to some reference scale. Formal justification of this transformation is not allways possible, since implies choosing a reference scale and the criteria and methods for adjusting the scales. We must also note that certain metrics, such as the ones that are given in the form of percentages, are trivially normalizable.

6.5.2 Round 1 – First Validation Session

This round was the one with more participation, and, moreover, the one where we received more feedback from the participants. There were two main causes for this: firstly, on the contrary to the online survey, in this session we could present the catalogue and background information in person, so a two-way communication could be established satisfactorily; and secondly, the allotted time for explaining the catalogue and its objectives was enough for the participants to participate adequately. For these reason, this was in our view the key round of the validation process.

As it happened on the on-line survey, response to the questionnaire was, in general, very satisfactory. The same pattern was repeated with little variation, as shown in Figure 13: question Q1 had a good rating (3.89), much like question Q3 (4.28); the average rating for question Q2 (3.44), although good, was also closer to the neutral value. These results, and the fact that there were similar to the on-line survey results, may indicate that the presented catalogue is in general well received, but there is a concern with the perceived difficulty of the respondents with respect to the feasibility of applying catalogue. After analyzing the free-text comments, this scepticism was due to the feeling that it would be difficult to encourage providers to adopt it. This is discussed further above.

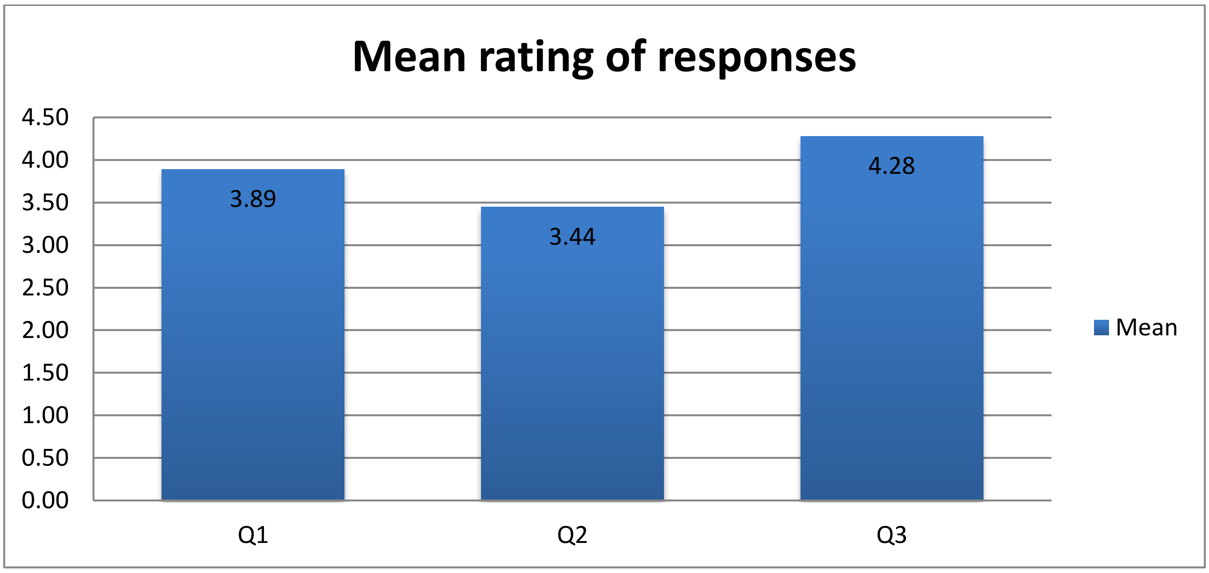

Figure 13: Mean rating of responses (Round 1)

In order to obtain a more detailed insight, Figure 14 shows the distribution of responses per option. This distribution is very similar to the result of the online survey, although slightly more diverse.

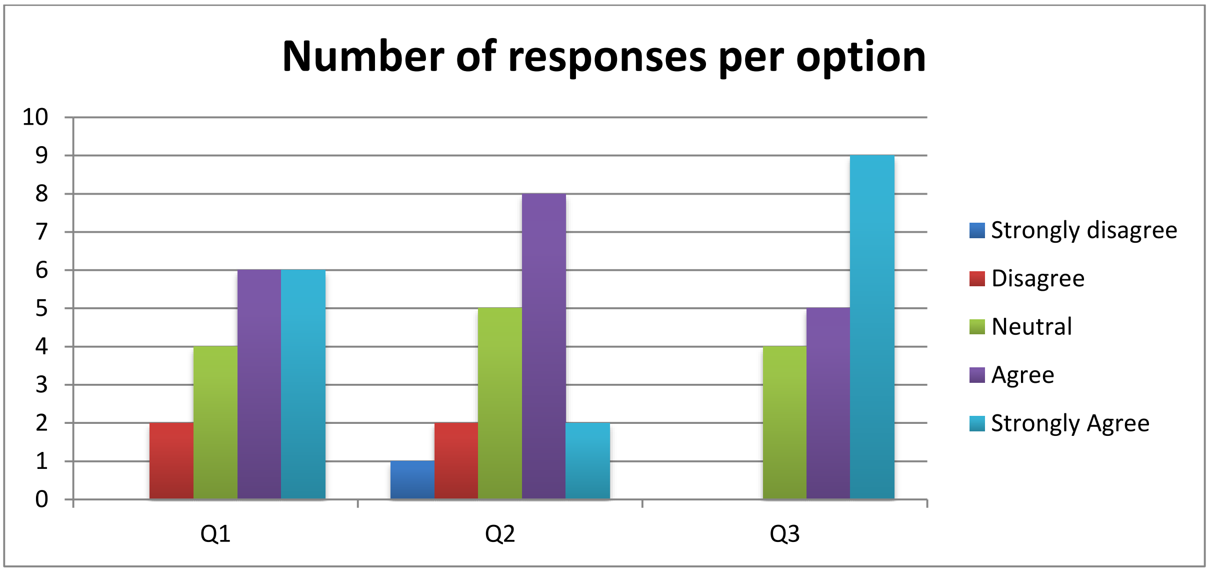

Figure 14: Distribution of responses (Round 1)

Some of the most remarkable comments received during this session were:

- "*The catalogue is very complete and reflects appropriately several facets of Accountability, however, the difficult part is to engage cloud service providers for utilizing these metrics. I wonder how are you going to tackle this*". Although this comment is not strictly directed to the catalogue itself, we believe is highly relevant for the success of the work package itself.
- *"Several metrics seem to be based on information coming from self assessments, which is not very useful*". Indeed, there are several metrics that are based on evidence that is usually self-assessed. To this end, the confidence on the metrics, as described in Section 5.1, tries to tackle this issue by expressing the level of independecy in the "source of assessment" factor.

6.5.3 Round 2 – Follow-up

Most of the participants of the first round of validation were willing to take part in a second round of validation. Since this round did not imply a huge variation of the catalogue with respect to the previous round, there was no need for repeating an in-person meeting with all the participants. Instead, a refined version of the catalogue, together with better explanation of its objectives and motivation, was distributed individually, and responses were gathered one at a time, as well.

As shown in Figure 15, and with more detail in Figure 16, results were very similar to the previous round, although slightly higher ratings were obtained. This time, question Q1 was rated higher (4.07) than the "Agree" level, which corresponds to a rating of 4. Question Q2 also increased, although it did not surpass the agree level. Rating of question Q3 remained practically the same.

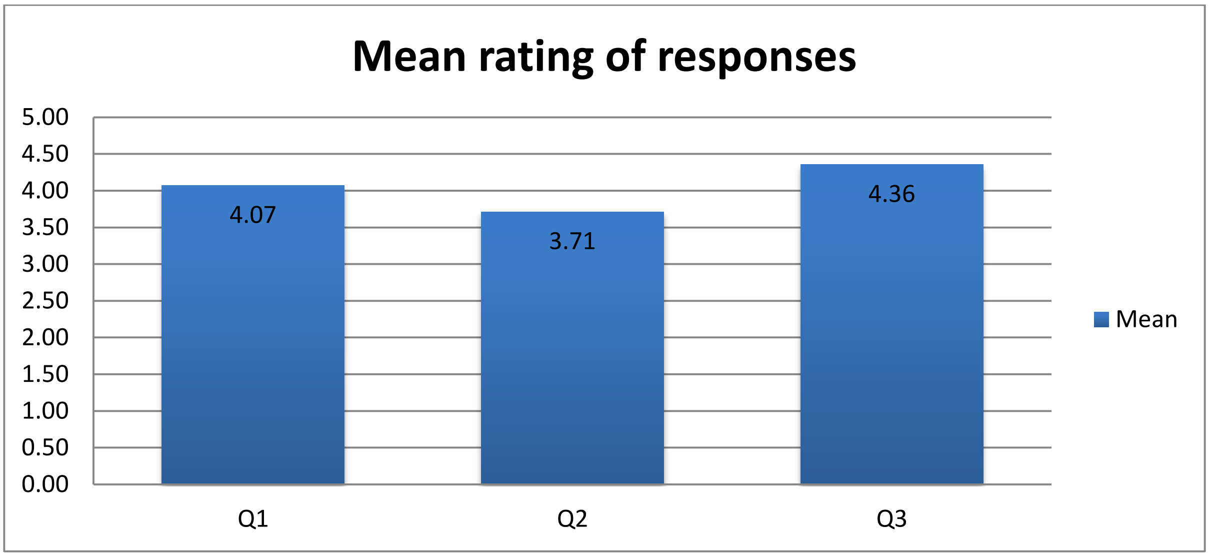

Figure 15: Mean rating of responses (Round 2)

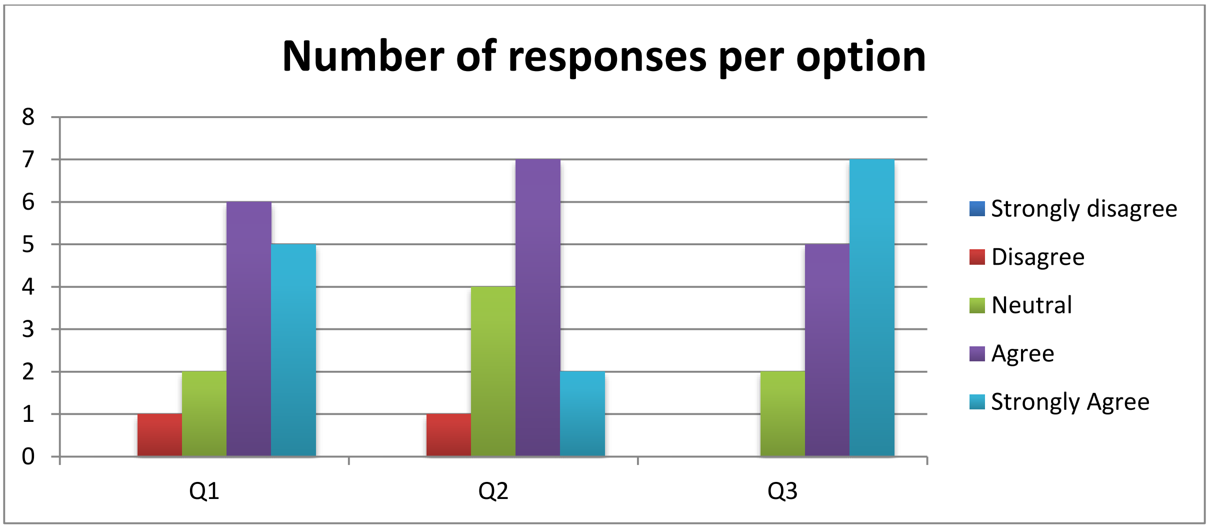

Figure 16: Distribution of responses (Round 2)

7 Conclusions

In this deliverable, we present the two principal outcomes of the last task of the WP:C-5, namely, the catalogue of accountability metrics, and the results of its validation. The catalogue is the product of the application of the elicitation methodology proposed by this work package, which was initiated by the definition of the Metrics Metamodel in the previous deliverable, and continued with the proposal of a complimentary approach based on the analysis of control frameworks, described in this deliverable.

The catalogue of accountability metrics is composed of 39 metrics, organized in three categories: Verifiability and Compliance, devoted to demonstrating compliance to good practices and regulations; Transparency, Responsibility and Attributability; which is related to measuring the characteristics about the internal processes that provide Accountability; and, Remediability and Incident Response, which encompasses metrics related to remediation, redress, and incident response.

Another important objective of this deliverable was to describe our approach for expressing confidence in the results of the metrics. To this respect, we described in Section 5.1 a proposal for extending the concept of measure to a bi-dimensional space, where the confidence in the metrics result is expressed as an orthogonal dimension to the measure performed. The confidence level is derived from analyzing the procedures that lead to the evaluation of particular metrics, with respect to certain quality criteria, namely, consistency and source of assessment.

As for future work, although this work package ends officially with the delivery of this document, it is of paramount importance to keep an effort in continuing the work done. In particular, there is the possibility of influencing standards relevant to the accountability metrics (such as the ISO/IEC 19086, and NIST Cloud Computing Service Metrics). Currently, we are collaborating with WP:A-5 in order to do this. This inclusion will give us the possibility to perform a more reliable and exhaustive validation of the metrics catalogue as they will be used by real cloud providers and for real certification authorities.

Another important application of the work done in this workpackage will be using the metrics for designing an Accountability Maturity Model (AMM). We are currently working on the definition of this AMM where metrics will be used as a guide to determine when an organisation has reached a suitable maturity with respect to accountability.

References

[1] A4Cloud project. D:C-2.1 – Report detailing conceptual framework. To appear.

[2] A4Cloud project. D:C-5.1 – Metrics for Accountability. September 2013.

[3] A4Cloud project. D:C-8.1 – Framework of Evidence. June 2014.

[4] A4Cloud project. MS:C-2.3 – Conceptual Framework. March 2013.

[5] A4Cloud project. MS:C-5.1 – Initial report on metrics for accountability. February 2013.

[6] AICPA/CICA Privacy Task Force. Generally Accepted Privacy Principles. http://www.aicpa.org/INTERESTAREAS/INFORMATIONTECHNOLOGY/RESOURCES/PRIVACY/GENERALLYACCEPTEDPRIVACYPRINCIPLES/Pages/default.aspx

[7] Article 29 Data Protection Working Party. Opinion 10/2004 on More Harmonised Information Provisions. 2004. http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2004/wp100_en.pdf

[8] Centre for Information Policy Leadership (CIPL), “Implementing Accountability in the Marketplace – A Discussion Document. Accountability Phase III – The Madrid Project”. November 2011

[9] Cloud Security Alliance. Cloud Control Matrix (CCM) v3. https://cloudsecurityalliance.org/research/ccm/

[10] Cloud Security Alliance. Consensus Assessments Initiative Questionnaire (CAIQ). https://cloudsecurityalliance.org/research/cai/

[11] Cloud Security Alliance. Open Certification Framework (OCF). https://cloudsecurityalliance.org/research/ocf/

[12] Cloud Security Alliance. Security, Trust and Assurance Registry (STAR). https://cloudsecurityalliance.org/star/

[13] Florida Statutes – Title XXXVII Chapter 627 Insurance Rates and Contracts – Section 627.4145 Readable language in insurance policies. Online: http://law.onecle.com/florida/insurance/627.4145.html

[14] H. Ghani, J. Luna and N. Suri, “Quantitative Assessment of Software Vulnerabilities Based on Economic-Driven Security Metrics” In Proc. of the IEEE International Conference on Risks and Security of Internet and Systems. 2013

[15] H. Ghani, J. Luna and N. Suri, “User-Centric Security Assessment of Software Configurations: A Case Study” In Proc. of the Engineering Secure Software and Systems Conference. 2014

[16] L. Godo, and V. Torra. "On aggregation operators for ordinal qualitative information." Fuzzy Systems, IEEE Transactions on 8.2 (2000): 143-154.

[17] D. S. Herrman. Complete Guide to Security and Privacy Metrics. 2007.

[18] International Organization for Standardization. ISO/IEC 15408 – Information technology – Security techniques – Evaluation criteria for IT security. 2009

[19] International Organization for Standardization. ISO/IEC 27002 – Information Technology, Security Techniques, Code of Practice for Information Security Management. 2005.

[20] International Organization for Standardization. ISO/IEC 27004:2009 (E) – Information Technology – Security techniques – Information Security Management – Measurement. 2009

[21] A. Kolesárová, G. Mayor, and R. Mesiar. "Weighted ordinal means." Information Sciences 177.18 (2007): 3822-3830.

[22] J. Luna, H. Ghani, D. Germanus, and N. Suri, “A Security Metrics Framework for the Cloud”. In Proc. of the International Conference on Security and Cryptography. 2011.

[23] J. Luna, et. al., “Quantitative Assessment of Cloud Security Level Agreements: A Case Study” In Proc. of the International Conference on Security and Cryptography. 2012.

[24] J. Luna, I. Krontiris and N. Suri, “Privacy-by-Design Based on Quantitative Threat Modelling” In Proc. of the IEEE International Conference on Risks and Security of Internet and Systems. 2012.

[25] J. Luna, R. Langenberg and N. Suri, “Benchmarking Cloud Security Level Agreements Using Quantitative Policy Trees” In Proc. of the ACM Cloud Computing Security Workshop. 2012.

[26] F. Moyano, K. Beckers, and C. Fernandez-Gago. "Trust-Aware Decision-Making Methodology for Cloud Sourcing." Advanced Information Systems Engineering. Springer International Publishing, 2014.

[27] J. Mylopoulos, L. Chung, S. Liao, H. Wang, and E. Yu. Exploring alternatives during requirements analysis. Software, IEEE, 18(1), 92-96. 2001

[28] National Institute of Standards and Technology. NIST SP 800-145 – The NIST Definition of Cloud Computing. 2011.

[29] National Institute of Standards and Technology. NIST SP 800-53 – Security and Privacy Controls for Federal Information Systems and Organizations. Revision 4. 2013.

[30] National Institute of Standards and Technology. NIST SP 800-55 – Performance Measurement Guide for Information Security. 2008.

[31] D. Nunez, C. Fernandez-Gago, S. Pearson, and M. Felici, "A metamodel for measuring accountability attributes in the cloud." Cloud Computing Technology and Science (CloudCom), 2013 IEEE 5th International Conference on. Vol. 1. IEEE, 2013.

[32] M. Ouedraogo, et al. "Appraisal and reporting of security assurance at operational systems level." Journal of Systems and Software 85.1 (2012): 193-208.

[33] M. Ouedraogo, et al. "Taxonomy of quality metrics for assessing assurance of security correctness." Software Quality Journal 21.1 (2013): 67-97.

[34] Project Management Institute. A Guide to the Project Management Body of Knowledge (PMBOK® Guide). 5th ed. Newtown Square, PA, USA: Project Management Institute (PMI). 2013.

[35] R. Savola. “Quality of security metrics and measurements”. Computers & Security 37 (2013): 78-90.

[36] R. Savola, and P. Heinonen. "A visualization and modeling tool for security metrics and measurements management." Information Security South Africa (ISSA), 2011. IEEE, 2011.

[37] S. Siegel,“Non parametric statistics,” The American Statistician, vol.11, no. 3, pp. 13–19, 1957.

[38] N. Siegmund. Measuring and Predicting Non-Functional Properties of Customizable Programs. PhD Thesis. 2012

[39] S. Stevens, “On the theory of scales of measurement,” Science, vol. 103, no. 2684, pp. 677–680, 1946.

[40] R. Yager, "On ordered weighted averaging aggregation operators in multicriteria decisionmaking." Systems, Man and Cybernetics, IEEE Transactions on 18.1 (1988): 183-190.

Index of figures

Index of tables

Appendix

1. Describe your level of agreement or disagreement regarding the following statements about the metrics you reviewed. If you have comments regarding specific metrics, you can use the textbox in the next question, indicating the identifier of the concerned metrics. (Available options: Strongly disagree / Disagree / Neither agree nor disagree / Agree / Strongly agree)
    - Q1: This set of metrics contains meaningful and relevant measures for Accountability in the Cloud
    - Q2: The use and application of this set of metrics would be easy, in general
    - Q3: This set of metrics can be easily understood by a professional audience

2. Additional comments. If you have comments regarding specific metrics, you can write them here indicating the identifier of the concerned metrics. Finally, please let us know if you miss any relevant metric.

3. The next questions are purely optional, but your responses would be very helpful for us.
    - What is the title that best describes your job?
    - In your work, you are best described as a: Cloud customer / Cloud provider / Other (please specify)
    - Would you be interested in a follow up inquiry? If so, please provide your email address.