Enhanced Assurance About Cloud Service Provision Promises

Michela D'errico, Siani Pearson

Enhanced Assurance about Cloud Service Provision Promises

Michela D'Errico and Siani Pearson

Hewlett Packard Enterprise, Security and Manageability Lab, Bristol, UK
{michela.derrico, siani.pearson}@hpe.com

Abstract. It is envisaged that in future cloud service providers will increasingly be using a Privacy Level Agreement (PLA) to disclose their data protection practices. This is essentially a self-assessment relating to data protection compliance. Many cloud customers may wish for greater ease in comparing PLAs from different providers, as well as increased assurance about what is being claimed. We tackle this issue by proposing: a standardised representation for PLAs that can be used in a number of ways, including automated comparison by software tools; an ontological approach that can be used as a basis for such automated analysis; a way of expressing evidence that supports statements made in the PLA. Evidence plays a core role when obtaining assurance and building trust, so we also present an ontology for evidence and show how the linkage between evidence elements and data protection aspects in PLAs can be realised through an ontology-aware tool prototype we have developed.

Keywords: assurance • evidence • policy enforcement • Privacy Level Agreement • privacy policy.

1 Introduction

Cloud Service Providers (CSPs) can disclose the level of data privacy and protection offered in a Privacy Level Agreement (PLA) [1], which is a Cloud Security Alliance (CSA) standard intended to be used by potential customers to assess data protection related offerings. In this work we envisage the modelling and representation of key information required to be provided in a PLA for different data protection related aspects.

An important source of requirements for CSPs is data protection legislation, which imposes obligations that have to be considered and complied with when offering cloud services. A PLA specifies how the CSP will be compliant with the applicable data protection law; therefore, PLAs are produced by privacy officers of a CSP in the form of natural language documents. Information about how CSPs address different aspects relating to data protection and privacy are reflected in the various sections in which the document is structured. The standardised structure of the agreement is of great help as it provides a way to group the statements found in a privacy policy by the aspect addressed (such as Data Transfer or Breach Notification). PLAv2 [1] has specifically been

developed with the aim of creating an agreement template suitable for containing statements about how a CSP is going to meet obligations set out by the European Union (EU) Data Protection Directive 95/46/EC [2] (DPD in short, hereafter) and other current European data protection requirements[1]. The guidelines provided for each section in the PLA inform CSPs about which information is to be provided. This information will enable a customer to analyse and evaluate how a particular CSP has planned to comply with the European legal data protection framework.

Although the PLA is an important achievement, the evaluation of the information contained within it involves humans analysing the agreement's statements. If we consider an organisational customer wanting to compare services based on the provided PLAs, this task may take a long time as it requires humans to read and compare, section by section, the data protection-related statements. With respect to this issue we see the utility of having a machine readable representation of the policy statements. In fact, the information modelling behind the representation of the PLA policy statements constitutes the basis on which tools supporting the comparison among PLAs can be built. The idea here is to automate as many human-performed tasks as possible, primarily for efficiency reasons (as would be needed for example if hundreds of agreements were available). What we seek to model is the set of core information that can be extracted from the agreements and that are likely to be looked for by consumers when searching for a suitable service. The representation of key PLA privacy policy statements will thus highlight the main offerings from the data protection perspective.

We have been researching the topic of privacy policies and machine readable versions within the A4Cloud project [3], which has been developing a set of tools enabling an accountability based-approach that includes managing policies for fulfillment of obligations. As shown in Fig. 1, the implementation of privacy policies can be seen as a two-phase process. During the first phase (*design time*) CSPs define the set of policies setting out the different aspects of the service provision. This policy definition step can result in the production of formal agreements such as Service Level Agreements (SLA) [4] or PLAs. In the following step the CSP selects the controls that are more suitable to fulfil the defined policies. At the end of this step enforcement systems should be appropriately configured to make the overall service components work as they should.
At runtime the components should behave as they have been instructed during the design time phase, and appropriate monitoring components should run to keep track of actions performed and the results of those actions. This is necessary to verify whether operationally the systems are behaving effectively as planned. From an accountability perspective, as we will expand on in Section 2, CSPs need to be prompt to prove deployed controls and performed actions; therefore, each step of the implementation process needs to produce evidence of the specific tasks that have been carried out. This is represented in Fig. 1 with the arrows linking each single step with an exemplary evi-

[1] On December 15th 2015 the European Commission, the European Parliament and the Council agreed on the draft text of the General Data Protection Regulation that, once approved, will update and replace the DPD. It is likely to come into force in spring 2016 with a two-year transition period for organisations to comply. The PLA WG will continue to work on the PLA to keep it aligned with current privacy laws in Europe.

dence repository. If not appropriately described, evidence and its handling can be complex. This is why we propose an ontology for the semantic annotations of evidence elements that can specify the nature of the evidence element and what this latter is evidence of.

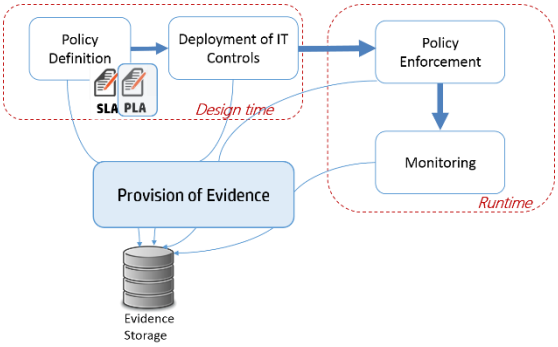

Fig. 1. Provision of Evidence for Accountable Implementation of Privacy Policies

We have analysed the capability of software components in a specific instance of a cloud environment to enforce privacy statements. For statements enforceable by the means of software, we extend the PLA ontology that we have developed by adding properties that allow us to enrich the agreement representation. Specifically, we want to augment the representation with the technical policies created for policy enforcement. We see a technical policy, in itself, as an element that may need to be provided as evidence for demonstration purposes. The technical policy is one of the elements of evidence needed to prove that a policy statement has not just been claimed in an agreement but has also been turned into a software artifact. Evidence produced at runtime contributes to further back up the fulfilment of the policy statements. To establish a link between evidence elements and the policy statements whose demonstration they contribute to support, we propose an ontologically based-approach. Specifically, we create a machine understandable representation of the PLA and then we propose a model to semantically describe evidence elements and link them to the privacy statements.

The remainder of the paper is structured as follows. In Section 2 we present the concept of evidence, along with some definitions, highlighting its importance for the sake of accountability. In Section 3 we present the approach we have taken to turn PLA into a software-exploitable tool. We illustrate an example of ontology modelling applied to the data transfer section of PLA. Then we propose an ontology modelling the concept of evidence (about data processing in the cloud), proposing in particular an approach to link evidence elements to the data protection aspect they underpin. In Section 4 we present the tool that we developed within the context of the A4Cloud project to facilitate the translation of policy statements into instances of the PLA-related ontology models. The tool also enables the creation of system level policies for the setup and configuration of enforcement components that are added to the ontology-based representation of

the PLA as evidence elements. In Section 5 related work is presented. Section 6 concludes the paper by providing some considerations about the work done and outlining directions for future work.

2 Evidence and Assurance in Relation to CSPs' Promises

As considered above, the information provided in the PLA represents data protection related promises that a CSP commits to keep with regard to the provision of a service. A customer selects a service based on an evaluation of the PLAs of the available services. Once a specific service is selected, the related CSP needs to set up and configure the systems so that the PLA terms are met when the customer starts to use the service.

Moving towards the adoption of an accountability-based approach in the provision of a service [5], the CSP must be able to provide the customer with further assurance that the policies are being enforced as stated in the agreement. This assurance can be built upon the evidence elements that the provider can produce and make available to authorised and interested parties in different phases of the cloud service provision.

From the customer point of view, the result of the evaluation of evidence is meant to be used as an indicator of what they can expect from the service in terms of adherence to the promised policies. From the provider standpoint, being able to produce evidence can be an advantage in service markets. Enterprise businesses are more likely to select services provided by CSPs that can guarantee a certain level of assurance with regard to the fulfillment of obligations, especially when the latter are required legally. Provision of evidence is not just a business-driven choice, though. In fact, for providers that want to disclose their practices by using the last release of CSA PLA [1], provision of evidence has been turned into a requirement to be addressed to demonstrate the CSPs' accountability in fulfilling obligations.

2.1 The Role of Evidence

Evidence is strictly bound to the concept of accountability, as it is one of the elements that must be produced and provided to appropriate parties by an organisation that wants to adopt an accountable approach for the provision of a cloud service. A strong accountability approach requires moving from accountability of policies and procedures to accountability of practices [5]. This move requires an organisation to be prompt in providing evidence about how obligations have been fulfilled and not just producing reports based on elements that have been analysed and elaborated by the provider itself. Elements that can be provided as proof of the correct (or incorrect) behaviour of a provider are likely to play an important role. In the PLA the way the provision of evidence has to be disclosed is explained in the guidelines accompanying the accountability section[2] within it, as the provision of evidence is central to the concept of accountability. Evidence is viewed as encompassing different levels. Specifically, "*Evidence elements*

[2] The A4Cloud project, through the authors, has contributed to the text of the accountability section in the PLA.

need to be provided at the (i) Organizational policies level to demonstrate that policies are correct and appropriate; at (ii) IT Controls level, to demonstrate that appropriate controls have been deployed; at iii) Operations level, to demonstrate that systems are behaving (or not) as planned" [1]. Therefore the availability of evidence elements can support the demonstration of the fulfilment of an obligation at different levels, from a declaration level through a documented policy to an operational level through the production of logs or any other system level tangible representation of the processing carried out by the systems. The willingness to produce evidence also reflects the transparency of an organisation and can contribute to building the trust of the customer in the provider.

The need to provide evidence is explicitly stated in the definition of accountability from the EDPS glossary [6] (see text we have underlined), which reads: "*Accountability requires that controllers put in place internal mechanisms and control systems that ensure compliance and provide evidence – such as audit reports – to demonstrate compliance to external stakeholders, including supervisory authorities*."

The Article 29 WP (Opinion 05/12, 2012) [7] also introduces the notion of (documentary) evidence to be provided to back up the asserted compliance to the data protection principles, "[…] *cloud providers should provide documentary evidence of appropriate and effective measures that deliver the outcomes of the data protection principles*".

CSPs need therefore to support assertions about practices adopted with related and relevant evidence artifacts. Example of these types of elements are certifications, logs and technical policies. These elements can be evaluated by customers who will make their choice by taking into account, in addition to the more common privacy policy statements, the information that a provider discloses about the evidence that it is able to produce. Some obligations may matter more than others for a specific customer. For those obligations, a customer is likely to view more favourably a provider that is able to produce a larger set, or more compelling, evidence elements.

2.2 Evaluation of Evidence Upfront and During Service Provision

As considered above, CSPs may want to provide a tangible demonstration to customers and/or auditors that they adopt an accountable approach. Such accountable CSPs would need to design their systems in such a way that they would be prompted to produce the evidence elements that will be required and analysed for demonstration purposes. The evidence can take different forms and can be used to demonstrate accountability at different levels, namely the organizational, controls and operations levels [8].

Evidence has to be produced by CSPs before entering into an agreement with a customer and also just after the contract is signed. In the first phase, CSPs need to provide evidence so that customers can make a more informed service selection. During this phase, customers can evaluate tangible evidence elements, such as documented policies, certifications and privacy seals. These elements represent a type of evidence that has been already produced. A different type of evidence is *promised evidence*, which refers to the elements that CSPs commit to produce. Logs are an example of promised evidence, as they can be produced when the service is in operation. Access to logs and

other forms of evidence produced at runtime may be required to assess whether the provider and the systems set up have behaved as agreed or not. *Promised evidence* is thus key for holding a provider accountable. The two different set of evidence elements that result from this distinction can be used for the evaluation of the accountability attributes *appropriateness* and *effectiveness*. According to the definition of these attributes[3] [9], appropriateness evaluates the *capability* of contributing, therefore evidence provided upfront shall be used for this purpose; effectiveness evaluates the *actual* contribution to accountability, therefore for the purpose of this attribute *promised evidence* elements shall be evaluated when produced at runtime.

Information about the evidence elements available and to be produced can be difficult to analyse and evaluate. We would like to model some aspects of the evidence related disclosure that can be of use for its evaluation. We want to create a model that allows the building of tools that, with reference to a service, can answer questions like "Which types of elements can be produced to demonstrate the fulfilment of this data transfer policy?" To achieve this goal we create an evidence ontology and then link it to the ontology models we created for the PLA sections. In this way it will be possible for a provider to specify whether a specific evidence element has to be used for evaluating the fulfilment of a specific data privacy aspect. Establishing this link will be of help for customers, as there may be some data protection aspects that matter more than others for a customer, and knowing that a CSP can provide a larger set of evidence elements than another CSP, with respect to that aspect, may determine the choice that the customer will make.

The ontology model we build would be used by a provider to create instances of classes and properties that reflect information disclosed about the service. As mentioned above, the duty to produce evidence does not end with the signing of an agreement. The ontology-based representation of the PLA can be enriched with new elements as they are created within the enforcement environment. Ontology classes can be used for tagging the evidence produced at runtime by the enforcement components. Let us consider the case of data transfer policy statements, and let us assume there is a component in charge of monitoring the fulfilment of the restriction about the data processing location. The component can be configured so that the logs for the monitoring of those statements are tagged with the instance of the Data Transfer section being monitored. This would facilitate gathering logs about the monitoring of a specific policy. Our work on PLA formal representation seeks then to structure the concept of evidence into an ontology model and link it to the policy statements that the specific evidence elements aim to prove. This modelling constitutes the base element that can enable the development of a tool that help cloud customers to query, for a specific service, the type of evidence that can be produced to demonstrate the fulfilment of a specific policy.

---

[3] **Appropriateness**: the extent to which the technical and organisational measures used have the capability of contributing to accountability.
Effectiveness: the extent to which the technical and organisational measures used actually contribute to accountability.

In the following section we will present the approach taken for the modelling of the PLA ontology. We will give details about the modelling of the Data Transfer section, as we will be referring to it as an example in Section 4 when we will present the ontology-enabled tool to create the PLA representation.

3 PLA and Evidence Ontology Models

We present in the following subsections a set of classes and properties forming part of our PLA ontology. We present the Data Transfer section as an example of section ontology modelling that shows the approach taken. Then we move to the modelling of evidence.

3.1 PLA Model Overview

We would like the PLA ontology model to reflect the structure of the agreement; thus we model the top level class `PLA` and link it to the `Service` class and to the `DataProtectionAspect` class. A PLA is associated with one and only one service; a service can have just one PLA, therefore we model this two-way relation through the property `isProvidedFor` and its inverse `provides`. The `DataProtectionAspect` class has a list of subclasses which correspond to the sections in the PLA. To give an example, `DataTransfer` and `PersonalDataBreach` are two subclasses.

Data Transfer.
In the Data Transfer section of the PLA, CSPs are required to set out the following set of information:

- Whether data will be transferred
- The reason of the transfer (regular operations or emergency)
- The country where data are transferred to (EEA or outside)
- The legal grounds for the transfer
- The Data Protection role of the recipient of the data being transferred

The Data Transfer ontology is then modelled by turning the above information into the set of properties described below. We give just two examples of the properties we have modelled to represent the information above, as we will be using them in Section 4.

Object property `toCountry`: this specifies the country where data will be transferred. This property ranges over instances of the class `Country`, which can be described by exhaustively listing all the possible instances of the class. The individuals of `Country` class are grouped into relevant areas so that we are able to retrieve additional information from the specific recipient country specified. As we know, for the implementation of DPD the knowledge about the area the country belongs to plays an important role. Knowing whether the country is within or outside the European Economic

Area (EEA) is important as it establishes whether additional safeguards have to be guaranteed by the recipient organisation. For this reason we create an `EEACountry` subclass so that we can also directly obtain the information about the area. For completeness we also model a `EUCountry` class, a subclass of EEACountry, as it can be useful in other contexts.

Modelling the data transfer policy by specifying not just the area (within or outside EEA) but also the specific country (or set of countries) is of importance for organisations that have stricter requirements in this regard. In fact, there are countries where organisations have to maintain stronger data protection policies because it is required by tougher local laws. Having this piece of information represented in the machine readable version of the Data Transfer section allows customers to search for services able to address their specific need about the location.

Object property `hasAdequacyBase`: this property specifies the means by which data transfer adequacy criteria are met. An instance of `DataTransfer` class is linked to an instance of the class `AdequacyBase`. We describe the class by enumerating some individuals belonging to this class, among which there well known legal grounds enabling the transfer of data, namely `ECApprovedModelContractClauses`, `BindingCorporateRules`, `EEAInternalTranfer`, `OtherContractualAgreement`, `Consent`, `Exception`. The instantiation of this property is of high importance when the transfer is to be done towards a country outside the EEA.

3.2 An Ontology for Evidence

Cloud customers evaluate capabilities of CSPs in demonstrating accountability by taking into consideration the set of evidence artifacts that have been produced in advance. The set of evidence elements is enriched as new evidence artifacts are produced while the service is in operation. This type of evidence may be requested at any time for the purpose of monitoring the behaviour of the provider and therefore assessing whether actions (and their effects) are compliant with what is expected.

Before building our evidence ontology, let us review a few definitions drawn for Evidence and Accountability Evidence. This initial analysis will allow us to extract the main properties that characterise an evidence element and possible connections between an evidence element and other concepts. The result of this analysis will be a set of elements which will drive the modelling of the evidence ontology, which will make the knowledge about evidence explicit.

A4Cloud has defined Accountability Evidence [10] as "*collection of data, metadata, 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 support an argument shown to a third party about the validity of claims about the appropriate and effective functioning (or not) of an observable system*". This definition is broad enough to account for evidence provided in different forms, whether raw or derived. What we highlight from this definition is the established linkage between the evidence and the *claims*. In our context, where claims are made in the PLA's sections and the capabilities of producing evidence

are stated in the PLA Accountability section, we model the link between a specific evidence artifact and the specific data protection aspect that the evidence backs up.

From the guidelines accompanying the accountability section in PLA [1] we gather a view on the different forms of evidence, namely evidence "*can take different forms, such as attestations, certifications, seals, third-party audits, logs, audit trails, system maintenance records, or more general system reports and documentary evidence of all processing operations*". This characterisation of the nature of evidence can be translated into an ontology that can be used for classification purposes. Referring to the same section in PLA, evidence can be provided to demonstrate the level of depth that the implementation of the policies has reached. The information about the level is additional information than can be used to augment the description of an evidence artifact. Organisations wanting to adopt an accountable approach need to be prepared to provide evidence at all the levels mentioned above, to prove that policies have not only been declared but are actually being followed in practice. The property of being able to provide evidence at different levels is another feature we aim to model in our ontology so that the populated evidence ontology (i.e. the evidence knowledgebase) can be queried to return the data protection aspects for which all levels have been implemented. It is not the case though that evidence can be provided at all the levels mentioned for all policy statements. For example, operationally, there may not be software-based mechanisms for policy enforcement. This consideration should be taken into account by the actors tasked with evaluation of the evidence-related guarantees.

Evidence Ontology Concepts. In the ontology we aim to model the evidence elements and their use for assessments. At the core of our ontology there is the broad *Evidence* concept that we want to better characterise by specifying specific evidence elements along with their definitions. Our context is cloud service provision and evidence is used for demonstration of effective implementation of organisational policies, IT controls and operations; therefore, in our ontology we model the evidence elements that can be used for these purposes.

A first characterisation can be done by introducing the concept of *Derived Evidence*, which is a form of evidence that has been created by examining a set of evidence elements. We introduce the *Assessment* concept as the general term to express the evaluation of evidence elements that generate *Assessment Reports*, which are therefore classified as *Derived Evidence*. We distinguish two main types of assessment, namely *Audit-Based Assessment* and *Continuous Monitoring Assessment*. The knowledge about the assessment type affects the level of confidence that stakeholders have about the result of the assessment, therefore it is important to be transparent about this type of information. Audit, according to the ISO/IEC 27000 definition [11], is a "*systematic, independent and documented process for obtaining audit evidence and evaluating it objectively to determine the extent to which the audit criteria are fulfilled*". Audit can be performed in different ways, based on what the audit results are to be used for. We distinguish the type of audit based on the actor playing the role of the auditor. *Internal Audit* (also called first party audit) is conducted by the organisation itself for internal purposes, such as process management review. The results of internal audits, in the form of *Audit Reports*, may be requested by customers, and therefore CSPs may wish to include this among the type of (derived) evidence generated. Internal audits can be

used to self-assess conformance of organisations to standards/best practices. *Third-Party Audit* is another type of audit-based assessment which is performed by parties external to the organisation. The results of this type of audit may result in certification, or be used for legal and/or regulatory purposes. They are typically conducted by accredited auditors following specific standard audit procedures. Being produced by actors that have no interest in the organisations being audited, third party audit reports assume importance as they provide a higher level of assurance and consequently customers would show greater trust in them. This is also due to the different factors that may affect the trust relationships between CSPs and customers [12].

Certification is defined in our ontology as a type of derived evidence, whose issuance is based on the result of an assessment. As specific type of certifications we include *Attestation* and *Privacy Seal*. Following the classification by ENISA [13], the *Cloud Certifications* class includes the certification schemes relevant for cloud customers. Each certification has a set of underlying standards or best practices. We model this connection through the relationship *underlyingPractices* between *Certification* and *Practices*, which includes *Standards* and *Best Practices*.

There are many artifacts produced during a cloud service lifecycle, which encompasses phases such as service design, development, deployment, advertisement, operation. Each produced artifact can be seen as evidence associated with a specific activity carried out during a specific phase. It can be very complex to classify every possible evidence element. The objective within this paper is to include most known and used types of evidence, based on their possible exploitation during the service procurement phase. We model the *Policies* evidence type and its two main subclasses *Privacy Policy* and *Security Policy*. Evidence elements can be classified as belonging to this category if they document respectively the privacy policy and the security policy adopted in an organisation. As an example of privacy policy evidence we have added the *Notice* class, which can be also be associated to an URL pointing to the resource location where the notice have been published.

Agreements represent evidence of what has been promised to customers. We distinguish *Legal Agreements*, which are legally binding, and *Service Specification Agreements*, which are documents describing the features of the service. As agreements addressing specific aspects of a service description we include *Privacy Level Agreements* (PLAs) and *Service Level Agreements* (SLAs), which focus on, respectively, data protection aspects and quality of service aspects. A SLA may refer to a PLA, therefore this link should be represented (although it is not explicitly modelled in this ontology).

Enforcement systems, which are in charge of the technical enforcement of what has been described in policies and agreements, need to be configured and set up appropriately before being deployed. The main artifacts produced at this stage can be classified as *Technical Policies* and *Configuration Files*. Technical policies are the representation of documented policies and procedures in low level policy languages. An example is the policy language denoted as A-PPL [14]; policies in A4Cloud are represented by using A-PPL and then enforced by an A-PPL Engine (A-PPLE) [15] able to process A-PPL policies. A component provided with a set of technical policies should act in a way consistent with that established by those policies. The accountability-driven view of the provision of a cloud service requires provision of proof also about the correct behaviour

of the system corresponding to a set of correctly specified policies. Proof of the appropriateness of the policies specified does not imply their effectiveness once they are enforced. Logs will be produced to give evidence of that. *Logs* are the main source of evidence to prove effectiveness of measures. As a specific type of logs we mention *Audit Trails*, which represent records of the sequence of operations leading to a relevant event.

Communications also constitute evidence. The main distinction is between *Informal Communications*, which includes messages exchanged in the form of emails, and *Formal Communications*. This latter, in turn, includes *Notifications*, which represent formal reporting of relevant events, and *Account*, which is defined as a report or description of an event or a process and may be used to communicate audit results and system state. The account provides answers to the "six reporters' questions" by using evidence elements [16].

We want to code the implicit linkage between policy statements in a PLA and the evidence that can be evaluated upfront by a customer to assess to which extent the CSP can be considered accountable for specific data protection aspects. This linkage makes explicit the relation between an evidence artifact and a specific data protection aspect. This linkage can be useful to advise customers on the correct use of the evidence elements available. Evidence elements can apply to different data protection aspects but may also target the demonstration of a specific part of the privacy policy. This is modelled in the ontology through the relationship *isEvidenceOf* which maps evidence elements to the data protection aspect which is the target of the evidence-based demonstration.

This linkage created can also be exploited at runtime. Components creating evidence as a result of the enforcement of a privacy policy statement should be enabled to exploit this link to enrich the knowledge base with a reference, such as a Uniform Resource Identifier (URI), to the evidence being created. The reference can also be an accessible Uniform Resource Locator (URL) that points to a web service endpoint to gather evidence elements of that type being continuously collected. We used the Protégé tool [17] to draw the evidence ontology, which is shown in Fig. 2. This ontology shows the classes to be used for describing the evidence artifacts, along with some relationships that hold among some concepts.

The ontology produced should be seen as a living document to be updated as the knowledge about evidence to be used for accountability purposes is enriched.

3.3 Discussion about Use of the Ontologies

The use of the ontology can help to gather and classify the evidence artifacts produced. Metadata describing a specific evidence element can help answer the following questions: "What is the data protection aspect whose demonstration this evidence is meant to contribute to?" About this question, we remark that in the guidelines of the Accountability section of the PLA there is no linkage suggested to be established between the evidence elements provided and a specific data protection aspect. The evidence ontology model we are going to design will also be of use in this respect. The reasoning capabilities of the ontology will be exploited as the semantic description of instances

of classes such as Logs will automatically be classified as instances of *Evidence* holding the properties of having been produced by *Software Tools*. Information about the type of evidence element can be used to handle this in the right way, based for example on the format of the element provided.
About the use of ontologies there are two main issues that arise which we try to address in the following paragraphs by proposing approaches we may take, in a later stage of our work.

Change Management. Privacy Policies can change over time to adapt the service offerings to updated laws or to reflect changes made in the service implementation. Providers need to track changes and inform the effected customers to let them check that the terms still meet their requirements. Providers need then to have processes in place that track changes and promptly require them to take actions on different tasks that are affected by detected changes. Changes can be handled as events, which can be automatically or manually created (as in the case of an updated law), and the change management process will trigger the execution of tasks which involve the active interaction of the provider. Signatu [18] is an example of a tool that creates the natural language policy and implements a process that alerts providers in case of changes to the law. Iubenda [19], another tool for privacy policy creation, also promises to keep tracking the privacy policy for necessary adaptation to current legislation.

Regulatory Compliance. Use of the Data Protection Policies Tool (DPPT) facilitates the creation of a privacy policy compliant with the DPD. It does so by presenting the provider with options in UI elements that reflect the knowledge we have about possible practices used by providers. However, an additional layer specifically designed to verify the compliance should be introduced to take into account dependencies between different statements that may render the privacy policy not compliant. Application specific compliance checker modules can be designed to verify the compliance with more stringent requirements than the ones derived from the DPD. We see that the development of these compliance checkers can be built by adding rules to be verified over the statements produced. For the time being we have focused on the use of the ontologies for creation of statements to show what the result would be like and how we envisage to use it.

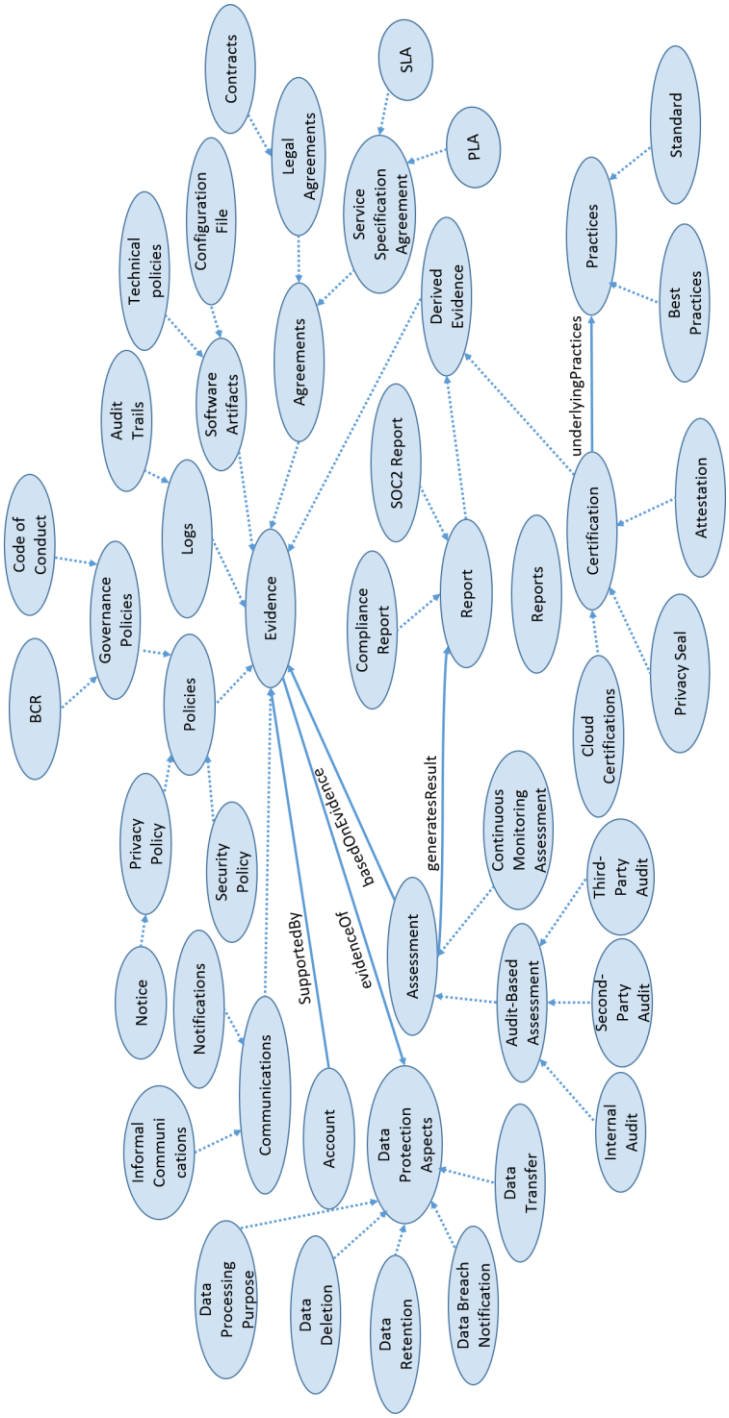

Fig. 2. Evidence Ontology

4 Data Protection Policies Tool

We illustrate how instances of the data transfer ontology concepts introduced in subsection 3.1 can be created with the aid of a Data Protection Policies Tool (DPPT), which is a user-friendly tool that we have developed. Data introduced by the user (who typically would be a policy manager or policy administrator) can be used to create the technical representation of the policy that is then sent to the component tasked with the policy enforcement. This technical policy (written in A-PPL as this is the A4Cloud reference policy language) will then be semantically described as evidence of the data transfer policy. The result will be a Web Ontology Language (OWL) [20] file containing the data transfer policy and the A-PPL policy.

4.1 Creation of Data Transfer Policy Statements

The GUI of the tool presents different panels and reflects the structure of the PLA. The Data Transfer panel is shown in Fig. 3. The elements available in the GUI are bound to the Data Transfer ontology concepts. Actions performed through the GUI, such as typing values in text fields and selecting an option from a combo box, result in the instantiation of corresponding ontology concepts. The GUI layer hides the ontology layer and helps the user in handling ontology-related operations. To clarify this point, let us consider the text field with the label "Country where personal data will be transferred". The name of the country entered by the user, "US" in this example, is an individual of `Country` class and is linked through the object property `toCountry` to the instance of the `DataTransfer` class.

Once the user has provided all the data, an A-PPL policy can be created. We have created A-PPL templates for expressing various policy statements. When we click the "Translate into A-PPL" button, the template for data transfer is used and filled in with the needed data. Depending on the language used and the capability of the enforcement components, all or only a subset of the data may be used. Once the A-PPL policy is created, the tool generates an assertion that declares the policy as an instance of the class `TechnicalPolicy` that is evidence of the Data Transfer instance, as shown in Fig. 4. The A-PPL policy can also be sent to the A-PPLE engine by clicking the related button on the GUI. The policy is sent by using the web service APIs provided by A-PPLE, therefore an URL identifying the endpoint of the web service is used. This information can be integrated into the ontology-based representation of Data Transfer if we add classes describing the enforcement components. In this case we can model an `EnforcementComponent` class and create an instance identifying the A-PPLE engine. We can provide more specific information about A-PPLE by adding the object property `hasEnpoint` which links the A-PPLE instance with an instance of the class `WebServiceEndpoint`, which is a URL.

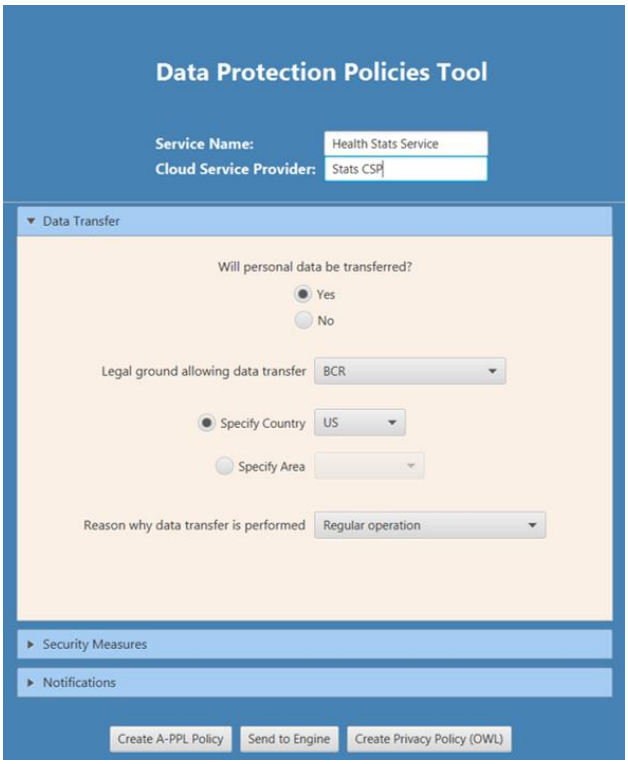

Fig. 3. GUI of Data Transfer Section

```
- <xacml:Actions>
  - <xacml:Action>
    - <xacml:ActionMatch MatchId="urn:oasis:names:tc:xacml:1.0:function:string-equal">
        <xacml:AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">data transfer</xacml:AttributeValue>
        <xacml:ActionAttributeDesignator DataType="http://www.w3.org/2001/XMLSchema#string" AttributeId="action:action-id"/>
      </xacml:ActionMatch>
    </xacml:Action>
  </xacml:Actions>
    <!-- Data must be transfered only the following locations -->
- <xacml:Environments>
  - <xacml:Environment>
    - <xacml:EnvironmentMatch MatchId="urn:oasis:names:tc:xacml:1.0:function:string-equal">
        <xacml:AttributeValue DataType="http://www.w3.org/2001/XMLSchema#string">Germany</xacml:AttributeValue>
        <xacml:EnvironmentAttributeDesignator DataType="http://www.w3.org/2001/XMLSchema#string" AttributeId="environment:environment-
          id"/>
      </xacml:EnvironmentMatch>
    </xacml:Environment>
  </xacml:Environments>
```

Fig. 4. Technical Policy as Evidence of Data Transfer Fulfillment

Having specified that data will be transferred to US, the CSP also needs to select the legal ground allowing the transfer. The CSP in this case selects Binding Corporate Rules (BCRs), which is classified in the ontology as a subclass of Governance Policies. To instantiate this class the CSP needs to provide an URI identifying the BCR text that can then be mapped to Data Transfer class, which is a subclass of Data Protection Aspects. The CSP can also specify that Compliance Reports will be produced during service operation to substantiate the legal compliance of the transfer being performed. This type of evidence further substantiate the data transfer aspect and its production may be

based on logs. This information can also been specified at the moment manually (that is, not through the aid of a GUI).

The instances created will be exploited to answer queries about the evidence available for the data transfer data protection aspect.

If we want to know which evidence elements are available for the Data Transfer policy, we need to query the populated ontology by using the DL Query Tab in the Protégé tool, and we obtain references to BCR, compliance report and technical policy (this latter is added to the knowledge base by the DPPT tool). The query and the result are shown in Fig. 5.

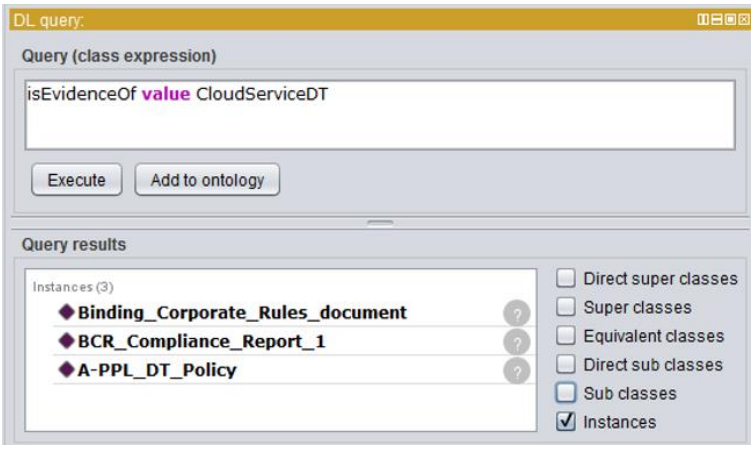

Fig. 5. Evidence Elements Query

5 Related Work

PLA as a research initiative was launched only in 2012 and PLA v2 was released in June 2015. We are not aware of available examples of real instances of PLA, nor of published work about software-based exploitable PLA, which we introduce with our work. However, there has been significant research in the area of SLAs [4], of which PLA is positioned to be supplementary and privacy-focused, and there are a few projects that have addressed the modelling of a machine readable SLA to be exploited by software tools. In particular, the SLA@SOI project [21] has created a SLA model for service lifecycle management [22] [23]. The SPECS project [24] addresses the topic of automating the management of security-oriented SLA. To this aim, the problem of the definition of a machine readable format for the SLA is tackled. SPECS introduces a SLA security conceptual model and proposes an XML schema for this model [25]. As an example of an ontology-based approach to enable SLA management we cite [26]. Significant research has been carried out on the representation of privacy policies in a machine readable format. Among relevant background in this area we cite W3C P3P [27], which developed a platform enabling web sites to express their data collection privacy practices in an XML standard format known as a *P3P policy*, and extensions of

this approach in the PRIME project [28] and PRIMELife project [29], although many other approaches have been taken.

There has also been prior work in ontology supported policy generation [30] based on the mapping between single eXtensible Access Control Markup Language (XACML) syntax elements and legal requirements modelled in an ontology. The advantage of the approach we propose is that we can utilize a task that needs to be completed by organisations in any case (namely the provision of SLAs and PLAs), and then automatically generate information that has a clear business benefit (namely, provision of assurance that can be used to generate trust).

Evidence-related topics have been tackled in different work for different purposes that relate to evidence generation, the gathering/collection of evidence, secure storage of produced evidence, protocols for evidence retrieval and evidence analysis. For example, Ruebsamen et al. in [31], as part of work carried out within the A4Cloud project, have tackled the design of a system for secure collection and storage of digital evidence to address the requirements imposed for the purpose of accountability audits. Various evidence sources are considered for evidence collection, with software agents specifically developed for collection of evidence from a specific source. The knowledge about the features of the source and the evidence produced has been used to design the software agents but has not been made explicit once the evidence is generated. This is one of the purposes of our work, to add metadata to evidence elements by means of ontologies so that they are given meaning by different systems being made to be ontology-enabled.

The role of logs and the importance they assume as accountability evidence is discussed by Ruebsamen et al. in [32], where the need for mappings between evidence data and high level requirements is also raised.

Semantic description of evidence has been tackled in the field of digital investigations for automating the process of integration of evidence [33], which often relies on expertise of expert practitioners who manually perform this task. Dosis et al. [33] in their work describe their ontology-based method to integrate digital evidence. They have specified a number of ontologies for describing sources of evidence commonly used in digital investigations, such as storage media and network traffic. With respect to this work, our ontology can be considered as an upper ontology of evidence, whose concepts can be expanded into additional ontologies in which the knowledge of a specific component can be modeled. Brady at al. [34] also propose an ontology – the Digital Evidence Semantic Ontology (DESO) – to describe devices that are sources of digital evidence. DESO has been developed to support digital evidence examiners in their job of entailing the classification and comparison of digital evidence artifacts.

The Cloud Trust Protocol [35] is a research initiative launched by CSA that aims to provide cloud customers with mechanisms to make queries about *elements of transparency* that can help build evidence-based trust towards the CSP. The link with accountability is not explicitly mentioned but, according to the view developed in the A4Cloud project [2], this type of mechanism contributes to displaying accountability.

Evidence for demonstrating accountability is addressed by the Privacy Office Guide produced by the Nymity Research Initiative [36]. Evidence is seen as one of the three fundamental elements of accountability (together with ownership and responsibility).

Nymity has developed a Privacy Management Accountability Framework which identifies 13 processes for which accountability has to be supported through assessments based on collected evidence.

6 Conclusions

In this paper we have presented an ontology-based approach to create a machine readable representation of the PLA. To ease the creation of PLA ontology instances we have developed a prototype of a GUI-based tool which presents the user with policy statements that have to be disclosed for a specific section of the PLA, and automatically generates a corresponding machine readable representation that can be used in a number of ways.

Provision of evidence is key for an organisation that wants to adopt an accountable approach for service provision. We propose an ontology modelling the concept of evidence and its linkage with privacy policy statements. This modelling allows a semantic description of the evidence elements produced according to their nature. Information about evidence is added to the ontology-based representation of the PLA and can be processed and exploited by customer side tools to extract information about the evidence produced to demonstrate the fulfilment of a data protection aspect. We have shown a specific example related to a Data Transfer policy where the evidence is provided in the form of a technical policy and BCRs, but additional elements could be added in the same way.

Future directions of this research include that we seek to keep working on the tool and the ontologies to keep these aligned with current legal obligations and formal evidence documents being produced and used. We also plan to integrate the tool into a specific real enforcement environment to test the usefulness of the ontology-based tagging.

Acknowledgment. This work has been partly funded from the European Commission's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 317550 (A4CLOUD) Cloud Accountability Project.

References

1. CSA Privacy Level Agreement, https://downloads.cloudsecurityalliance.org/assets/research/pla/downloads/2015_05_28_PrivacyLevelAgreementV2_FINAL_JRS5.pdf
2. European Commission (EC): Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (1995)
3. Cloud Accountability Project (A4Cloud), http://www.a4cloud.eu/
4. Patel, P., Ranabahu, A. H., Sheth, A. P.: Service Level Agreement in Cloud Computing. (2009)
5. Pearson, S.: Privacy Management and Accountability in Global Organisations. In: Hansen, M., Hoepman, J., Leenes, R., Whitehouse, D. (eds.) Privacy and Identity Management for

Emerging Services and Technologies. IFIP Advances in Information and Communication Technology, vol. 421, pp. 33-52. Springer Berlin Heidelberg (2014)
6. European Data Protection Supervisor (EDPS): Glossary of terms (2012). http://www.edps.europa.eu/EDPSWEB/edps/site/mySite/pid/71#accountability
7. European DG of Justice (Article 29 Working Party): Opinion 05/12 on Cloud Computing (2012), http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp196_en.pdf.
8. Felici, M., Pearson, S. (eds.): D:C-2.1 Report detailing conceptual framework. Deliverable D32.1, A4CLOUD (2014)
9. Jaatun, M., Pearson, S., Gittler, F., Leenes R.: Towards Strong Accountability for Cloud Service Providers. IEEE 6th International Conference on Cloud Computing Technology and Science (CloudCom), pp. 1001-1006. IEEE (2014)
10. Felici, M.: Cloud Accountability: Glossary of Terms and Definitions. In: Felici, M., Fernández-Gago, C. (eds.) Accountability and Security in the Cloud. LNCS, vol. 8937, pp. 291-306. Springer International Publishing (2015)
11. ISO/IEC 27000, Information *technology — Security techniques — Information security management systems — Overview and vocabulary*European DG of Justice (Article 29 Working Party), "Binding Corporate Rules", http://ec.europa.eu/justice/data-protection/document/international-transfers/binding-corporate-rules/index_en.htm
12. Pearson, S.: Privacy, Security and Trust in Cloud Computing. In: Pearson, S., Yee, G. (eds.) Privacy and Security for Cloud Computing. Computer Communications and Networks, pp. 3-42. Springer London (2013)
13. Cloud Computing Certification - CCSL and CCSM, https://resilience.enisa.europa.eu/cloud-computing-certification
14. M. Azraoui et al., "A-PPL: An Accountability Policy Language", Technical report, Eurecom, 2014.
15. Azraoui, M., Elkhiyaoui, K., Önen, M., Bernsmed, K., Santana de Oliveira, A., Sendor, J.: A-PPL: An Accountability Policy Language. In: Garcia-Alfaro, J., Herrera-Joancomartí, J., Lupu, E., Posegga, J., Aldini, A., Martinelli, F., Suri, N. (eds) Data Privacy Management, Autonomous Spontaneous Security, and Security Assurance. LNCS, vol. 8872, pp. 319-326. Springer, Heidelberg (2015)
16. A4Cloud Consortium: Cloud Accountability Reference Architecture, http://www.a4cloud.eu/content/cloud-accountability-reference-architecture
17. Protege Ontology Editor, http://protege.stanford.edu
18. Signatu (Beta version as of November 18, 2015), https://signatu.com/home
19. Iubenda, http://www.iubenda.com/en
20. Owl 2 web ontology language document overview (second edition), http://www.w3.org/TR/owl2-overview/
21. SLA@SOI Project, http://sla-at-soi.eu/
22. Kearney, K.T., Torelli, F., Kotsokalis, C.: SLA*: An abstract syntax for Service Level Agreements. In: 2010 11th IEEE/ACM International Conference on Grid Computing (GRID), pp. 217-224. IEEE (2010)
23. Kotsokalis, C., Yahyapour, R., Gonzalez, M.A.R.: SAMI: The SLA Management Instance. In: 2010 Fifth International Conference on Internet and Web Applications and Services (ICIW), pp. 303-308. IEEE (2010)
24. SPECS Project, http://www.specs-project.eu/
25. SPECS Consortium: Report on conceptual framework for Cloud SLA negotiation – Initial. Technical Report D2.2.1 (2014)

26. Labidi, T., Mtibaa, A., Gargouri, F.: Ontology-Based Context-Aware SLA Management for Cloud Computing. In: Ait Ameur, Y., Bellatreche, L., Papadopoulos, G. A. (eds) Model and Data Engineering. LNCS, vol. 8748, pp. 193-208. Springer International Publishing (2014)
27. Cranor, L.: Web Privacy with P3P. O'Reilly & Associates (2002)
28. Camenisch, J., Leenes, R., Sommer, D. (eds.) Digital Privacy: PRIME – Privacy and Identity Management for Europe, LNCS 6545, Springer (2011)
29. Ardagna, C.A., Bussard, L., Di, S.D.C., Neven, G., Paraboschi, S., Pedrini, E., Preiss, S., Raggett, D., Samarati, P., Trabelsi, S.: Primelife policy language (20009)
30. Rahmouni, H.B., Solomonides, T., Casassa Mont, M., Shiu, S.: Privacy compliance in european healthgrid domains: An ontology-based approach. In: 22nd IEEE International Symposium on Computer-Based Medical Systems, 2009. CBMS 2009, pp.1-8. IEEE (2009)
31. Ruebsamen, T., Pulls, T., Reich, C.: Secure Evidence Collection and Storage for Cloud Accountability Audits. In: Proceedings of the 5th International Conference on Cloud Computing and Services Science, pages 321-330. ISBN 978-989-758-104-5 (2015)
32. Rübsamen, T., Reich, C., Taherimonfared, A., Wlodarczyk, T., Rong, C.: Evidence for Accountable Cloud Computing Services. Pre-Proceedings of International Workshop on Trustworthiness, Accountability and Forensics in the Cloud (TAFC) (2013)
33. Dosis S., Homem I., Popov O.: Semantic Representation and Integration of Digital Evidence. Procedia Computer Science, vol. 22, pp. 1266 – 1275 (2013)
34. Brady, O., Overill, R., Keppens, J.: Addressing the Increasing Volume and Variety of Digital Evidence Using an Ontology. In: 2014 IEEE Joint Intelligence and Security Informatics Conference (JISIC), pp.176-183. IEEE (2014)
35. CSA Cloud Trust Protocol Initiative, https://cloudsecurityalliance.org/research/ctp/
36. Nymity Research Initiative, A Privacy Office Guide to Demonstrating Accountability (2014), https://www.nymity.com/data-privacy-resources/data-privacy-research/privacy-accountability-book.aspx